Polynomials and operator orderings

Carl M. Bender and Gerald V. Dunne
The Blackett Laboratory, Imperial College, London SW7 2BZ, United Kingdom

(Received 16 October 1987; accepted for publication 30 March 1988)

It is shown that there is an exact one-to-one correspondence between all possible sets of polynomials (both orthogonal and nonorthogonal) and rules for operator orderings. Operator orderings that are Hermitian give polynomials with definite parity. Most of the standard classical orthogonal polynomials are associated with operator orderings that are not particularly simple. However, there is a special one-parameter class of Hermitian operator orderings that corresponds to a class of elegant but little-known orthogonal polynomials called continuous Hahn polynomials.

I. INTRODUCTION

There is a well-known ambiguity that arises when one attempts to quantize a classical system: there is an infinite number of quantum operators corresponding to the classical function $q^n p^n$. It is conventional to specify the possible operator orderings O as a sum

$$O(q^n p^n) \equiv \frac{\sum_{k=0}^{n} a_k^{(n)} q^k p^n q^{n-k}}{\sum_{k=0}^{n} a_k^{(n)}}, \tag{1}$$

where the coefficients $a_k^{(n)}$ may be chosen arbitrarily. Note that if $a_k^{(n)} = a_{n-k}^{(n)*}$ then the ordering is Hermitian.[1] Some of the better-known correspondence rules are[2]

(i) symmetric ordering, for which $a_0^{(n)} = a_n^{(n)} = 1$ and $a_k^{(n)} = 0$, for $0<k<n$;

(ii) Born–Jordan ordering, for which $a_k^{(n)} = 1$;

(iii) Weyl ordering,[3] for which $a_k^{(n)} = \binom{n}{k}$.

In this paper, we point out that with every operator correspondence rule O one can associate a class of polynomials $P_n(x)$ defined by

$$O(q^n p^n) \equiv P_n[O(qp)]. \tag{2}$$

Note that when we are given a particular rule O it is necessary to reorder the operators p and q to put $O(q^n p^n)$ into the polynomial form $P_n[O(qp)]$. The reordering of the operators p and q requires the use of the commutation relation $[q,p] = i$. This commutation relation is characteristic of the Heisenberg algebra, and therefore the polynomials that are generated by this procedure must reflect the structure of the Heisenberg algebra. A different algebra would give rise to other sets of polynomials.

II. THEORY

It is easy to see that Hermitian orderings will give rise to polynomials having definite parity. To illustrate the association of a set of polynomials with a Hermitian correspondence rule we write down the first seven most general rules of the form (1):

$$\begin{aligned}
O(q^0p^0) &= 1,\\
O(q^1p^1) &= \tfrac{1}{2}(qp+pq),\\
O(q^2p^2) &= \frac{ap^2q^2 + bqp^2q + aq^2p^2}{2a+b},\\
O(q^3p^3) &= \frac{cp^3q^3 + dqp^3q^2 + dq^2p^3q + cq^3p^3}{2(c+d)},\\
O(q^4p^4) &= \frac{ep^4q^4 + fqp^4q^3 + gq^2p^4q^2 + fq^3p^4q + eq^4p^4}{2e+2f+g},\\
O(q^5p^5) &= \frac{hp^5q^5 + jqp^5q^4 + kq^2p^5q^3 + kq^3p^5q^2 + jq^4p^5q + hq^5p^5}{2(h+j+k)},\\
O(q^6p^6) &= \frac{rp^6q^6 + sqp^6q^5 + tq^2p^6q^4 + uq^3p^6q^3 + tq^4p^6q^2 + sq^5p^6q + rq^6p^6}{2r+2s+2t+u},
\end{aligned} \tag{3}$$

where $a,b,...,u$ are arbitrary real numbers. The polynomials corresponding to these ordering rules are

$$\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= x^2 - (6a-b)/4(2a+b),\\
P_3(x) &= x^3 - ((23c-d)/(4c+4d))x,\\
P_4(x) &= x^4 - \left(\frac{86e+14f-5g}{4e+4f+2g}\right)x^2 + \left(\frac{210e-30f+9g}{16(2e+2f+g)}\right),\\
P_5(x) &= x^5 - \left(\frac{115h+35j-5k}{2(h+j+k)}\right)x^3 + \left(\frac{1689h-71j+9k}{16(h+j+k)}\right)x,
\end{aligned} \tag{4}$$

[a] On sabbatical from Department of Physics, Washington University, St. Louis, Missouri 63130.

 0022-2488/88/081727-05$02.50

$$P_6(x) = x^6 - \left(\frac{1010r + 410s + 50t - 35u}{4(2r+2s+2t+u)}\right)x^4 + \left(\frac{24278r + 1478s - 682t + 295u}{16(2r+2s+2t+u)}\right)x^2 - \left(\frac{20790r - 1890s + 630t - 225u}{64(2r+2s+2t+u)}\right).$$

Notice that there is an exact one-to-one correspondence here between polynomials and operator ordering rules. For example, to specify a degree-2 polynomial which has even parity requires one parameter apart from overall normalization. Similarly, apart from overall normalization, $O(q^2p^2)$ is determined by one parameter. In the above formulas this parameter is b/a. In general, apart from overall normalization, both a definite-parity polynomial P_n of degree n and a Hermitian ordering rule $O(q^np^n)$ are uniquely determined by precisely $[n/2]$ parameters, where $[m]$ means *integer part of m*.

A convenient shorthand representation of the order rule O is to list the values of $a_k^{(n)}$ in a pyramidlike table, where $a_k^{(n)}$ is the k th entry in the nth row. The pyramid corresponding to (3) is

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & a & & b & & a & & & & \\
 & & & c & & d & & d & & c & & & \\
 & & e & & f & & g & & f & & e & & \\
 & h & & j & & k & & k & & j & & h & . \\
r & & s & & t & & u & & t & & s & & r \\
\cdot^{\cdot^{\cdot}} & & & & & & & & & & & & \ddots
\end{array} \qquad (5)$$

Notice that if all the entries in any horizontal row are multiplied by an arbitrary number then the operator ordering rule is unchanged.[4]

There are now two natural questions to ask. First, given an ordering rule, i.e., given a pyramid, are the corresponding polynomials orthogonal? Second, if we require that the polynomials be orthogonal, what kind of constraint on the pyramid entries does this give?

To discuss the question of orthogonality, consider an arbitrary set of definite-parity polynomials:

$$\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= x^2 + \alpha,\\
P_3(x) &= x^3 + \beta x,\\
P_4(x) &= x^4 + \gamma x^2 + \delta,\\
P_5(x) &= x^5 + \epsilon x^3 + \zeta x,\\
P_6(x) &= x^6 + \eta x^4 + \theta x^2 + \kappa,\\
&\vdots
\end{aligned} \qquad (6)$$

Suppose these polynomials belong to an orthogonal set. The orthogonality condition is that

$$\int dx\, w(x) P_n(x) P_m(x) = 0 \quad (n \neq m), \qquad (7)$$

where $w(x)$ is the (assumed positive) weight function with respect to which these polynomials are orthogonal and we do not specify the range of integration. We use the notation

$$\langle x^n \rangle \equiv \int dx\, w(x) x^n. \qquad (8)$$

The orthogonality of P_0 and P_2 implies that

$$\langle x^2 \rangle + \alpha \langle 1 \rangle = 0, \qquad (9)$$

the orthogonality of P_0 and P_4 implies that

$$\langle x^4 \rangle + \gamma \langle x^2 \rangle + \delta \langle 1 \rangle = 0, \qquad (10)$$

and the orthogonality of P_1 and P_3 implies that

$$\langle x^4 \rangle + \beta \langle x^2 \rangle = 0. \qquad (11)$$

Eliminating the positive quantities $\langle x^{2n} \rangle$ from (9)–(11) gives a condition that the polynomial coefficients must satisfy:

$$\delta + \alpha\beta - \alpha\gamma = 0. \qquad (12)$$

Equation (12) is actually the first of an infinite tower of algebraic constraints which the coefficients of a set of orthogonal polynomials must satisfy. The next three constraints analogous to (12) are[5]

$$\delta + \beta\epsilon - \beta\gamma - \zeta = 0, \qquad (13)$$

$$\alpha\theta - \alpha\eta\beta + \alpha\beta\gamma - \alpha\delta - \kappa = 0, \qquad (14)$$

$$\eta\delta - \eta\beta\gamma + \theta\beta - \kappa - \epsilon\delta + \epsilon\beta\gamma - \zeta\beta = 0. \qquad (15)$$

Because there is a one-to-one correspondence between polynomials and operator orderings we can immediately translate the constraints in (12)–(15) into conditions which must be satisfied by the coefficients $a_k^{(n)}$ in the definition of the operator ordering rule in (1). Using the notation in (3) we see that the pyramid entries must satisfy nontrivial orthogonality conditions, the first two of which are

$$3adg - bcg + 9acg - 10adf - 2bcf + 2acf + 16bde - 26ade + 14bce - 14ace = 0, \qquad (16)$$

$$9ceh + 17cfh + 10cgh - cej + 7cfj + 5cgj - 11cek - 3cfk - 9deh - 10dfh - 5dgh + dej + dek = 0. \qquad (17)$$

We have only listed these two orthogonality conditions as the others are excessively complicated.

Let us now consider the three commonly used orderings mentioned earlier in this paper. For the case of symmetric ordering the pyramid of coefficients has the form

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 0 & & 1 & & & & \\
 & & & 1 & & 0 & & 0 & & 1 & & & \\
 & & 1 & & 0 & & 0 & & 0 & & 1 & & \\
 & 1 & & 0 & & 0 & & 0 & & 0 & & 1 & \\
1 & & 0 & & 0 & & 0 & & 0 & & 0 & & 1\ , \\
\cdot^{\cdot^{\cdot}} & & & & & & \vdots & & & & & & \ddots
\end{array}$$

and the first seven polynomials defined by (2) are

$$
\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= x^2 - \tfrac{3}{4},\\
P_3(x) &= x^3 - \tfrac{23}{4}x,\\
P_4(x) &= x^4 - \tfrac{43}{2}x^2 + \tfrac{105}{16},\\
P_5(x) &= x^5 - \tfrac{115}{6}x^3 + \tfrac{563}{16}x,\\
P_6(x) &= x^6 - \tfrac{505}{4}x^4 + \tfrac{12139}{16}x^2 - \tfrac{10395}{64}.
\end{aligned}
$$

These polynomials do not satisfy any of the orthogonality conditions (12)–(15). For the case of Born–Jordan ordering the pyramid of coefficients has the form

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 1 & & 1 & & & & \\
 & & & 1 & & 1 & & 1 & & 1 & & & \\
 & & 1 & & 1 & & 1 & & 1 & & 1 & & \\
 & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & ,\\
1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1\\
\iddots & & & & & & \vdots & & & & & & \ddots
\end{array}
$$

and the first seven polynomials are

$$
\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= x^2 - \tfrac{5}{12},\\
P_3(x) &= x^3 - \tfrac{11}{4}x,\\
P_4(x) &= x^4 - \tfrac{19}{2}x^2 + \tfrac{189}{80},\\
P_5(x) &= x^5 - \tfrac{145}{6}x^3 + \tfrac{1627}{48}x,\\
P_6(x) &= x^6 - \tfrac{205}{4}x^4 + \tfrac{3619}{16}x^2 - \tfrac{19305}{448}.
\end{aligned}
$$

One again, we see that these polynomials do not form an orthogonal set.

Now consider the case of Weyl ordering, for which the pyramid of coefficients is Pascal's triangle of binomial coefficients:

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & .\\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1\\
\iddots & & & & & & \vdots & & & & & & \ddots
\end{array}
$$

The first seven associated polynomials are

$$
\begin{aligned}
P_0(x) &= 1,\\
P_1(x) &= x,\\
P_2(x) &= x^2 - \tfrac{1}{4}\\
P_3(x) &= x^3 - \tfrac{5}{4}x,\\
P_4(x) &= x^4 - \tfrac{7}{2}x^2 + \tfrac{9}{16},\\
P_5(x) &= x^5 - \tfrac{15}{2}x^3 + \tfrac{89}{16}x,\\
P_6(x) &= x^6 - \tfrac{55}{4}x^4 + \tfrac{439}{16}x^2 - \tfrac{225}{64}.
\end{aligned}
\tag{18}
$$

Remarkably, these seven polynomials satisfy all four orthogonality conditions (12)–(15). Indeed, the full set of polynomials generated in this way is orthogonal. These polynomials were previously discovered in the context of the quantum finite-element approximation.[6] They are a special case of a four-parameter class of orthogonal polynomials known as continuous Hahn polynomials.[7–11] We briefly review the properties of the polynomials (18) below[12]: (i) the weight function $w(x) = \mathrm{sech}(\pi x)$; (ii) the orthogonality relation is

$$\int_{-\infty}^{\infty} dx\, w(x) P_n(x) P_m(x) = \delta_{nm} \frac{(n!)^2}{2^{2n}};$$

(iii) the recurrence relation is

$$P_n(x) = xP_{n-1}(x) - \tfrac{1}{4}(n-1)^2 P_{n-2}(x);$$

(iv) the generating function is

$$\sum_{n=0}^{\infty} \frac{t^n P_n(x)}{n!} = \frac{e^{2x \arctan(t/2)}}{\sqrt{1 + t^2/4}};$$

and (v) a closed form expression for the nth polynomial may be given in terms of a generalized hypergeometric function:

$$P_n(x) = [(-i)^n n!/2^n]\,{}_3F_2(-n, n+1, \tfrac{1}{4} - i(x/2); \tfrac{1}{2}, 1; 1).$$

One might wonder what the pyramid of coefficients looks like for some of the better known sets of classical orthogonal polynomials. For example, for the Chebyschev polynomials

$$
\begin{aligned}
T_0(x) &= 1,\\
T_1(x) &= x,\\
T_2(x) &= x^2 - \tfrac{1}{2},\\
T_3(x) &= x^3 - \tfrac{3}{4}x,\\
T_4(x) &= x^4 - x^2 + \tfrac{1}{8},\\
T_5(x) &= x^5 - \tfrac{5}{4}x^3 + \tfrac{5}{16}x,\\
T_6(x) &= x^6 - \tfrac{3}{2}x^4 + \tfrac{9}{16}x^2 - \tfrac{1}{32},
\end{aligned}
$$

and the corresponding pyramid of operator ordering coefficients is

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 3 & & 2 & & 3 & & & & \\
 & & & 5 & & 25 & & 25 & & 5 & & & \\
 & & 7 & & 84 & & 202 & & 84 & & 7 & & \\
 & 11 & & 327 & & 1582 & & 1582 & & 327 & & 11 & .\\
9 & & 586 & & 5447 & & 10956 & & 5447 & & 586 & & 9\\
\iddots & & & & & & \vdots & & & & & & \ddots
\end{array}
$$

The entries in this pyramid do not seem to follow any obvious pattern.

The pyramid for the Hermite polynomials,

$$H_0(x) = 1,$$
$$H_1(x) = x,$$
$$H_2(x) = x^2 - \tfrac{1}{2},$$
$$H_3(x) = x^3 - \tfrac{3}{2}x,$$
$$H_4(x) = x^4 - 3x^2 + \tfrac{3}{4},$$
$$H_5(x) = x^5 - 5x^3 + \tfrac{15}{4}x$$
$$H_6(x) = x^6 - \tfrac{15}{2}x^4 + \tfrac{45}{4}x^2 - \tfrac{15}{8},$$

is

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 3 & & 2 & & 3 & & & & \\
 & & & 7 & & 17 & & 17 & & 7 & & & \\
 & & 25 & & 76 & & 182 & & 76 & & 25 & & \\
 & 27 & & 159 & & 454 & & 454 & & 159 & & 27 & \\
331 & & 2582 & & 11653 & & 16948 & & 11653 & & 2582 & & 331 \\
\cdot^{\cdot^{\cdot}} & & & & & & \vdots & & & & & & \ddots
\end{array}$$

Also, the pyramid for the Legendre polynomials,

$$L_0(x) = 1,$$
$$L_1(x) = x,$$
$$L_2(x) = x^2 - \tfrac{1}{3},$$
$$L_3(x) = x^3 - \tfrac{3}{5}x,$$
$$L_4(x) = x^4 - \tfrac{6}{7}x^2 + \tfrac{3}{35},$$
$$L_5(x) = x^5 - \tfrac{10}{9}x^3 + \tfrac{15}{63}x,$$
$$L_6(x) = x^6 - \tfrac{15}{11}x^4 + \tfrac{5}{11}x^2 - \tfrac{5}{231},$$

is

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 7 & & 10 & & 7 & & & & \\
 & & & 17 & & 103 & & 103 & & 17 & & & \\
 & & 203 & & 2948 & & 7138 & & 2948 & & 203 & & \\
 & 583 & & 20091 & & 100286 & & 100286 & & 20091 & & 583 & \\
3491 & & 261462 & & 2511213 & & 5092148 & & 2511213 & & 261462 & & 3491 \\
\cdot^{\cdot^{\cdot}} & & & & & & \vdots & & & & & & \ddots
\end{array}$$

Once again, we observe that the pyramid entries are complicated numbers, which do not fit a recognizable scheme.

Apparently the classical orthogonal polynomials give pyramids with ugly entries. However, Weyl ordering is not the only ordering rule which has elegant pyramid entries and gives orthogonal polynomials. In fact, we have found an infinite one-parameter family of (Hermitian) operator ordering rules which generates a one-parameter class of orthogonal polynomials. The pyramid entries have a simple formula,

$$a_i^{(n)} = \binom{n+l}{i}\binom{n+l}{i+l}\binom{n+l}{l}^{-1}, \tag{19}$$

where l is an arbitrary real parameter. The corresponding polynomials also belong to the large class of continuous Hahn polynomials. The properties of these polynomials are as follows[13]: (i) the weight function $w(x)$ is

$$w(x) = (1/2\pi)|\Gamma(\tfrac{1}{2} + ix)|^2 |\Gamma(l + \tfrac{1}{2} + ix)|^2,$$

and when l is an integer

$$w(x) = \frac{\pi}{2}\operatorname{sech}^2(\pi x)\prod_{j=0}^{l-1}\left\{x^2 + \left[l - \left(j + \frac{1}{2}\right)\right]^2\right\}; \tag{20}$$

(ii) the orthogonality relation is

$$\int_{-\infty}^{\infty} dx\, w(x) P_n(x) P_m(x) = \frac{\pi (n!)^3}{(16)^{n+l}(2n+2l+1)}\left(\frac{\Gamma(n+2l+1)}{\Gamma(2n+2l+1)}\right)^2 \delta_{nm};$$

(iii) the recurrence relation is

$$P_n(x) = xP_{n-1}(x) - \frac{1}{4}\frac{(n-1)^2(n+2l-1)^2}{(2n+2l-1)(2n+2l-3)}P_{n-2}(x);$$

and (iv) a closed-form expression for the nth polynomial is a generalized hypergeometric function,

$$P_n(x) = (-i)^n \frac{\Gamma(n+2l+1)\Gamma(n+1)\Gamma(n+l+1)}{\Gamma(2n+2l+1)\Gamma(l+1)} \times {}_3F_2\left(-n, n+2l+1, \frac{1}{2}-ix; 1, l+1; 1\right).$$

The cases $l=0$ and $l=1$ have particulary elegant pyramids because the entries are all integers with the entries along the outside edges all being 1. When $l=0$ the pyramid is

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 4 & & 1 & & & & \\
 & & & 1 & & 9 & & 9 & & 1 & & & \\
 & & 1 & & 16 & & 36 & & 16 & & 1 & & \\
 & 1 & & 25 & & 100 & & 100 & & 25 & & 1 & \\
1 & & 36 & & 225 & & 400 & & 225 & & 36 & & 1 \\
\cdot^{\cdot^{\cdot}} & & & & & & \vdots & & & & & & \ddots
\end{array},$$

and when $l=1$ the pyramid is

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 3 & & 1 & & & & \\
 & & & 1 & & 6 & & 6 & & 1 & & & \\
 & & 1 & & 10 & & 20 & & 10 & & 1 & & \\
 & 1 & & 15 & & 50 & & 50 & & 15 & & 1 & \\
1 & & 21 & & 105 & & 175 & & 105 & & 21 & & 1 \\
\cdot^{\cdot^{\cdot}} & & & & & & \vdots & & & & & & \ddots
\end{array}.$$

III. CONCLUSION

In this paper we have shown that there is an exact one-to-one correspondence between the ordering of operators satisfying the Heisenberg algebra and the construction of polynomials. We also find that there is a very deep and natural connection between continuous Hahn polynomials, which are orthogonal, and certain classes of Hermitian orderings. The main purpose of this paper is to suggest a new way of revealing the mathematical structure of an operator algebra. By the methods we have used in this paper it is clear that we can examine other algebras besides the Heisenberg algebra. Given that the Hahn polynomials play a special role in the Heisenberg algebra one can ask what the analogous special polynomials are for other algebras. It was not the purpose of this paper to discuss the applications of the connection between operator ordering and polynomials because they are discussed elsewhere. The application of operator ordering and Hahn polynomials to the solution of the unitarity problem in the finite-element approximation to quantum field equations is discussed in Ref. 6. We consider the application of operator ordering and Hahn polynomials to the exact solution of continuum operator Heisenberg equations of motion arising from Euler Hamiltonians in Ref. 14.

ACKNOWLEDGMENTS

One of us (C.M.B.) wishes to thank the Theoretical Physics Group at Imperial College for its hospitality and support and the U.S. Department of Energy for partial financial support. The other (G.V.D.) wishes to acknowledge the support of a Commonwealth Scholarship from the British Council, and an ORS Award.

[1]A useful identity here is $q^kp^nq^{n-k} = p^{n-k}q^np^k$, $n>k$. Thus if $a_k^{(n)} = a_{n-k}^{(n)}$ then the ordering scheme in (1) can also be written as

$$O(q^np^n) = \frac{\sum_{k=0}^n a_k^{(n)}p^kq^np^{n-k}}{\sum_{k=0}^n a_k^{(n)}}.$$

[2]F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions* (Reidel, Boston, 1982), Chap. II.

[3]Weyl ordering is useful in path integration because it corresponds to a midpoint discretization (see Ref. 2). Also, Weyl ordering is identical to totally symmetric ordering which is necessary to preserve unitarity in the linear-finite-element approximation in quantum mechanics (see Ref. 6).

[4]There is a further transformation of the pyramid entries which leaves the identity of the polynomials unaltered. The rescaling $P_n(x) \to (1/\lambda^n)P_n(\lambda x)$ leaves unaltered the orthogonality properties of the polynomials and also leaves the coefficients of the highest power of x in each polynomials equal to 1.

[5]An alternative approach to the derivation of (12)–(15) is to use the well-known fact that a sequence of monic polynomials $P_n(x)$ is orthogonal with respect to a positive measure if and only if they satisfy a three-term recurrent relation of the form

$$P_{n+1}(x) = (x - C_n)P_n(x) - \lambda_n P_{n-1}(x), \quad n>0, \quad \lambda_n>0,$$

provided that $P_1(x) = 1$, $P_1(x) = x - c_1$.

[6]C. M. Bender, L. R. Mead, and S. S. Pinsky, Phys. Rev. Lett. **56**, 2445 (1986).
[7]W. Hahn, Math. Nachr. **2**, 4 (1949).
[8]T. S. Chihara, *An Introduction to Orthogonal Polynomials* (Gordon and Breach, New York, 1978), Chaps. 5 and 6.
[9]M. Atakishiyev and S. K. Suslov, J. Phys. A **18**, 1583 (1985).
[10]R. Askey, J. Phys. A **18**, L1017 (1985).
[11]G. E. Andrews and R. Askey, "Classical orthogonal polynomials," in *Lecture Notes in Mathematics*, edited by C. Brezinski, A. Draux, A. P. Magnus, P. Maroni, and A. Ronveaux (Springer, Berlin, 1985), pp. 36–62.
[12]C. M. Bender, L. R. Mead, and S. S. Pinsky, J. Math. Phys. **28**, 509 (1987).

[13]There is an interesting connection between the weight functions of the continuous Hahn polynomials discussed in this paper and the Euler and Bernoulli numbers. The nth moment of the weight function $w(x) = \operatorname{sech}(\pi x)$ for the polynomials corresponding to Weyl ordering is proportional to the nth Euler number (see Ref. 5). The nth moment of the weight functions in (19) can be expressed as linear combinations of the Bernoulli numbers.

[14]C. M. Bender and G. V. Dunne, Phys. Lett. B **200**, 520 (1988).

Geometrical classification scheme for weights of Lie algebras

Richard H. Capps
Department of Physics, Purdue University, West Lafayette, Indiana 47907

(Received 13 April 1987; accepted for publication 6 April 1988)

Every weight of a Lie algebra belongs to one particular Weyl orbit. The depth within an orbit has been defined as the minimum number of Weyl reflections associated with simple roots that are necessary to transform the weight into the dominant weight of the orbit. A simple set of classification parameters that measures the depth and distinguishes between different weights of the same depth is introduced. The parameters are geometrical in that they are related directly to the orientation of the weight vector. They correspond to a particular member of the Weyl group that transforms the weight into the dominant weight. A simple rule is given for listing all the weights of an orbit in terms of the parameters.

I. INTRODUCTION

In the 1950's Dynkin developed a method for constructing the irreducible representations (irreps) of a finite, simple Lie algebra by subtracting simple roots from the most positive weight.[1] When one uses this method, a useful concept is the level number of a weight M, defined as the number of simple roots that must be subtracted from the most positive weight to obtain M.[2]

This procedure has two serious defects. First, it is too time-consuming to be practical for large irreps. The second defect is that one must keep track of the result of previous steps in the procedure. As a result, one cannot start anywhere in constructing an irrep, but must start at the top or bottom.

One can remedy these defects by combining Weyl symmetry with the Dynkin method. All weights obtainable from a weight M by series of zero or more Weyl reflections comprise the Weyl orbit, or Weyl class, of M. All weights in an orbit have the same multiplicity in every irrep. Thus constructing an irrep may be divided into two parts. One finds the orbits in the irrep and their multiplicities, and one finds the weights in each orbit. One of the earlier authors to emphasize this approach was Humphreys.[3] Some detailed algorithms for finding the orbits in an irrep and their multiplicities have been given in the recent literature.[4,5]

Combining Weyl symmetry with the Dynkin method is useful also for the classification of weights. Each weight belongs to an infinite number of irreps, and so is not associated naturally with any one particular irrep. On the other hand, each weight belongs to only one Weyl orbit. We consider here the problem of classifying the weights of an orbit with a depth parameter that is analogous to the level number in the irrep construction procedure, and "position" parameters to distinguish different weights of the same depth. One straightforward method of constructing the weights in a Weyl orbit has been known and used for years. One starts with the most positive weight and generates the others by series of Weyl reflections associated with the simple roots. Each weight may be labeled by the series of reflections used to obtain it in the construction. This method is reviewed in Sec. II.

The purpose of this paper is to introduce and discuss an alternate, geometrical classification scheme. The constructional and geometrical methods are different; each has some advantages. The relation between the two methods, summarized by Theorem 2 of Sec. V, is illuminating. Two of the attractive features of the geometrical scheme are that it leads to a natural and unique set of position parameters, and it leads to an uncomplicated rule for listing all the weights in an orbit.

Some basic formulas and procedures are reviewed in Sec. II. The classification parameters are defined in Sec. III. Their most important properties are derived in Secs. IV and V. An example is discussed in Sec. VI.

II. BASIC FORMULAS AND PROCEDURES

The standard Cartan–Weyl definition of weights is made. A maximal set of commuting generators of a semisimple algebra is chosen and denoted by $H_1 \cdots H_n$, where n is the rank of the algebra. The H_i are diagonalized for each irrep. If M is a state in an irrep, $H_i M = f_i M$. The weight M is a vector in an n-dimensional Euclidean space with components f_i. A root is a weight of the adjoint irrep.

The orthogonal axes are ranked from 1 to n. A weight is defined as positive if its first nonzero component is positive. A weight A is more positive than B if $A - B$ is positive. A simple root is a positive root that cannot be written as a sum of two positive roots. There are n simple roots. The integral Dynkin components m_j of a weight M are defined by the scalar product equation,

$$m_j = \langle R_j, M \rangle (2/R_j^2) , \quad (2.1)$$

where R_j is a simple root.

The Weyl reflection S_α associated with the nonzero root α permutes the weights of the algebra. The effect of S_α on a weight M is given by[6]

$$S_\alpha (M) = M - \langle \alpha, M \rangle (2/\alpha^2) \alpha . \quad (2.2)$$

If α is a simple root R_i the reflection is called simple and denoted by S_i. It is seen from Eqs. (2.1) and (2.2) that

$$S_i (M) = M - m_i R_i . \quad (2.3)$$

The jth Dynkin component of the simple root R_i is given by

 0022-2488/88/081732-04$02.50

$(r_i)_j = A_{ij}$, where the Cartan matrix A is defined by $A_{ij} = \langle R_i, R_j \rangle (2/R_j^2)$. It follows that the jth Dynkin component of Eq. (2.3) is

$$S_i(M)_j = m_j - m_i A_{ij} . \tag{2.4}$$

The Weyl group consists of all products of any number of Weyl reflections. Therefore every member S of the Weyl group has the property,

$$\langle S(M_1), S(M_2) \rangle = \langle M_1, M_2 \rangle , \tag{2.5}$$

where M_1 and M_2 are arbitrary weights.

A dominant weight is one with no negative Dynkin components. The most positive weight in an orbit is dominant, and is the only dominant weight in the orbit.

It is seen from Eq. (2.3) that the simple reflection $S_i(M)$ leads to a more positive weight, a less positive weight, or the same weight if the Dynkin component m_i is negative, positive, or zero, respectively. I define a positive simple reflection series of a weight M as one in which each reflection corresponds to a negative Dynkin component, and so leads to a weight of greater positivity. Given an arbitrary weight M, one method of obtaining the dominant weight is by making a positive simple reflection series. If one wants to construct the complete orbit corresponding to a dominant weight M^{++}, one can try all possible negative simple reflection series.

Moody and Patera have suggested that the depth of a weight M be defined as the minimum number of simple reflections necessary to proceed from M^{++} to M.[7] This is certainly a reasonable definition.

Each weight of a particular depth may be labeled by listing roots of a minimal simple reflection series from M^{++} to M. There may be many minimal paths, in which case one has a choice of labels. Of course one may use a convention to make the choice unique, such as numbering all the simple roots and resolving all choices of negative simple reflections in favor of the root with smallest (or largest) tag number. However, such a convention is artificial and not very aesthetic.

III. GEOMETRICAL CLASSIFICATION PARAMETERS

In this section the geometrical classification parameters are introduced and some of their properties discussed. The proofs of the statements made are given in Secs. IV and V.

The orbit of a weight is specified in the conventional manner, by listing the Dynkin components m_i^{++} of the dominant weight M^{++}.

The parameters called here position parameters distinguish all weights in an orbit. Let Π^+ be the set of all positive roots. The position of a weight M is specified by listing all the roots π in the set Π^+ that satisfy the relation,

$$\langle \pi, M \rangle < 0 . \tag{3.1}$$

This is called here the signature list. The depth of M is the number of roots in the list. This is identical to the depth defined in Sec. II. Two different weights in an orbit necessarily have different signature lists. This method is geometrical because the roots in the signature list are related directly to the direction of M in weight space. Clearly, the depth is a rough measure of the angular distance between M and M^{++}.

Each orbit belongs to one of a finite number of patterns, where a pattern is defined by the set of Dynkin components m_i^{++} that are zero. The set of signature lists that denote the weights of an orbit is the same for all orbits of the same pattern.

IV. THE DEPTH MEASURE

The equivalence of the depths defined in the constructional procedure (Sec. II) and the geometrical procedure (Sec. III) is shown by the theorem given below. The symbol $N_-(M)$ is used to denote the number of positive roots π that satisfy Eq. (3.1).

Theorem 1: The number of reflections in any positive simple reflection series from M to M^{++} is equal to $N_-(M)$.

It is clear that if M is dominant, $N_-(M) = 0$ and the theorem is satisfied. Hence we consider an arbitrary nondominant weight M, and a simple Weyl reflection S_j associated with a negative Dynkin component m_j. It is sufficient to show that $N_-[S_j(M)] = N_-(M) - 1$. It follows from Eq. (2.1) that $\langle R_j, M \rangle < 0$. Since all diagonal elements of the Cartan matrix are equal to two, it follows from Eq. (2.4) that $S_j(M)_j = -m_j$. Consequently, $\langle R_j, S_j(M) \rangle > 0$. I define Π_j to be the set of all positive roots except R_j. Since the R_j scalar product produces the desired decrease in N_- [as one proceeds from M to $S_j(M)$], it is sufficient for the theorem if the number of negative scalar products $\langle \pi, S_j(M) \rangle$ is equal to the number of negative scalar products $\langle \pi, M \rangle$ when π ranges through the entire set Π_j. In order to show that this is the case, we first note that because of Eq. (2.5) and the condition $S_j^2 = 1$,

$$\langle \pi, S_j(M) \rangle = \langle S_j(\pi), M \rangle . \tag{4.1}$$

A lemma given by Jacobson states that if α is any positive (negative) root other than R_j $(-R_j)$, then $S_j(\alpha)$ is also a positive (negative) root.[8] Since R_j is excluded from Π_j, the transform $S_j(\pi)$ is positive when π is in Π_j. The set $S_j(\Pi_j)$ is identical to the set Π_j. Because of Eq. (4.1), this is sufficient to prove the theorem.

If $\mathscr{A}$ is the algebra under consideration, $\mathscr{A}_0$ is defined as the algebra obtained by writing the Dynkin diagram for $\mathscr{A}$ and deleting each circle (with its connecting lines) that corresponds to a positive Dynkin component of M^{++}. The proof of Theorem 1 can be extended to show that if M^{--} is the most negative weight of an orbit, then

$$N_-(M^{--}) = P(\mathscr{A}) - P(\mathscr{A}_0) , \tag{4.2}$$

where $P(\mathscr{A})$ is the number of positive roots in the algebra $\mathscr{A}$. This formula is given, without proof, in a previous paper.[9]

The algebra $\mathscr{A}_0$ was introduced (with a different notation) in Ref. 7, where it is shown that D_C, the number of weights in the orbit C, is given by

$$D_C = D(\mathscr{A})/D(\mathscr{A}_0) , \tag{4.3}$$

where $D(\mathscr{A})$ is the number of elements in the full Weyl group for the algebra $\mathscr{A}$.

V. CORRESPONDENCE OF PARAMETERS TO A WEYL TRANSFORMATION

In this section the depth and position parameters are related to a particular member S of the Weyl group with the property $S(M) = M^{++}$. First we need a convenient method of labeling all the members of the Weyl group. I define a half-root set to be a set of nonzero roots that contains exactly one of each pair of conjugate roots. Let α be a half-root set and S a Weyl transformation. The roots β_i, defined by $S(\alpha_i) = \beta_i$, must also be a half-root set.

It is convenient to choose the α_i to be π_i, the members of the set Π^+. The signature of each positive root π_i is defined to be $(+1)$ if $S(\pi_i)$ is a positive root, and (-1) if $S(\pi_i)$ is a negative root. Thus if β_1^+ denotes the positive member of the pair $(\beta_i, -\beta_i)$, one may write

$$S(\pi_i) = \beta_i = [\mathrm{Sig}\ \pi_i(S)]\beta_1^+ , \tag{5.1}$$

where $[\mathrm{Sig}\ \pi_i(S)]$ is the signature of π_i for the transformation S. The list of signatures of the π_i specifies the transformation uniquely.[10]

We consider a Weyl transformation S with the property $S(M) = M^{++}$. This condition does not specify S uniquely unless the pattern of M^{++} is $m_i^{++} > 0$ for all i. However, if S is a positive simple reflection series, it is determined uniquely. This is true despite the fact that often there are many different positive simple reflection series that lead from M to M^{++}. In order to prove this point we write the scalar product $\langle \pi_i, M \rangle$ of Eq. (3.1) in the form

$$\langle \pi_i, M \rangle = \langle \pi_i, S^{-1}(M^{++}) \rangle = \langle S(\pi_i), M^{++} \rangle , \tag{5.2}$$

where Eq. (2.5) has been used. If one combines Eq. (5.2) with Eq. (5.1), the result is

$$\langle \pi_i, M \rangle = [\mathrm{Sig}\ \pi_i(S)] \langle \beta_i^+, M^{++} \rangle . \tag{5.3}$$

The scalar product of a positive root and a dominant weight cannot be negative. Therefore, if $\langle \pi_i, M \rangle \neq 0$, the sign of this scalar product is that of $[\mathrm{Sig}\ \pi_i(S)]$.

Equation (5.3) cannot be used to determine the signature of π_i if $\langle \pi_i, M \rangle = 0$. However, the lemma of Jacobson,[8] discussed in Sec. IV, can be used to show that the number of π_i in the set Π^+ with negative signatures cannot exceed the number of terms in a positive simple reflection series from M to M^{++}. It follows from this and Theorem 1 of Sec. IV that if $\langle \pi_i, M \rangle = 0$, the signature of π_i is positive. All signatures are determined, so the S is unique.

The list of positive roots satisfying $\langle \pi_i, M \rangle < 0$, used to specify position and depth in Sec. III, is the list of negative signatures of the unique S defined as a positive simple reflection series that transforms M into M^{++}.

The Weyl reflection surfaces divide the weight space into sectors, sometimes called chambers. If all m_i^{++} are positive, the weights of the orbit are not on any sector boundaries. This is the maximal pattern; there is one weight in each sector. It is convenient to make a one-to-one correspondence between the sectors and the members of the Weyl group; each sector V_i corresponds to the transformation that takes V_i to the dominant sector V^{++}. If sector V corresponds to the transformation S, the sector corresponding to S^{-1} is called V^{-1}. If two sectors differ in the signature of only one positive root, they intersect in an $(n-1)$-dimensional boundary, and are called adjacent.

There is a subtle but uncomplicated relation between the constructional and geometrical labeling schemes. This is best illustrated by considering two consecutive sectors in a negative simple reflection series that starts with V^{++}. Let these two sectors be $T(V^{++})$ and $S_j T(V^{++})$, where T is any Weyl transformation and S_j is a simple reflection. These sectors may be far apart. Next we consider the inverses of these two sectors, $T^{-1}(V^{++})$ and $T^{-1}S_j(V^{++})$. These latter two sectors are adjacent because $S_j(V^{++})$ is adjacent to V^{++}, and the transformation T^{-1} preserves angles. Since the number of adjacent pairs is equal to the number of pairs related by a simple reflection, the following theorem is valid.

Theorem 2: Two sectors U and V are related by one simple reflection if and only if the inverse sectors U^{-1} and V^{-1} are adjacent.

We next consider the problem of finding the rules for the signature lists corresponding to the Weyl orbits. We start with the maximal orbits, which correspond to the full Weyl group. The positive roots π_i have the property that no three add up to zero. If $S(\pi_i) = \beta_i$, it follows that the β's also have this property, i.e.,

$$\beta_i + \beta_j + \beta_k \neq 0, \quad \text{for all } i, j, \text{ and } k . \tag{5.4}$$

It is easy to see that this rule may be written in terms of the signatures in the following way:

If $\pi_k = \pi_i + \pi_j$,

and $\mathrm{Sig}\ \pi_i = \mathrm{Sig}\ \pi_j$,

then $\mathrm{Sig}\ \pi_k = \mathrm{Sig}\ \pi_i$.

This signature condition is necessary for all Weyl transformations. It is also sufficient, as may be seen from the following proof. We consider the inverse transformation. Let β be a half-root set that satisfies Eq. (5.4). We need to show that a Weyl transformation exists that transforms the set β into the set Π^+. We assume that the set β is not identical to Π^+. Let π_m be the least positive member of Π^+ that is not in the β set. Thus $(-\pi_m)$ is in the β set. If π_m is not a simple root, it may be written $\pi_m = \pi_a + \pi_b$. Since π_a, π_b, and $(-\pi_m)$ are in the β set, this implies a violation of Eq. (5.4). Consequently, π_m must be simple.

Let n_- be the number of negative roots in the set β. If this set is transformed by the simple Weyl reflection S_m associated with π_m, the Jacobson lemma states that the number n_- is decreased by one.[8] One can construct a series of simple reflections of this type, each decreasing n_- by one, until n_- is zero. The product of these reflections is a Weyl group member then transforms the β set into Π^+.

Next I will write a simple rule for specifying the subset of the allowed signature lists that applies when the orbit is not maximal, i.e., when one or more m_i^{++} is zero.

Orbit Rule: If one or more of the m_i^{++} are equal to zero, the allowed signature lists correspond to the inverses of the transformations for which the signatures of the simple roots corresponding to the zero-valued m_i^{++} are all positive.

This rule is applied in Sec. VI.

If M^{++} has l different zero-valued Dynkin components, all the weights of the orbit are on $(n-l)$-dimensional intersections of two or more sectors. For any $(n-l)$-dimensional boundary, one of the depths of the sectors intersecting at the boundary is smaller than the others. The labeling prescription of Sec. III labels a weight on the boundary with the signature list of the intersecting sector of minimum depth.

VI. EXAMPLE

We discuss as an example the algebra $A_4[\mathrm{SU}(5)]$, in particular the weight W with Dynkin components $(-3\ 5\ -3\ 1)$. Since the number of positive roots of A_4 is only ten, one could determine the dominant weight of the orbit of W fairly quickly by making a simple reflection series. However, I will use the alternate method of working in a special basis, since this technique is fast even for more complicated algebras.[11] The quarks of SU(5) are numbered 1 through 5 in order of decreasing positivity. If $\bar{k}$ denotes the antiquark of quark k, the four simple roots are

$$A=(1\bar{2}),\quad B=(2\bar{3}),\quad C=(3\bar{4}),\quad D=(4\bar{5})\,. \tag{6.1a}$$

The other six positive roots are

$$\begin{aligned} &E=(1\bar{3}),\quad F=(2\bar{4}),\quad G=(3\bar{5})\,,\\ &H=(1\bar{4}),\quad I=(2\bar{5}),\quad J=(1\bar{5})\,. \end{aligned} \tag{6.1b}$$

The normalization is such that these roots are of length $\sqrt{2}$.

If m_i are the Dynkin components of a weight M, integral "quark components" f_i exist that satisfy the equation[12]

$$m_i=f_i-f_{i+1}\,. \tag{6.2}$$

One may choose any one of the five f_i arbitrarily and use Eq. (6.2) to determine the other four. The number f_i represents the number of quarks i minus the number of antiquarks $\bar{i}$ in the weight. The quark components are convenient for expressing the scalar product with a root, i.e.,

$$\langle (i\bar{k}),M\rangle=f_i-f_k\,. \tag{6.3}$$

It follows from Eqs. (2.2) and (6.3) that the different members of an orbit correspond to all distinct permutations of the five f_i. The dominant weight is the permutation such that $f_i \geqslant f_{i+1}$ for all i.

A set of f_i for the weight $W=(-3\ 5\ -3\ 1)$ is $[25032]$. The dominant weight of this orbit is $[53220]$ with Dynkin components $(2\ 1\ 0\ 2)$. It is seen from Eqs. (6.1a), (6.1b), and (6.3) that the set of positive roots with negative signatures for W is $(ACGH)$.

For A_4 it is not difficult to prepare a complete list of the 120 Weyl transformations and their inverses, from which the signature lists of all weights of all orbits may be read.[13] I illustrate by considering the 20 transformations of depth 4. These are listed in Eq. (6.4) below, together with their inverses. The notation $X^*=Y$ indicates that X and Y are inverses, while the subscript s denotes a transformation that is its own inverse. The list is

$$\begin{aligned} &ABDE_s,\quad ABEH^*=ACFH,\quad ACDG_s\,,\\ &ACEH^*=BCFH,\quad ACGH^*=BDEI,\quad ADEJ^*=CFGH\,,\\ &ADGJ^*=BEFI,\quad AEHJ^*=DGIJ,\quad BCFI^*=BDGI\,,\\ &BDFI^*=CDGI,\quad BEFH_s,\quad CFGI_s\,. \end{aligned} \tag{6.4}$$

For example, consider an orbit of the pattern $(+\ +\ 0\,0)$, i.e., $m_3^{++}=m_4^{++}=0$. It is seen from Eq. (6.4) and the orbit rule of Sec. V that the depth-4 weights in the orbit have the signature lists $ACFH$, $ADGJ$, $DGIJ$, and $BEFH$.

In the case of W there are five different positive simple reflection series leading to the dominant orbit weight. If one considers the most negative weight of a maximal orbit of A_4, all ten signatures are negative. For such a weight there are 768 different positive simple reflection series leading to the dominant weight.

VII. CONCLUDING REMARK

In any parametrization, it is desirable that as many parameters as possible are bounded absolutely, in the sense that the bound depends only on the algebra. It is impossible to classify all weights by a finite set of parameters, each of which is bounded in this way, since the number of weights is infinite. However, in the parameter scheme outlined in Sec. III, only the Dynkin components of the M^{++} are unbounded. The number of patterns is 2^n, where n is the rank. The number of weights in an orbit is bounded by the size of the Weyl group, and the number of negative signatures is bounded by the number of positive roots.

ACKNOWLEDGMENT

This paper was supported in part by the U.S. Department of Energy.

[1]E. B. Dynkin, in *American Mathematical Society Translations*, Series I, Vol. 9, pp. 328–469; Series 2, Vol. 6, pp. 111–244.
[2]R. Slansky, Phys. Rep. **79**, 1 (1981), Sec. 5.
[3]James E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972), pp. 107–125.
[4]R. V. Moody and J. Patera, Bull. Am. Math. Soc. 7, 237 (1982).
[5]R. H. Capps, J. Math. Phys. **28**, 1215 (1987), Sec. III.
[6]R. N. Cahn, *Semi-Simple Lie Algebras and Their Representations* (Benjamin/Cummings, Reading, MA, 1984), Chap. XII.
[7]R. V. Moody and J. Patera, SIAM J. Alg. Disc. Meth. **5**, 359 (1984), Sec. 3.
[8]N. Jacobson, *Lie Algebras* (Wiley–Interscience, New York, 1962), Lemma 1, p. 241.
[9]See Ref. 5. Eq. (5.8).
[10]This may be shown by using the theorems of pp. 240–243 of Ref. 8.
[11]Standard bases for many Lie algebras are listed by R. C. King and A. H. A. Al-Qubanchi, J. Phys. A: Math. Gen. **14**, 15 (1981). Many Weyl orbits of E_8 are listed in a standard basis in Ref. 5, Sec. IV.
[12]The basic properties of this basis are discussed by R. H. Capps, J. Math. Phys. **27**, 914 (1986), Sec. III A.
[13]Several methods exist for finding all the Weyl transformations and their inverses. In the case of SU(n) one may use the fact that the Weyl group is the permutation group of the n quarks.

Polynomial icosahedral invariants

C. J. Cummins and J. Patera
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec, Canada

(Received 4 June 1987; accepted for publication 23 March 1988)

In this paper the ring of polynomial invariants of the icosahedral group I is studied. It begins by reviewing the surprising connections that this group and its double cover II have with various areas of mathematics and physics of current interest. Information concerning the representation theory of these groups is then given. After a brief discussion of the methods involved, the integrity basis of the ring of polynomial invariants for each irreducible representation of I is given, together with syzygies. These are expressed in terms of the invariants of the tetrahedral subgroup T of I. Finally other methods of finding these invariants are discussed.

I. INTRODUCTION

The purpose of this paper is to contribute to breaking a "...conspiracy of silence about the icosahedron in the physics literature" (a complaint voiced in a recent review[1] on quasicrystals). Motivation for our work originates in that, within a very short space of time, the interest in icosahedral symmetries has started to grow in completely unrelated fields of mathematics (through its deep relation to the largest of the exceptional simple Lie groups) and in physics in two very different fields, quasicrystals, and (E_8-based) superstring models, probably the "hottest topics" at present.

The icosahedral group[2] I is the largest of the finite subgroups of the SO(3) group. Its distinguishing feature among these groups is that it contains elements of order 5 and therefore it cannot be a symmetry of a two- or three-dimensional crystal (lattice). Until a few years ago it had been of marginal physical interest. Now, however, the presence of fivefold symmetry in solids has been observed,[3] which may be explained in terms of quasicrystals[1] or icosahedral glasses.[4] Consequently there is a role for the icosahedron to play even in physics (in biology the fivefold symmetry of living organisms is long well known[5,6]). In spite of the long history of the subject, a number of obviously important properties of the icosahedral representation theory have not been worked out. This work partially fills the gap.

The discovery of quasicrystals was preceded in mathematics by the discovery of curious, apparently far reaching, but not yet fully understood relations between simple Lie algebras/groups and the finite subgroups of SU(2), in particular, the algebra of type E_8 and the icosahedral group. Therefore when the E_8 group later turned out to be of prime interest in elementary particle physics, one justly wondered about the role of the icosahedral group there.

In this paper we study the structure of the rings of polynomial invariants of I, with variables transforming under one of the five irreducible representations of $I \subset SO(3)$. We find the integrity bases of the invariants together with their syzygies. In a way this completes the results of Ref. 7 where the icosahedral case was left out for lack of motivation (in 1978) and because in itself it is larger than all the finite subgroups of SO(3) studied there.

In Sec. II we point out the amazing relations between E_8 and I. Section III contains relevant information about the tetrahedral subgroup T and its irreducible representations, decomposition of their tensor products, etc., and in Sec. IV we discuss the general structure of the rings of polynomial tensors of representations of any finite group G. The results relevant to our study are given. These three sections form a collection of known facts put together in a form suitable for our purpose. Our results are described in Sec. V (and summarized in Tables IX–XIV). This section also contains a description of the methods of our computation. Finally in Sec. VI we discuss other methods of calculating the integrity bases, and record some alternative conventions.

II. THE ICOSAHEDRAL GROUP

Just as the rotation group SO(3) has double-valued representations that are single-valued representations of the covering group SU(2), so the icosahedral group I has double-valued representations that are single-valued representations of the double icosahedral group $II \subset SU(2)$. The irreducible representations of II are denoted by $\Gamma_1,...,\Gamma_9$, and may be classified as either odd or even. The odd representations are faithful representations of II and are contained in SU(2) representations of even dimension. They are thus the "spinors" of I. The even representations are contained in the SU(2) representations of odd dimension and are not faithful representations of II, but rather of $II/\{1, -1\} \cong I$. They may thus be identified with the irreducible representations of I. In the numbering scheme adopted here the first five representations of II are even and the last four odd. This numbering is chosen to coincide with that of Refs. 1 and 2. Although our calculation will concern only the even representations we shall where convenient include details of the odd ones. In particular, the connections between the icosahedral group and E_8 involve the spin representations.

The character table of II is shown in Table I in which ω denotes the "golden ratio" $(1+\sqrt{5})/2$, a solution of the quadratic equation $x^2 - x - 1 = 0$. The SU(2)-conjugacy classes of each double icosahedral conjugacy class are also included. The notations we use and the properties of the SU(2)-conjugacy classes of elements of finite order were ex-

 0022-2488/88/081736-10$02.50

TABLE I. Character table of the double icosahedral group II, ω is the "golden ratio" $(1+\sqrt{5})/2$. The first three rows give the number of elements in each conjugacy class, the corresponding SU(2) conjugacy class and the orders of the elements in each class.

	C_1	C_1'	C_2	C_2'	C_3	C_3'	C_4	C_5	C_5'
	1	1	12	12	12	12	30	20	20
	[10]	[01]	[32]	[23]	[14]	[41]	[11]	[12]	[21]
	1	2	5	10	5	10	4	3	6
Γ_1	1	1	1	1	1	1	1	1	1
Γ_2	3	3	$1-\omega$	$1-\omega$	ω	ω	-1	0	0
Γ_3	3	3	ω	ω	$1-\omega$	$1-\omega$	-1	0	0
Γ_4	4	4	-1	-1	-1	-1	0	1	1
Γ_5	5	5	0	0	0	0	1	-1	-1
Γ_6	2	-2	$\omega-1$	$1-\omega$	$-\omega$	ω	0	-1	1
Γ_7	2	-2	$-\omega$	ω	$\omega-1$	$1-\omega$	0	-1	1
Γ_8	4	-4	-1	1	-1	1	0	1	-1
Γ_9	6	-6	1	-1	1	-1	0	0	0

haustingly described in Ref. 8. We do not repeat it here. For convenience we have also included the orders of the elements of each conjugacy class, although this information may be deduced from a knowledge of the corresponding SU(2)-conjugacy classes.

In Table II we display the decompositions of the tensor products of the representations $\Gamma_1,...,\Gamma_9$, while in Table III the generators of these irreducible representations are given. We have chosen a form of the generators in which the first two generate a tetrahedral subgroup. This enables us to make use of previous calculations of the invariants and covariants of this subgroup.[7] It should be noted, however, that I (or II) may be generated by just two elements. In the case of the two three-dimensional representations Γ_2 and Γ_3 we have displayed only the generators of Γ_2. Those of Γ_3 are obtained simply by replacing ω by $1-\omega$, that is, by selecting the other root of the equation $x^2-x-1=0$. This comment also applies to the invariants and syzygies described later.

In the case of the even representations of II there exists a method[9] for finding explicitly the required generators by making use of the fact that $\mathrm{I}\cong \mathrm{A}_5$, the alternating group of even permutations of five objects. Once these even generators have been found[10] we may obtain the generators of the two-dimensional fundamental spin representation in the following manner. Let g be an SU(2) matrix and σ^i, $i=1,2,3$, be the Pauli spin matrices, i.e.,

$$g=\begin{bmatrix} a & b \\ -b^* & a^* \end{bmatrix}, \quad |a|^2+|b|^2=1,$$

and

$$\sigma^1=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^2=\begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}, \quad \sigma^3=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The action on g on the σ^i by conjugation now yields an SO(3) transformation $D_{ij}(g)$,

$$g\sigma^i g^{-1}=\sum_j D_{ij}(g)\sigma^j.$$

If $a=a_1+ia_2$ and $b=b_1+ib_2$ then $D_{ij}(g)$ is given by

TABLE II. Decompositions of the tensor products of irreducible representations of II.

	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8	Γ_9
Γ_1	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8	Γ_9
Γ_2		$\Gamma_5\oplus\Gamma_2\oplus\Gamma_1$	$\Gamma_5\oplus\Gamma_4$	$\Gamma_5\oplus\Gamma_4\oplus\Gamma_3$	$\Gamma_5\oplus\Gamma_4\oplus\Gamma_3\oplus\Gamma_2$	$\Gamma_8\oplus\Gamma_6$	Γ_9	$\Gamma_9\oplus\Gamma_8\oplus\Gamma_6$	$2\Gamma_9\oplus\Gamma_8\oplus\Gamma_7$
Γ_3			$\Gamma_5\oplus\Gamma_3\oplus\Gamma_1$	$\Gamma_5\oplus\Gamma_4\oplus\Gamma_2$	$\Gamma_5\oplus\Gamma_4\oplus\Gamma_3\oplus\Gamma_2$	Γ_9	$\Gamma_8\oplus\Gamma_7$	$\Gamma_9\oplus\Gamma_8\oplus\Gamma_7$	$\Gamma_9\oplus\Gamma_8\oplus\Gamma_7\oplus\Gamma_6$
Γ_4				$\Gamma_5\oplus\Gamma_4\oplus\Gamma_3\oplus\Gamma_2\oplus\Gamma_1$	$2\Gamma_5\oplus\Gamma_4\oplus\Gamma_3\oplus\Gamma_2$	$\Gamma_9\oplus\Gamma_7$	$\Gamma_9\oplus\Gamma_6$	$2\Gamma_9\oplus\Gamma_8$	$2\Gamma_9\oplus2\Gamma_8\oplus\Gamma_7\oplus\Gamma_6$
Γ_5					$2\Gamma_5\oplus\Gamma_4\oplus\Gamma_3\oplus\Gamma_2\oplus\Gamma_1$	$\Gamma_9\oplus\Gamma_8$	$\Gamma_9\oplus\Gamma_8$	$2\Gamma_9\oplus2\Gamma_8\oplus\Gamma_7\oplus\Gamma_6$	$3\Gamma_9\oplus2\Gamma_8\oplus\Gamma_7\oplus\Gamma_6$
Γ_6						$\Gamma_2\oplus\Gamma_1$	Γ_4	$\Gamma_5\oplus\Gamma_2$	$\Gamma_5\oplus\Gamma_4\oplus\Gamma_3$
Γ_7							$\Gamma_3\oplus\Gamma_1$	$\Gamma_5\oplus\Gamma_3$	$\Gamma_5\oplus\Gamma_4\oplus\Gamma_2$
Γ_8								$\Gamma_5\oplus\Gamma_4\oplus\Gamma_3\oplus\Gamma_2\oplus\Gamma_1$	$2\Gamma_5\oplus2\Gamma_4\oplus\Gamma_3\oplus\Gamma_2$
Γ_9									$3\Gamma_5\oplus2\Gamma_4\oplus2\Gamma_3\oplus2\Gamma_2\oplus\Gamma_1$

TABLE III. Generators of II for each irreducible representation. Here A_1, A_2, and A_3 correspond to the three permutations (123), (12)(34), and (12)(45) under the identification of II with A_5 (Ref. 10) and $\sigma = e^{2\pi i/3}$.

	A_1	A_2	A_3
$\Gamma_2(\Gamma_3)$	$\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&-1&0\\0&0&-1\end{bmatrix}$	$-\frac{1}{2}\begin{bmatrix}1&\omega&1-\omega\\\omega&1-\omega&1\\1-\omega&1&\omega\end{bmatrix}$
Γ_4	$\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&1&0&0\\0&0&1&0\end{bmatrix}$	$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&-1&0\\0&0&0&-1\end{bmatrix}$	$-\frac{1}{4}\begin{bmatrix}1&\sqrt{5}&\sqrt{5}&\sqrt{5}\\\sqrt{5}&-3&1&1\\\sqrt{5}&1&1&-3\\\sqrt{5}&1&-3&1\end{bmatrix}$
Γ_5	$\begin{bmatrix}0&0&1&0&0\\1&0&0&0&0\\0&1&0&0&0\\0&0&0&\sigma&0\\0&0&0&0&\sigma^2\end{bmatrix}$	$\begin{bmatrix}1&0&0&0&0\\0&-1&0&0&0\\0&0&-1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}0&1&1&-1&-1\\1&1&0&-\sigma&-\sigma^2\\1&0&1&-\sigma^2&-\sigma\\-1&-\sigma^2&-\sigma&0&1\\-1&-\sigma&-\sigma^2&1&0\end{bmatrix}$
Γ_6	$\left(\frac{1}{2}\right)(1+i)\begin{bmatrix}1&i\\1&-i\end{bmatrix}$	$\begin{bmatrix}0&-i\\-i&0\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}-i(1-\omega)&\omega+i\\-\omega+i&i(1-\omega)\end{bmatrix}$
Γ_7	$\left(\frac{1}{2}\right)(1+i)\begin{bmatrix}1&i\\1&-i\end{bmatrix}$	$\begin{bmatrix}0&-i\\-i&0\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}-i\omega&1+i-\omega\\-1+i+\omega&i\omega\end{bmatrix}$
Γ_8	$\left(\frac{1}{2}\right)(1+i)\begin{bmatrix}\sigma\begin{bmatrix}1&i\\1&-i\end{bmatrix}&\begin{matrix}0&0\\0&0\end{matrix}\\\begin{matrix}0&0\\0&0\end{matrix}&\sigma^2\begin{bmatrix}1&i\\1&-i\end{bmatrix}\end{bmatrix}$	$\begin{bmatrix}0&-i&0&0\\-i&0&0&0\\0&0&0&-i\\0&0&-i&0\end{bmatrix}$	$\frac{1}{4}\begin{bmatrix}\sigma^2-\sigma-i&1-i&-i\sigma(\omega+\sigma)&(i\sigma^2-i-1)(\sigma^2\omega+1)\\-i-1&\sigma^2-\sigma+i&(i\sigma^2-i+1)(\sigma^2\omega+1)&i\sigma(\omega+\sigma)\\-i\sigma(\sigma\omega+1)&(i\sigma-i-1)(\sigma\omega+1)&-\sigma^2+\sigma-i&1-i\\(i\sigma-i+1)(\sigma\omega+1)&i\sigma(\sigma\omega+1)&-i-1&-\sigma^2+\sigma+i\end{bmatrix}$
Γ_9	$\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}\otimes\left(\frac{1}{2}\right)(1+i)\begin{bmatrix}1&i\\1&-i\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&-1&0\\0&0&-1\end{bmatrix}\otimes\begin{bmatrix}0&-i\\-i&0\end{bmatrix}$	$-\frac{1}{2}\begin{bmatrix}1&\omega&1-\omega\\\omega&1-\omega&1\\1-\omega&1&\omega\end{bmatrix}\otimes\frac{1}{2}\begin{bmatrix}-i\omega&1+i-\omega\\-1+i+\omega&i\omega\end{bmatrix}$

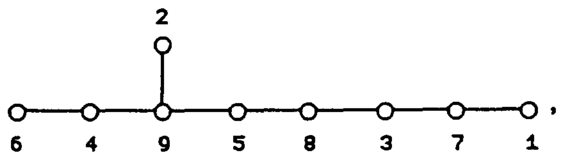
$$\begin{bmatrix} a_1^2 - a_2^2 - b_1^2 + b_2^2 & -2(a_1a_2 + b_1b_2) & -2(a_1b_1 - a_2b_2) \\ 2(a_1a_2 - b_1b_2) & a_1^2 - a_2^2 + b_1^2 - b_2^2 & -2(a_1b_2 + a_2b_1) \\ 2(a_1b_1 + a_2b_2) & 2(a_1b_2 - a_2b_1) & a_1^2 + a_2^2 - b_1^2 - b_2^2 \end{bmatrix}.$$

Identifying this matrix with each of the generators of Γ_2 in turn then allows us to find the corresponding elements of SU(2) up to a choice of sign. This gives us the generators of Γ_6, and those of Γ_7 are obtained similarly from the generators of Γ_3 or equivalently by substituting $1-\omega$ for ω. This is essentially the method described by Hamermesh.[11] The remaining odd generators may be obtained by taking the tensor product of the basic spin representation with suitable even representations (thanks are due to H. Zassenhaus for this observation).

From Table II arises the first remarkable connection[12] between the icosahedral group and E_8. We first define the 9×9 matrix m_{jk} to be the multiplicity of the representation Γ_k in the Kronecker product of Γ_7 with Γ_j, i.e.,

$$\Gamma_7 \otimes \Gamma_j = \bigoplus_k m_{jk}\Gamma_k.$$

From this matrix a directed graph Λ of nine nodes is now constructed and with edges of multiplicity m_{jk} from the j th to the k th nodes. As usual the convention is adopted that two edges of opposite orientation and the same multiplicity between the same pair of nodes are replaced by an undirected edge. The result of this procedure is the following graph:



which is the Dynkin diagram of affine E_8 [if we also label the i th node by the dimension of the i th representation we obtain the marks of E_8 (see Ref. 13)]. This is a particular case of the following result of Ford and McKay.[12]

Proposition: Each of the five types of finite groups, the cyclic group (of order $r+1$), the dicyclic group (of order $4\{r-2\}$), the double tetrahedral group, the double octahedral group, and the double icosahedral group, have a two-dimensional representation such that corresponding graph Λ is the Dynkin diagram of the affine algebras A_r, D_r, E_6, E_7, and E_8.

The second relation between the icosahedral group and E_8, or more generally between a finite subgroup of SU(2) and the simple Lie algebras of types A, D, and E, can be described as follows. If G is a finite subgroup of SU(2) then G acts naturally on $\mathbb{C}^2$ and the orbifold $\mathbb{C}^2/G$ has a singularity at 0. These singularities are Kleinian singularities. The way in which these singularities are related to simple Lie algebra $\mathfrak{g}$ of types A, D, and E was recognized by Grothendieck *et al.*, and was worked out in detail by Slodowy.[14] The nilpotent cone $\mathfrak{B}$ in $\mathfrak{g}$ is an algebraic variety of dimension $\dim \mathfrak{g} - \text{rank}\, \mathfrak{g}$. Inside it there lies a subvariety $\mathfrak{B}'$ of codimension 2 that is formed by the elements called subregular (these actually form a single orbit under the action of the adjoint group). Any two-dimensional complex space in $\mathfrak{B}$ transverse to $\mathfrak{B}'$ in $\mathfrak{B}$ has a singularity at 0. This is precisely of the same form $\mathbb{C}^2/G$ for the appropriate G. One has $G = \text{II}$ in the case $\mathfrak{g} = E_8$.

III. THE TETRAHEDRAL SUBGROUP

The tetrahedral group T, being the largest subgroup of I, is hence a convenient tool in our work. Therefore, in this section we recall some of the relevant properties of the irreducible representations of T and the corresponding double point group TT.

We denote the irreducible representations of TT by $\gamma_1,\ldots,\gamma_7$. As in the icosahedral case these representations may be classed as either odd or even: $\gamma_1,\ldots,\gamma_4$ are the even representations and correspond to the irreducible representations of T, and $\gamma_5,\ldots,\gamma_7$ are the odd representations. The character table of TT is shown in Table IV; σ stands for $e^{2\pi i/3}$. Each TT-conjugacy class is also identified by the SU(2)-conjugacy class to which it belongs. Table V contains the decomposition of tensor products of irreducible representations of TT.

The nontrivial generators of TT are given in Table VI, the generators of the odd representations are obtained in the same manner as described above for II. Table VII contains the reduction (branching rules) of the irreducible representations of II to a direct sum of representations of TT. With the exception of Γ_9 these branchings may be obtained directly from the forms of the generators of II and TT in Tables III and VI.

TABLE IV. Character table of the double tetrahedral group TT; $\sigma = e^{2\pi i/3}$. The first three rows give the number of elements in each conjugacy class, the corresponding SU(2) conjugacy class and the orders of the elements in each class.

	C_1	C_1'	C_2	C_3	C_3'	C_4	C_4'
	1	1	6	4	4	4	4
	[10]	[01]	[11]	[12]	[21]	[12]	[21]
	1	2	4	3	6	3	6
γ_1	1	1	1	1	1	1	1
γ_2	1	1	1	σ	σ	σ^2	σ^2
γ_3	1	1	1	σ^2	σ^2	σ	σ
γ_4	3	3	-1	0	0	0	0
γ_5	2	-2	0	-1	1	-1	1
γ_6	2	-2	0	$-\sigma$	σ	$-\sigma^2$	σ^2
γ_7	2	-2	0	$-\sigma^2$	σ^2	$-\sigma$	σ

TABLE V. Decompositions of the tensor products of irreducible representations of TT.

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
γ_1	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7
γ_2		γ_3	γ_1	γ_4	γ_6	γ_7	γ_5
γ_3			γ_2	γ_4	γ_7	γ_5	γ_6
γ_4				$2(\gamma_4)\oplus\gamma_3\oplus\gamma_2\oplus\gamma_1$	$\gamma_5\oplus\gamma_6\oplus\gamma_7$	$\gamma_5\oplus\gamma_6\oplus\gamma_7$	$\gamma_5\oplus\gamma_6\oplus\gamma_7$
γ_5					$\gamma_4\oplus\gamma_1$	$\gamma_4\oplus\gamma_2$	$\gamma_4\oplus\gamma_3$
γ_6						$\gamma_4\oplus\gamma_3$	$\gamma_4\oplus\gamma_1$
γ_7							$\gamma_4\oplus\gamma_2$

TABLE VII. Reduction (branching rules) of the irreducible representations of II to a direct sum of irreducible representations of TT.

Γ_1	↓	γ_1	Γ_6	↓	γ_5
Γ_2	↓	γ_4	Γ_7	↓	γ_5
Γ_3	↓	γ_4	Γ_8	↓	$\gamma_6\oplus\gamma_7$
Γ_4	↓	$\gamma_4\oplus\gamma_1$	Γ_9	↓	$\gamma_5\oplus\gamma_6\oplus\gamma_7$
Γ_5	↓	$\gamma_4\oplus\gamma_3\oplus\gamma_2$			

IV. STRUCTURE OF THE RING OF POLYNOMIAL INVARIANTS

For any group G we may introduce a set of variables $x,y,z,\ldots$ that carry a representation Γ_m of G. The set of all polynomials in $x,y,z,\ldots$ with complex coefficients form a ring $\mathbf{R}[x,y,z,\ldots]$ that is naturally graded by degree. The polynomials of degree k also carry a representation of G which may be identified with the fully symmetrized component of the tensor product $\{\Gamma_m\}^k$. These polynomial tensors may be decomposed into the direct sum of irreducible representations of G and the number of times the irreducible representation Γ_r appears in this decomposition is given by the coefficient of λ^k in the Taylor series expansion (known in the mathematical literature as the Poincaré series) of the following generating function[7,15]:

$$B(\Gamma_r,\Gamma_m,\lambda) = \frac{1}{N}\sum_s \frac{N_s\chi_{rs}^*}{\det(1-\lambda A_s)}, \qquad (1)$$

where N is the order of G, N_s is the number of elements in the conjugacy class s, χ_{rs} the character of this class in the representation Γ_r (* denotes complex conjugation), and A_s is a matrix in Γ_m representing any element of the class s.

The set of all polynomials invariant under the action of G form a subring $\mathbf{J}$ of $\mathbf{R}$ that is also graded by degree. This is known as the ring of invariant polynomials of Γ_m. The generating function (also known as the Molien series) for the number of linearly independent invariant polynomials of degree k in $\mathbf{J}$ is given by $B(\Gamma_1,\Gamma_m,\lambda)$, Γ_1 being the identity representation of G. These generating functions for all the finite subgroups of SU(2) have been computed in Refs. 7 and 16. In the cases T and I they take the forms

TABLE VI. Generators of TT for each irreducible representation. Here A_1 and A_2 correspond to the two permutations (123) and (12)(34) under the identification of TT with A_4 and $\sigma = e^{2\pi i/3}$.

	A_1	A_2
γ_4	$\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}$	$\begin{bmatrix}1&0&0\\0&-1&0\\0&0&-1\end{bmatrix}$
γ_5	$\left(\frac{1}{2}\right)(1+i)\begin{bmatrix}1&i\\1&-i\end{bmatrix}$	$\begin{bmatrix}0&-i\\-i&0\end{bmatrix}$
γ_6	$\left(\frac{1}{2}\right)\sigma^2(1+i)\begin{bmatrix}1&i\\1&-i\end{bmatrix}$	$\begin{bmatrix}0&-i\\-i&0\end{bmatrix}$
γ_7	$\left(\frac{1}{2}\right)\sigma(1+i)\begin{bmatrix}1&i\\1&-i\end{bmatrix}$	$\begin{bmatrix}0&-i\\-i&0\end{bmatrix}$

$$\gamma_1:\ 1/(1-\lambda), \qquad (2a)$$

$$\gamma_2,\gamma_3:\ 1/(1-\lambda^3), \qquad (2b)$$

$$\gamma_4:\ (1+\lambda^6)/(1-\lambda^2)(1-\lambda^3)(1-\lambda^4), \qquad (2c)$$

and

$$\Gamma_1:\ 1/(1-\lambda), \qquad (3a)$$

$$\Gamma_2,\Gamma_3:\ \frac{1+\lambda^{15}}{(1-\lambda^2)(1-\lambda^6)(1-\lambda^{10})}, \qquad (3b)$$

$$\Gamma_4:\ \frac{1+\lambda^{10}}{(1-\lambda^2)(1-\lambda^3)(1-\lambda^4)(1-\lambda^5)}, \qquad (3c)$$

$$\Gamma_5:\ \frac{1+\lambda^5+2\lambda^6+\lambda^7+\lambda^{12}}{(1-\lambda^2)(1-\lambda^3)^2(1-\lambda^4)(1-\lambda^5)}. \qquad (3d)$$

We now make the following definition (see Ref. 17 for a recent review).

Definition: A set $\mathbf{B} = \{i_1,\ldots,i_l\}$, where $i_s \in \mathbf{J}$, $1 \leqslant s \leqslant l$, is called an integrity (or polynomial) basis of $\mathbf{J}$ if every element of $\mathbf{J}$ may be written as a polynomial in this set.

In general this expression for a general invariant in terms of an integrity basis is not unique. This nonuniqueness is measured by syzygies which are defined as follows.

Definition: Given an integrity basis $\mathbf{B} = \{i_1,\ldots,i_l\}$ a nonzero polynomial p such that $p(i_1,\ldots,i_l) = 0$ is called a syzygy.

Note that two polynomial expressions for an invariant differ by a syzygy.

In the case of representations of finite groups it is always possible to choose an integrity basis $\mathbf{B}$ such that the form of a general invariant and the independent syzygies have a particularly nice form.

Definition: As above consider a representation Γ_m of a finite group G, and let Γ_m have dimension k. Then a good integrity basis $\mathbf{B}$ is one such that $\mathbf{B} = \mathbf{B}_f \amalg \mathbf{B}_c$ (disjoint union), where $\mathbf{B}_f = \{I_1,\ldots,I_k\}$ are called free invariants and $\mathbf{B}_c = \{E_1,\ldots,E_{l-k}\}$ are called constrained invariants and we have the decomposition

$$\mathbf{J} = \mathbb{C}[I_1,\ldots,I_k] \oplus E_1\mathbb{C}[I_1,\ldots,I_k] \oplus \ldots \oplus E_{l-k}\mathbb{C}[I_1,\ldots,I_k].$$

In other words, every $i \in \mathbf{J}$ may be written in the form

$$i = p_0 + E_1 p_1 + \ldots + E_{l-k} p_{l-k}, \qquad (*)$$

where the p_j, $j = 0,\ldots,l-k$, are polynomials in the free invariants $I_1,\ldots,I_k$.

Since each product $E_r E_s$, $1 \leqslant r \leqslant s \leqslant l-k$, is an element of $\mathbf{J}$, from the above definition we have

$$E_r E_s - (q_0 + E_1 q_1 + \ldots + E_{l-k} q_{l-k}) = 0,$$

for some set of polynomials q_j, $1\leqslant j\leqslant l-k$, that will depend on r and s. It can be shown that any other syzygy may be built up from this basic set.

There is usually a certain amount of choice as to the elements of **B**; in most cases a possible choice of their number, type, and degree can, however, be found from the relevant generating function. To each factor $(1-\lambda^k)$ in the denominator of the generating function we associate an element I_i of degree k and to each term $c\lambda^k$ in the numerator we associate c linearly independent elements E_i of degree k. This information is of great help in suggesting how to search for syzygies and invariants. In particular it allows us to know when we have found the whole of **B**.

As an example consider the representation γ_4. From the denominator of the corresponding generating function we deduce that there are three free members of **B** of degrees 2, 3, and 4, which may be chosen to be[7]

$$\{2\} = x^2+y^2+z^2, \tag{4a}$$

$$\{3\} = xyz, \tag{4b}$$

$$\{4\} = x^4+y^4+z^4. \tag{4c}$$

The first two of which are unique, while to $\{4\}$ we may add any multiple of the square of the second-order invariant. For simplicity we write $\{a^n b^m\}$ for $\{a\}^n\{b\}^m$, etc., so that the square of the second-order invariant is written as $\{2^2\}$. The sixth-order invariant may be chosen to be[7]

$$\{6\} = (x^2-y^2)(y^2-z^2)(z^2-x^2). \tag{4d}$$

Once again this choice is not unique since we may add multiples of $\{2^3\}$, $\{3^2\}$, and $\{42\}$. Since the numerator of (2c) contains only λ^6 we deduce that $\{6^2\}$ must be expressible in terms of the other elements of **B**. This syzygy takes the form

$$\begin{aligned}&-2\{4^3\}+5\{4^2 2^2\}-4\{42^4\}+36\{43^2 2\}\\&\quad+\{2^6\}-20\{3^2 2^3\}+108\{3^4\} = -4\{6^2\}.\end{aligned} \tag{5}$$

In this example we have just a single constrained invariant and hence the situation is quite simple. In practice, however, there will exist several constrained invariants and the situation will be more complex. To illustrate the points that may arise, suppose we have three invariants a, b, and c that satisfy the syzygy (we assume that a, b, and c are going to be in our integrity basis)

$$a^n+b^n+c^n=0.$$

Now consider a general polynomial in a, b, and c. This is an invariant, but clearly in general its expression in terms of a polynomial of a, b, and c will not be unique. For example, we may eliminate any one of a^n, b^n, or c^n in favor of the other two variables. Suppose we choose a^n, then we have expressed an arbitrary invariant i in the form

$$i=p_0(b,c)+p_1(b,c)a+\cdots+p_{n-1}(b,c)a^{n-1}.$$

Choosing $\mathbf{B}_f=\{b,c\}$ and $\mathbf{B}_c=\{a,a^2,...,a^{n-1}\}$ (assuming that there are no other nontrivial syzygies satisfied by a, b, and c) defines a good integrity basis. This raises two points: first we could have chosen $\mathbf{B}_f=\{a,b\}$ and $\mathbf{B}_c=\{c,c^2,...,c^{n-1}\}$ or $\mathbf{B}_f=\{a,c\}$ and $\mathbf{B}_c=\{b,b^2,...,b^{n-1}\}$ as good integrity bases. In this sense our original nontrivial syzygy $a^n+b^n+c^n=0$ is more fundamental than our choice of basis. For this reason in our results we have given a set of independent nontrivial syzygies and a possible choice of invariants (which is, of course, far from unique). The second point is that certain trivial syzygies arise once one has made a choice of **B**. The reader may have noticed that it was claimed above that a complete set of independent syzygies was given in the case of a good integrity basis by the products E_rE_s of the constrained invariants, whereas in this example we started with just a single syzygy, but obtain $n-1$ constrained invariants and consequently their products. The solution to this apparent problem is of course that the constrained invariants are related by being powers of a single invariant. To be more precise if $r+s=xn+y$ $(1\leqslant y\leqslant n-1)$ then the product of $E_r=a^r$ and $E_s=a^s$ is simply

$$\begin{aligned}E_rE_s&=a^ra^s=a^{r+s}=(a^n)^xa^y\\&=(-b^n-c^n)^xa^y=(-I_1^n-I_2^n)^xE_y.\end{aligned}$$

These trivial syzygies may be determined easily after one has made a choice of integrity basis and we shall not discuss them further.

By using the result[7]

$$B(\Gamma_1,\Gamma_r\oplus\Gamma_s;\lambda_1\lambda_2)=\sum_t B(\Gamma_t,\Gamma_r;\lambda_1)B(\Gamma_t^*,\Gamma_s;\lambda_2), \tag{6}$$

the Molien series for any reducible representation may be calculated, and hence the integrity basis found. The integrity bases for the representations of T that occur in the branchings from the irreducible representations of I are shown in Table VIII. In this table we adopt the convention that the representation γ_2 acts on the indeterminate p, γ_3 acts on q, and γ_4 acts on the three indeterminates x, y, and z. This convention is adhered to in the case of the direct sum $\gamma_2\oplus\gamma_3\oplus\gamma_4$ and hence the invariants of this representation are polynomials in p, q, x, y, and z. In the next section we express the icosahedral invariants in terms of these tetrahedral invariants.

V. METHODS AND RESULTS

In principle, the icosahedral invariants we wish to calculate may be found directly by using the following procedure.

For any particular degree first construct the most general homogeneous polynomial $P(x,y,z,...)$ in the variables $x,y,z,...$. This polynomial will in general contain a large number of arbitrary parameters c_i. For it to be invariant under icosahedral transformations we require that, for each generating element A of I,

$$P(x,y,z,...)=P(D(A)x,D(A)y,D(A)z,...), \tag{7}$$

$D(A)$ being the matrix representing A in Γ_m, which acts on $x,y,z,...$ as a linear transformation. Equating the coefficients

TABLE VIII. Integrity bases for the tetrahedral group representations occurring in the branching of irreducible icosahedral group representations.

Representation	Free	Constrained
γ_2	$\{3_1\} = p^3$	none
γ_3	$\{\bar{3}_1\} = q^3$	none
γ_4	$\{2_0\} = x^2 + y^2 + z^2$	$\{6_0\} = (x^2 - y^2)(y^2 - z^2)(z^2 - x^2)$
	$\{3_0\} = xyz$	
	$\{4_0\} = x^4 + y^4 + z^4$	
$\gamma_2 \oplus \gamma_3 \oplus \gamma_4$	$\{2_0\}$	$\{2_1\} = pq$
	$\{3_1\}$	$\{3_2\} = (x^2 + \sigma y^2 + \sigma^2 z^2)q$
	$\{\bar{3}_1\}$	$\{\bar{3}_2\} = (x^2 + \sigma^2 y^2 + \sigma z^2)p$
	$\{3_0\}$	$\{4_1\} = (x^2 + \sigma y^2 + \sigma^2 z^2)p^2$
	$\{4_0\}$	$\{\bar{4}_1\} = (x^2 + \sigma^2 y^2 + \sigma z^2)q^2$
		$\{2_1^2\}$
		$\{5_1\} = (x^4 + \sigma y^4 + \sigma^2 z^4)q$
		$\{\bar{5}_1\} = (x^4 + \sigma^2 y^4 + \sigma z^4)p$
		$\{3_2 2_1\},\{\bar{3}_2 2_1\}$
		$\{6_0\}$
		$\{3_2^2\},\{\bar{3}_2^2\}$
		$\{5_1 2_1\},\{\bar{5}_1 2_1\}$
		$\{6_0 2_1\},\{6_0 2_1^2\}$

of each monomial in the variables $x,y,z,...$ yields a large number of homogeneous linear equations for the c_i. The solution of this set then yields the most general invariant of degree r.

In many cases this calculation is not difficult to perform. For the present case, however, the computations involved would be formidable. Fortunately a number of steps may be taken to simplify the problem. The first of these is to make use of the tetrahedral invariants of Sec. IV.

Since $\mathrm{T} \subset \mathrm{I}$, any polynomial invariant under the icosahedral group must also be invariant under the tetrahedral subgroup. Hence each icosahedral invariant may be written as a sum of tetrahedral invariants of the same degree. The most general sum of tetrahedral invariants of any degree may be found from Table VIII for each irreducible representation of I. Furthermore, since two of the generators of I generate the tetrahedral subgroup we only have to impose invariance under the third generator to ensure invariance for the whole of I. The number of arbitrary coefficients is further reduced by calculating first the invariants of lowest degree and excluding products of these invariants when finding those of higher degree.

Finally we note that it is usually inefficient to expand the whole of Eq. (7) in order to generate the required linear system as this leads to a large number of dependent equations. A better procedure is to specialize some or all of the variables $x,y,z,...$ to particular values (usually roots of unity), thus generating a smaller number of hopefully independent equations. This procedure can be systematized[18]; however, the icosahedral case is sufficiently simple that a few trial choices lead more quickly to the desired set of independent linear equations for the coefficients.

Although these steps reduce to a large extent the amount of work involved in finding the required invariants, the result is still far from a hand calculation. We have thus made use of the computer algebra system MACSYMA. The results of this calculation are given in Tables IX, X, and XI. In these tables the symbol $[n]$ denotes an nth degree icosahedral invariant (if necessary additional labels are used to distinguish invariants of the same order) and the tetrahedral invariants are as in Table VIII. As above the notation $[n^a m^b]$ is used for $[n]^a[m]^b$. In the case of Γ_4 (Tables X and XIII) $\{1\}$ is used to denote an indeterminate, t say, that is left invariant by the action of the tetrahedral subgroup. This corresponds to the fact that the identity representation of T occurs in the branching rule of Γ_4 in Table VII.

When classifying the invariants of Γ_5 we have found it useful to introduce "symmetric" and "antisymmetric" invariants. The former are denoted with a subscript s and the latter with a subscript a. They differ in their transformation properties under the $\mathbb{Z}_2$ transformation given by

$$p \to q, \quad q \to p, \quad \sigma \to \sigma^2, \quad \sigma^2 \to \sigma.$$

The symmetric invariants are left unchanged by this transformation, while the antisymmetric representations are multiplied by -1.

The syzygies satisfied by these invariants (for Γ_5 those syzygies of degree $\leqslant 14$) were found in a similar manner by equating the most general polynomial in the elements of **B** of some particular degree to zero and then solving for the unknown coefficients. The results of this calculation are given in Tables XII, XIII, and XIV.

As was noted above the distinction between free and constrained invariants is to a certain extent arbitrary, the more important thing being the nontrivial syzygies. For convenience, however, we give here sets of constrained invariants consistent with the syzygies of Tables XII, XIII, and XIV:

TABLE IX. The integrity basis for the representation Γ_2; that of Γ_3 is obtained by substituting $1 - \omega$ for ω.

$$[2] = \{2\}$$
$$[6] = (4\omega - 2)\{6\} + 22\{3^2\} + \{42\}$$
$$[10] = 3[6]\{4\} - 8\{4^2 2\} + 9\{42^3\} - 256\{43^2\} + 128\{3^2 2^2\}$$
$$[15] = [6](15\{432\} + 290\{3^3\} - 11\{32^3\}) - 225\{4^3 3\} + 425\{4^2 3 2^2\} - 80\{43^3 2\} - 270\{432^4\} - 9728\{3^5\} + 54\{3^3 2^3\} + 58\{32^6\}$$

TABLE X. The syzygy satisfied by the invariant [15] of the representation Γ_2 and Γ_3.

$$
\begin{aligned}
&80[15^2] + 50[10^3] - 550[10^2 6 2^2] - 66[10^2 2^5] + 450[10 6^3 2] \\
&\quad + 360[10 6^2 2^4] + 2458[10 6 2^7] - 740[10 2^{10}] - 135[6^5] \\
&\quad - 215[6^4 2^3] - 1200[6^3 2^6] - 776[6^2 2^9] - 2625[6 2^{12}] \\
&\quad + 1495[2^{15}] = 0
\end{aligned}
$$

$$
\begin{aligned}
&\Gamma_2, \Gamma_3 \quad [15], \\
&\Gamma_4 \quad [10], \\
&\Gamma_5 \quad [5_s],\ [6_s],\ [6_a],\ [7_s],\ [7_s 5_s].
\end{aligned}
$$

In the first two cases it is clear that these particular choices are fairly natural; for the representation Γ_5, however, the situation is less clear-cut. In particular it should be noted that the 12th-order invariant is somewhat special. For this case there are no new invariants of this order, but the four products $[7_s 5_s]$, $[6_s 6_a]$, $[6_s^2]$, and $[6_a^2]$ are related by three independent syzygies that may be used to eliminate any three. We may thus choose any of them as the 12th-order invariant. Of course, this procedure also depends on the previous choice of lower-order constained invariants.

VI. OTHER CONSTRUCTIONS

In this section we would like to describe some other methods of calculating the integrity bases of I that do not make use of the tetrahedral subgroup.

The invariants of the three-dimensional representations Γ_2 and Γ_3 are long well known.[2,19] They are constructed by making use of the fact that it is via these two representations that the icosahedral group acts on the icosahedron. For more details the reader should consult Table III and Appendix D of Ref. 19.

The integrity basis of Γ_4 may be found by making use of the isomorphorism $I \cong A_5 \subset S_5$ and the theory of symmetric polynomials. Quite generally any symmetric group S_n has a permutation representation P_n that may be realized by the permutations of a set of indeterminates $x_1, x_2, \ldots, x_n$. This representation is the direct sum of an irreducible $(n-1)$-dimensional representation and a one-dimensional identity representation. Labeling these representations by partitions in the usual way we have

$$
P_n \cong (n) \oplus (n-1,1). \tag{8}
$$

TABLE XI. The integrity basis for the representation Γ_4.

$$
\begin{aligned}
[2] &= \{2\} + \{1^2\} \\
[3] &= -2\sqrt{5}\{3\} - [2]\{1\} + 2\{1^3\} \\
[4] &= 5\{4\} + 18[2]\{1^2\} + 6[3]\{1\} - 23\{1^4\} \\
[5] &= 10[4]\{1\} - 80[3]\{1^2\} - 25[2^2]\{1\} - 160[2]\{1^3\} + 256\{1^5\} \\
[10] &= 2[4]\{6\} - 5[2^2]\{6\} - 32\{61\}[3] - 96\{61^2\}[2] + 256\{61^4\}
\end{aligned}
$$

TABLE XII. The syzygy satisfied by the invariant [10] of the presentation Γ_4.

$$
\begin{aligned}
&-[5^4] - 60[5^3 3 2] + 40[5^2 4^2 2] - 180[5^2 4 3^2] - 380[5^2 4 2^3] \\
&\quad - 210[5^2 3^2 2^2] + 916[5^2 2^5] + 160[5 4^3 3] - 1760[5 4^2 3 2^2] \\
&\quad - 2520[5 4 3^3 2] + 6520[5 4 3 2^4] - 1728[5 3^5] + 6940[5 3^3 2^3] \\
&\quad - 7800[5 3 2^6] - 32[4^5] + 560[4^4 2^2] + 720[4^3 3^2 2] - 3800[4^3 2^4] \\
&\quad + 540[4^2 3^4] - 5600[4^2 3^2 2^3] + 12\,500[4^2 2^6] - 2700[4 3^4 2^2] \\
&\quad + 14\,500[4 3^2 2^5] - 20\,000[4 2^8] + 3375[3^4 2^4] - 12\,500[3^2 2^7] \\
&\quad + 12\,500[2^{10}] + 2000[10^2] = 0
\end{aligned}
$$

In the case of S_5 the representation (4,1) under restriction to I yields the single irreducible representation Γ_4. Polynomials in the x_i's that are invariant under permutations are called symmetric polynomials and it is well known that the integrity basis consists of n free invariants of degrees 1 to n. Convenient choices are the elementary symmetric polynomials $e_1, \ldots, e_n$, the complete symmetric polynomials $h_1, \ldots, h_n$, or the power sum symmetric polynomials $p_1, \ldots, p_n$.[20] There is also a relative invariant given by Vandermonde's determinant

$$
D_n = \det(x_i^{n-j}) = \prod_{i<j} (x_i - x_j)
$$

of degree $\frac{1}{2}n(n+1)$. Under restriction to A_n, D_n may be considered to be a constrained invariant whose square may be expressed as a polynomial in the other elements of the integrity basis.

Thus for the representation $\Gamma_4 \oplus \Gamma_1$ of I we may choose, for example, $p_1, \ldots, p_5$ and D_5 as an integrity basis. Clearly the vector (1,1,1,1,1) generates the one-dimensional invariant subspace. On restriction to the orthogonal subspace the invariant $p_1 = x_1 + \cdots + x_5$ vanishes and the other invariants provide the required integrity basis of Γ_4.

We may obtain the integrity basis of Γ_5 by considering another permutation representation, this time of S_6. This is

TABLE XIII. The integrity basis for the representation Γ_5.

$$
\begin{aligned}
[2_s] &= \{2_0\} + 2\{2_1\} \\
[3_a] &= \{\bar{3}_1 - 3_1\} + 3\{\bar{3}_2 - 3_2\} \\
[3_s] &= \{\bar{3}_1 + 3_1\} + \{\bar{3}_2 + 3_2\} - 4\{3_0\} \\
[4_s] &= \{4_0\} - 2\{\bar{4}_1 + 4_1\} - 6\{2_1^2\} + 4[2_s]\{2_1\} \\
[5_a] &= \{\bar{5}_1 - 5_1\} - 2[2_s]\{\bar{3}_2 - 3_2\} - [3_a]\{2_1\} + 5\{\bar{3}_2 2_1 - 3_2 2_1\} \\
[5_s] &= \{\bar{5}_1 + 5_1\} - 2[2_s]\{\bar{3}_2 + 3_2\} - [3_s]\{2_1\} + 7\{\bar{3}_2 2_1 + 3_2 2_1\} \\
&\quad + 12\{3_0 2_1\} \\
[6_s] &= -21[4_s]\{2_1\} + 8[2_s]\{\bar{4}_1 + 4_1\} - 32[2_s]\{2_1^2\} + 23[2_s^2]\{2_1\} \\
&\quad + 28[3_s]\{3_0\} - 48[3_s]\{\bar{3}_2 + 3_2\} - 192\{\bar{3}_2 3_0 + 3_2 3_0\} \\
&\quad - 16\{3_0^2\} + 4[3_a]\{\bar{3}_2 - 3_2\} + 80\{\bar{3}_2^2 + 3_2^2\} \\
[6_a] &= 6(\sigma - \sigma^2)\{6_0\} - 23[3_s]\{\bar{3}_2 - 3_2\} + 72[3_a]\{3_0\} \\
&\quad - 284\{\bar{3}_2 3_0 - 3_2 3_0\} - 17[3_a]\{\bar{3}_2 + 3_2\} \\
&\quad + 70\{\bar{3}_2^2 - 3_2^2\} + 8[2_s]\{\bar{4}_1 - 4_1\} \\
[7_s] &= 7[2_s^2]\{3_0\} - 4[3_s 2_s]\{2_1\} - 32[2_s]\{3_0 2_1\} \\
&\quad - 2[5_s]\{2_1\} + 4[3_s]\{2_1^2\} \\
&\quad + 64\{3_0 2_1^2\} - 2[3_s]\{\bar{4}_1 + 4_1\} - 16\{\bar{4}_1 3_0 + 4_1 3_0\} - 5[4_s]\{3_0\}
\end{aligned}
$$

TABLE XIV. The syzygies satisfied by the constained elements of the integrity basis of Γ_5.

$$36[7_s3_s] + 5[6_s4_s] - 7[6_s2_s^2] - 375[5_a^2] - 9[5_s^2] - 240[5_a3_a2_s] - 5[4_s3_s^2] - 31[4_s3_a^2] + 5[3_a^2 2_s^2] + 7[3_s^2 2_s^2] = 0$$

$$- 480[7_s4_s] + 672[7_s2_s^2] - 10[6_s5_s] - 200[6_a5_a] - 128[6_s3_s2_s] - 64[6_a3_a2_s] - 456[5_s4_s2_s] - 755[5_s3_s^2] + 35[5_s3_a^2] + 720[5_a3_a3_s] + 552[5_s2_s^3] + 600[4_s^2 3_s] - 1680[4_s3_s2_s^2] - 1024[3_s^3 2_s] + 1024[3_s3_a^2 2_s] + 1176[3_s2_s^4] = 0$$

$$- 24[6_a5_s] + 50[6_s5_a] + 16[6_s3_a2_s] + 360[5_a4_s2_s] + 319[5_a3_s^2] - 175[5_a3_a^2] - 144[5_s3_s3_a] - 72[5_a2_s^2] + 24[4_s^2 3_a] + 48[4_s3_a2_s^2] - 56[3_a^3 2_s] + 56[3_a3_s^2 2_s] + 24[3_a2_s^4] = 0$$

$$576[7_s5_s] + 9216[7_s3_s2_s] - 64[6_a^2] + 976[6_s4_s2_s] - 450[6_s3_s^2] - 18[6_s3_a^2] - 144[6_a3_a3_s] - 1424[6_s2_s^3] - 19\,200[5_a^2 2_s] + 5328[5_s4_s3_s] - 15\,120[5_a4_s3_a] - 1008[5_s3_s2_s^2] + 8880[5_a3_a2_s^2] - 3456[4_s^3] + 8640[4_s^2 2_s^2] + 176[4_s3_s^2 2_s] - 11\,840[4_s3_a^2 2_s] - 6912[4_s2_s^4] + 63[3_a^4] - 2871[3_s^4] + 2808[3_a^2 3_s^2] + 14\,656[3_a^2 2_s^3] + 2576[3_s^2 2_s^3] + 1728[2_s^6] = 0$$

$$- 4320[7_s5_s] - 13\,824[7_s3_s2_s] - 50[6_s^2] - 2280[6_s4_s2_s] - 719[6_s3_s^2] + 485[6_s3_a^2] - 72[6_a3_a3_s] + 2760[6_s2_s^3] + 144\,000[5_a^2 2_s] + 1512[5_s4_s3_s] - 1800[5_a4_s3_a] - 6264[5_s3_s2_s^2] + 94\,680[5_a3_a2_s^2] - 960[4_s3_s^2 2_s] + 13\,560[4_s3_a^2 2_s] - 1085[3_a^4] - 2552[3_s^4] + 3637[3_a^2 3_s^2] - 1560[3_a^2 2_s^3] - 2112[3_s^2 2_s^3] = 0$$

$$2304[7_s3_a2_s] + 5760[5_a5_s2_s] + 10\,440[5_a3_s2_s^2] + 1944[5_s3_a2_s^2] - 1080[4_s3_a3_s2_s] + 3672[3_a3_s2_s^3] + 7200[7_s5_a] + 40[6_a6_s] + 320[6_a3_s^2] - 248[6_a3_a^2] + 234[6_s3_a3_s] - 72[5_s4_s3_a] - 2520[5_a4_s3_s] - 1467[3_a^3 3_s] + 1467[3_s^3 3_a] = 0$$

$$- 40[7_s6_s] + 5856[7_s4_s2_s] - 3020[7_s3_s^2] + 140[7_s3_a^2] - 8544[7_s2_s^3] - 334[6_s4_s3_s] + 2106[6_s3_s2_s^2] + 2560[6_a5_a2_s] + 8[6_a4_s3_a] + 808[6_a3_a2_s^2] + 1920[5_s5_a3_a] + 432[5_s4_s^2] + 5088[5_s4_s2_s^2] + 8640[5_s3_s^2 2_s] + 960[5_s3_a^2 2_s] - 6864[5_s2_s^4] + 28\,800[5_a^2 3_s] + 9600[5_a3_a3_s2_s] - 7320[4_s^2 3_s2_s] + 703[4_s3_s^3] + 2177[4_s3_s3_a^2] + 20\,928[4_s3_s2_s^3] + 12\,123[3_s^3 2_s^2] - 13\,083[3_s3_a^2 2_s^2] + 14\,952[3_s2_s^5] = 0$$

$$96[7_s6_a] + 576[7_s3_s3_a] + 640[6_s5_a2_s] + 2[6_s4_s3_a] + 202[6_s3_a2_s^2] - 120[6_a4_s3_s] + 168[6_a3_s2_s^2] + 3456[5_s5_a3_s] + 1152[5_s3_a3_s2_s] - 9600[5_a^2 3_a] - 2160[5_a4_s^2] + 3744[5_a4_s2_s^2] + 5120[5_a3_s^2 2_s] - 10\,112[5_a3_a^2 2_s] - 1008[5_a2_s^4] - 600[4_s^2 3_a2_s] - 569[4_s3_s^2 3_a] - 7[4_s3_a^3] + 960[4_s3_a2_s^3] + 2435[3_s^2 3_a2_s^2] - 2243[3_a^3 2_s^2] - 168[3_a2_s^5] = 0$$

$$144[7_s^2] + 576[7_s5_s2_s] + 1152[7_s4_s3_s] + 80[6_s5_a3_a] + 18[6_s4_s^2] + 212[6_s4_s2_s^2] + 360[6_s3_s^2 2_s] + 40[6_s3_a^2 2_s] - 286[6_s2_s^4] + 576[6_a5_a3_s] + 192[6_a3_s3_a2_s] + 1008[5_s4_s3_s2_s] + 1944[5_s3_s^3] + 72[5_s3_s3_a^2] - 720[5_s3_s2_s^3] - 15\,360[5_a^2 2_s^2] + 144[5_a4_s3_a2_s] - 1520[5_a3_s^2 3_a] - 496[5_a3_a^3] - 10\,032[5_a3_a2_s^3] - 1521[4_s^2 3_s^2] - 111[4_s^2 3_a^2] + 4342[4_s3_s^2 2_s^2] - 1270[4_s3_a^2 2_s^2] + 2880[3_s^4 2_s] - 2632[3_s^2 3_a^2 2_s] - 2729[3_s^2 2_s^4] - 248[3_a^4 2_s] + 137[3_a^2 2_s^4] = 0$$

TABLE XV. Alternative conventions for TT irreducible representations. We include the partition notation which allows contact to be made with the standard notations of the representation theory of the symmetric group[21] and the notations used by the Atlas of finite groups.[22]

γ_1	[4]	γ_5	[4]′
γ_2	$[2^2]_+$	γ_6	$[31]'_+$
γ_3	$[2^2]_-$	γ_7	$[31]'_-$
γ_4	[31]		

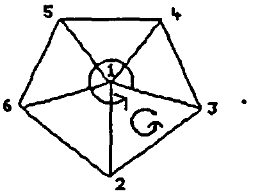
TABLE XVI. Alternative conventions for II irreducible representations. Notation as in Table XV.

Γ_1	[5]	1*a*	Γ_6	$[5]'_+$	2*a*
Γ_2	$[31^2]_+$	3*a*	Γ_7	$[5]'_-$	2*b*
Γ_3	$[31^2]_-$	3*b*	Γ_8	[32]′	4*b*
Γ_4	[41]	4*a*	Γ_9	[41]′	6*a*
Γ_5	[32]	5*a*			

TABLE XVII. Alternative conventions for TT conjugacy classes. Notation as in Table XV.

C_1	(1^4)
C_2	(2^2)
C_3	$(31)_+$
C_4	$(31)_-$

possible because of the fact that there exists a nonstandard embedding of A_5 in S_6 such that the restriction of the representation (5,1) is Γ_5. To obtain the generators of A_5 in this representation consider the action of the icosahedral group on an icosahedron. Joining each opposite pair of vertices of this icosahedron we obtain six lines permuted by the action of the icosahedral group. Consider in plan view the situation is as follows:



The two arrows represent the action of two elements of the icosahedral group, one of order 5 and one of order 3, which hence generate the whole of I. Their effects as permutations on the lines joining the vertices (numbered 1 to 6) are (23456) and (123)(465).

We may rewrite the form of the generating function for Γ_5 in the form

$$\frac{(1+x^3)(1+x^5+2x^6+x^7+x^{12})}{(1-x^2)(1-x^3)(1-x^4)(1-x^5)(1-x^6)}.$$

Thus as for Γ_4 we may take the free intervals as $p_2,\dots,p_6$. The

TABLE XVIII. Alternative conventions for II conjugacy classes. Notation as in Table XV.

C_1	(1^5)	1*A*
C_2	$(5)_+$	5*A*
C_3	$(5)_-$	5*B*
C_4	(2^21)	2*A*
C_5	(31^2)	3*A*

constrained invariants are, however, much more complicated. Although it is possible to find all invariants of the required degrees by demanding invariance under the two permutations (23456) and (123)(465) there does not seem to be a natural choice as is the case for the other representations.

To summarize, we have in this paper investigated the polynomial invariants of the icosahedral group. The integrity basis for each irreducible representation has been found by making use of the invariants of the tetrahedral subgroup T of I. We have also indicated how the geometry of the icosahedron and the use of permutation representations may also be used to construct the integrity bases in a different manner. In conclusion we note that a number of different conventions are available for labeling the irreducible representations and classes of II and TT. Some of these are summarized in Tables XV–XVIII.

ACKNOWLEDGMENTS

It is a pleasure to thank J. Mckay, R. V. Moody, L. Soicher, and H. Zassenhaus for useful and stimulating conversations. Thanks are also due to D. Rand for his help and advice concerning MACSYMA.

The work for this paper was supported in part by the National Science and Engineering Council of Canada and by the Ministère de l'Education du Québec.

[1]D. R. Nelson, "Quasicrystals," Sci. Am. **255**(8), 42 (1986).
[2]F. Klein, *Lectures on the Icosahedron and the Solutions of Equations of the Fifth Degree* (Dover, New York, 1956).
[3]D. Shechtman, I. A. Blech, D. Gratias, and J. W. Cahn, "Metallic phase with long-range orientational order and no translational symmetry," Phys. Rev. Lett. **53**, 1951 (1984).
[4]D. Shechtman and I. A. Blecht, "The microstructure of rapidly solidified Al_6Mn," Metall. Trans. A **16**, 1005 (1985).
[5]J. V. Hogle, M. Chow, and D. J. Filman, "The Structure of Poliovirus," Sci. Am. **256**(3), 42 (1987).
[6]S. V. Petuchov, *Biomechanics, Bionics and Symmetry* (Nauka, Moscow, 1981) (in Russian).
[7]J. Patera, R. T. Sharp, and P. Winternitz, "Polynomial irreducible tensors for point groups," J. Math. Phys. **19**, 2362 (1978).
[8]R. V. Moody, J. Patera, and R. T. Sharp, "Character generators for elements of finite order in simple Lie groups A_1,A_2,A_3,B_2, and G_2," J. Math. Phys. **24**, 2387 (1983).
[9]M. Puttaswamaiah and G. de B. Robinson, "Induced representations and alternating groups," Can. J. Math. **16**, 587 (1964).
[10]J. S. Lomont, *Applications of Finite Groups* (Academic, New York, 1959).
[11]M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Addison–Wesley, Reading, MA, 1962).
[12]D. Ford and J. McKay, "Representations and Coxeter graphs," in *The Geometric Vein. Festschrift for Coxeter*, edited by C. Davis, B. Grünbaum, and F. A. Sherk (Springer, New York, 1981), pp. 549–554; J. McKay, "Graphs, singularities and finite groups," in *Santa Cruz Conference in Finite Groups*, Am. Math. Soc. Pure Math. Symp. **37**, 183 (1980).
[13]R. Steinberg, "Finite subgroups of SU(2), Dynkin diagrams and affine Coxeter elements," Pac. J. Math. **118**, 587 (1985).
[14]P. Slodowy, "Simple singularities and simple algebraic groups," *Lecture Notes in Mathematics*, Vol. 815 (Springer, New York, 1980).
[15]W. Burnside, *Theory of Groups of Finite Order* (Dover, New York, 1955), 2nd ed.
[16]P. Desmier and R. T. Sharp, "Polynomial tensors for double point groups," J. Math. Phys. **20**, 74 (1979).
[17]N. J. A. Sloane, "Error correcting codes and invariant theory," Am. Math. Mon. **84**, 82 (1977).
[18]R. V. Moody and J. Patera, "Character decomposition of class functions on simple Lie groups," to be published in Math. Comput.
[19]M. V. Jaric, L. Michel, and R. T. Sharp, "Zeros of covariant vector fields for the point groups; invariant formulation," J. Phys. (Paris) **45**, 1 (1984).
[20]I. G. Macdonald, *Symmetric Functions and Hall Polynomials* (Clarendon, Oxford, 1979).
[21]L. Dehuai and B. G. Wybourne, "The alternating group: Branching rules, products and plethysms for ordinary and spin representations," J. Phys. A **14**, 1835 (1981).
[22]J. H. Conway, R. A. Wilson, R. T. Curtis, S. P. Norton, and R. P. Parker, *Atlas of Finite Groups* (Clarendon, Oxford, 1985).

SL(3,R) as the group of symmetry transformations for all one-dimensional linear systems. II. Realizations of the Lie algebra

M. Aguirre
Instituto de Física, Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

J. Krause
Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 6177, Santiago 22, Chile

(Received 26 January 1988; accepted for publication 30 March 1988)

The two-dimensional space-time realizations of the Lie algebra of SL(3,R) are obtained, when the group acts as the maximal point symmetry group of any given one-dimensional Newtonian linear system. It is shown that these realizations are isomorphic with the realization of the Lie algebra of the projective group in the plane. Next an active point of view is introduced, and SL(3,R) is interpreted as a group of mappings that transform one admissible world line of the system into another. Thus a new mechanical realization of the sl(3,R) algebra comes to the fore. Some miscellaneous examples are included.

I. INTRODUCTION

In a previous paper[1] we have solved the converse problem of similarity analysis[2,3] for finite point symmetry transformations of any inhomogeneous ordinary linear differential equation of the second order. In that paper, the eight-parameter realizations of the symmetry group were obtained in the form of conjugated diffeomorphisms of the form $\mathscr{F}^{-1}\mathscr{P}_2\mathscr{F}$, where $\mathscr{F}$ stands for some parameter-free transformations of the (t,x) variables (which depend exclusively on the fundamental solutions of the equation), and where $\mathscr{P}_2$ is an arbitrary projective transformation in the plane. In this fashion, without recourse to the Lie algebra, it was shown that the full point symmetry group of *all* such equations corresponds to SL(3,R) indeed. (This fact seems to be not so well known to most physicists,[4] although it figures in the literature[3] and, moreover, it was a fact well known to Lie himself.[5])

Notwithstanding this fact, we wish to remark that the main interest of our previous work on this issue[1,6] stems from its usefulness in classical mechanics as well as in quantum kinematics.[7] In effect, we have obtained in Ref. 1 a *technique* for *calculating the specific finite realizations* of SL(3,R) for any given one-dimensional linear Newtonian system, in terms of a set of basic solutions of the equation of motion. (We have been unable to find this technique in the current literature.) Hence in this article, as an indispensable complement of our previous work, we tackle the problem of finding the physically meaningful *realization of the Lie algebra* of SL(3,R) [i.e., sl(3,R)] for a given linear system in two-dimensional space-time.[8] Since the motivation underlying this paper is the same already formulated in Ref. 1, we would like to refer the reader to the Introduction of that paper.

The organization of this paper is as follows. In Sec. II we obtain the desired space-time realizations of sl(3,R), and we compare them with the realization of the Lie algebra of the projective group within the present formalism. Next, in Sec. III, we adopt the active point of view and interpret SL(3,R) as a group of mappings that transform one admissible world line of the system into another. Thus a new mechanical realization of SL(3,R) comes to the fore, which takes place in the classical *state-space* of the system. Then we get the general realization of the associated sl(3,R) algebra in this space. Section IV contains some miscellaneous examples of our technique.

II. SPACE-TIME REALIZATIONS OF sl(3,R)

In Ref. 1 it was shown that the full point symmetry group of the general second-order linear differential equation

$$L(t)x \equiv \ddot{x} + f_2(t)\dot{x} + f_1(t)x = f_0(t) \tag{2.1}$$

becomes realized by the following local diffeomorphisms of the space $\{t,x\}$:

$$t' = \tau^{-1}\left(\frac{q^1u_1(t) + q^3u_2(t) + q^2(x-u_P(t))}{q^7u_1(t) + q^8(x-u_P(t)) + u_2(t)}\right),$$
$$x' = \frac{q^4u_1(t) + q^6u_2(t) + q^5(x-u_P(t))}{q^7u_1(t) + q^8(x-u_P(t)) + u_2(t)} \times u_2\left[\tau^{-1}\left(\frac{q^1u_1(t) + q^3u_2(t) + q^2(x-u_P(t))}{q^7u_1(t) + q^8(x-u_P(t)) + u_2(t)}\right)\right] + u_P\left[\tau^{-1}\left(\frac{q^1u_1(t) + q^3u_2(t) + q^2(x-u_P(t))}{q^7u_1(t) + q^8(x-u_P(t)) + u_2(t)}\right)\right]. \tag{2.2}$$

Here $u_1(t)$ and $u_2(t)$ are two linearly independent solutions of the corresponding homogeneous equation, $L(t)x = 0$, and $u_P(t)$ is a particular solution of Eq. (2.1). The function τ^{-1} denotes the inverse function of $\tau(t) = u_1(t)/u_2(t)$, and the q's are the eight parameters of SL(3,R). (For details, see Ref. 1.) Indeed, in Ref. 1 it was proved that this rather formidable scheme of transformations entails a local, albeit finite, realization of SL(3,R) over the configuration space-time arena of the system whose equation of motion is (2.1). In order to handle these transformations in an easy manner, we had better introduce a change of variables, say

$$\hat{t} = u_1(t)/u_2(t), \quad \hat{x} = (x-u_P(t))/u_2(t). \tag{2.3}$$

(This is, in fact, an $\mathscr{F}$ *transformation*, according to the terminology used in Ref. 1.) Hence upon substitution of (2.3)

 0022-2488/88/081746-07$02.50

into (2.2) these diffeomorphisms read, briefly,

$$t' = \dot{\hat{t}}^{-1}\left(\frac{q^1\hat{t}(t) + q^2\hat{x}(t,x) + q^3}{1 + q^7\hat{t}(t) + q^8\hat{x}(t,x)}\right),$$
$$x' = \frac{q^4\hat{t}(t) + q^5\hat{x}(t,x) + q^6}{1 + q^7\hat{t}(t) + q^8\hat{x}(t,x)}u_2(t') + u_P(t'). \qquad (2.4)$$

[By the way, in this fashion it becomes apparent that Eq. (2.2) corresponds to the conjugations $\mathscr{F}^{-1}\mathscr{P}_2\mathscr{F}$ of the projective group $\mathscr{P}_2$ of the plane $\{t,x\}$ by the local, parameter-free diffeomorphisms $\mathscr{F}$ defined in Eq. (2.3). Cf. Theorem III in Ref. 1.]

We next discuss the general Lie algebra of the point symmetry group of Eq. (2.1) from the standpoint of this formalism. Clearly, in this endeavor one considers the monoparametric transformations of variables obtained from Eq. (2.2) to the first order of approximation in the parameter one handles. Now Lie's operators $Z_a(t,x)$, $a = 1,...,8$ (attached to the infinitesimal transformations in the present realization of the group), are given by

$$Z_a(t,x) = \eta_a(t,x)\partial_t + \theta_a(t,x)\partial_x, \qquad (2.5)$$

where one has

$$\eta_a(t,x) = \lim_{q\to e}\left(\frac{\partial}{\partial q^a}\right)t'(t,x;q),$$
$$\theta_a(t,x) = \lim_{q\to e}\left(\frac{\partial}{\partial q^a}\right)x'(t,x;q), \qquad (2.6)$$

as usual. The point e denotes the "identity point" in the group manifold, which, in the adopted parametrization of $\mathscr{P}_2$, has coordinates $q^1 = q^5 = 1$, and $q^2 = q^3 = q^4 = q^6 = q^7 = q^8 = 0$. Therefore, applying these standard manipulations to Eq. (2.4), we obtain the generators in the following form:

$$\eta_a = \dot{\hat{t}}^{-1}(\hat{t}\delta_{a1} + \hat{x}\delta_{a2} + \delta_{a3} - \hat{t}^2\delta_{a7} - \hat{t}\hat{x}\delta_{a8}),$$
$$\theta_a = u_2(\hat{t}\delta_{a4} + \hat{x}\delta_{a5} + \delta_{a6} - \hat{t}\hat{x}\delta_{a7} - \hat{t}^2\delta_{a8}) - u_2\hat{x}_t\eta_a, \qquad (2.7)$$

wherefrom the desired realization of the Lie algebra immediately follows. In fact, we get

$$Z_1 = \hat{t}\dot{\hat{t}}^{-1}\partial_t - u_2\hat{t}\dot{\hat{t}}^{-1}\hat{x}_t\partial_x,$$
$$Z_2 = \hat{x}\dot{\hat{t}}^{-1}\partial_t - u_2\dot{\hat{t}}^{-1}\hat{x}\hat{x}_t\partial_x,$$
$$Z_3 = \dot{\hat{t}}^{-1}\partial_t - u_2\dot{\hat{t}}^{-1}\hat{x}_t\partial_x,$$
$$Z_4 = u_2\hat{t}\partial_x,$$
$$Z_5 = u_2\hat{x}\partial_x, \qquad (2.8)$$
$$Z_6 = u_2\partial_x,$$
$$Z_7 = -\hat{t}^2\dot{\hat{t}}^{-1}\partial_t + u_2\hat{t}(\hat{t}\dot{\hat{t}}^{-1}\hat{x}_t - \hat{x})\partial_x,$$
$$Z_8 = -\hat{t}\hat{x}\dot{\hat{t}}^{-1}\partial_t + u_2\hat{x}(\hat{t}\dot{\hat{t}}^{-1}\hat{x}_t - \hat{x})\partial_x.$$

In the applications one substitutes from Eq. (2.3) into these expressions and thus one obtains the Z operators in terms of (t,x). The final, completely general, form of these operators is exhibited explicitly in Table I.

With the aim of obtaining the Lie algebra obeyed by these operators, it would be rather lengthy and cumbersome

TABLE I. Point symmetry realization of sl(3,R) for $\ddot{x} + f_2\dot{x} + f_1 x = f_0$. The Wronskian $w = \dot{u}_1 u_2 - u_1\dot{u}_2$ corresponds to the independent solutions $u_1(t)$ and $u_2(t)$, when $f_0 = 0$, and $u_P(t)$ is a particular solution of the inhomogeneous differential equation.

Z_a	$\eta_a(t,x)\partial_t + \theta_a(t,x)\partial_x$
Z_1	$w^{-1}u_1\{u_2\,\partial_t + (\dot{u}_2(x-u_P) + u_2\dot{u}_P)\partial_x\}$
Z_2	$w^{-1}(x-u_P)\{u_2\,\partial_t + (\dot{u}_2(x-u_P) + u_2\dot{u}_P)\partial_x\}$
Z_3	$w^{-1}u_2\{u_2\,\partial_t + (\dot{u}_2(x-u_P) + u_2\dot{u}_P)\partial_x\}$
Z_4	$u_1\,\partial_x$
Z_5	$(x-u_P)\partial_x$
Z_6	$u_2\,\partial_x$
Z_7	$-w^{-1}u_1\{u_1\,\partial_t + (\dot{u}_1(x-u_P) + u_1\dot{u}_P)\partial_x\}$
Z_8	$-w^{-1}(x-u_P)\{u_1\,\partial_t + (\dot{u}_1(x-u_P) + u_1\dot{u}_P)\partial_x\}$

to work out their commutators directly from Table I. So we shall follow a general approach already used in one of our previous works on this issue.[8] Since the Lie algebra must be of the form

$$[Z_a,Z_b] = f^c_{ab}Z_c, \qquad (2.9)$$

it follows that the structure constants f^c_{ab} have to be consistent with the following identities:

$$f^c_{ab}\eta_c = [\eta_a,\eta_{bt}] + [\theta_a,\eta_{bx}],$$
$$f^c_{ab}\theta_c = [\eta_a,\theta_{bt}] + [\theta_a,\theta_{bx}], \qquad (2.10)$$

where, of course, the square brackets denote antisymmetrization of the indices a and b only. [Let us note that Eq. (2.9) is *not* an ansatz, since we are certainly handling a Lie group and the algebra must be finite and closed.] Hence according to Eqs. (2.3), (2.4), and (2.7) a straightforward calculation yields

$$\eta_{at} = (\dot{\hat{t}})^{-1}\{(\dot{\hat{t}}\delta_{a1} + \hat{x}_t\delta_{a2} - 2\hat{t}\dot{\hat{t}}\delta_{a7} - (\dot{\hat{t}}\hat{x} + \hat{t}\hat{x}_t)\delta_{a8}) - \ddot{\hat{t}}\eta_a\},$$
$$\eta_{ax} = (u_2\dot{\hat{t}})^{-1}\{\delta_{a2} - \hat{t}\delta_{a8}\},$$
$$\theta_{at} = u_2\{(\dot{\hat{t}}\delta_{a4} + \hat{x}_t\delta_{a5} - (\dot{\hat{t}}\hat{x} + \hat{t}\hat{x}_t)\delta_{a7} - 2\hat{x}\hat{x}_t\delta_{a8}) - \hat{x}_t\eta_{at} - \hat{x}_{tt}\eta_a\} + u_2\dot{u}_2^{-1}\theta_a, \qquad (2.11)$$
$$\theta_{ax} = (\delta_{a5} - \hat{t}\delta_{a7} - 2\hat{x}\delta_{a8}) - u_2\hat{x}_t\eta_{ax} + u_2\dot{u}_2^{-1}\eta_a;$$

in consequence, after some steps we get

$$[\eta_a,\eta_{bt}] + [\theta_a,\eta_{bx}]$$
$$= \dot{\hat{t}}^{-1}\{([\delta_{a3},\delta_{b1}] + [\delta_{a6},\delta_{b2}]) - (2[\delta_{a3},\delta_{b7}] + [\delta_{a6},\delta_{b8}] + [\delta_{a2},\delta_{b5}])\hat{t}$$
$$- ([\delta_{a3},\delta_{b8}] + [\delta_{a1},\delta_{b2}] + [\delta_{a2},\delta_{b5}])\hat{x} - ([\delta_{a1},\delta_{b7}] + [\delta_{a4},\delta_{b8}])\hat{t}^2$$
$$- ([\delta_{a2},\delta_{b7}] + [\delta_{a5},\delta_{b8}])\hat{t}\hat{x}\}, \qquad (2.12)$$

TABLE II. The nonzero structure constants of the Lie algebra associated with the differential equation $\ddot{x}+f_2\dot{x}+f_1x=f_0$.

	$f^c_{ab}\neq 0$
f^1_{ab}	$-2[\delta_{a3},\delta_{b7}]-[\delta_{a6},\delta_{b8}]-[\delta_{a2},\delta_{b4}]$
f^2_{ab}	$-[\delta_{a3},\delta_{b8}]-[\delta_{a1},\delta_{b2}]-[\delta_{a2},\delta_{b5}]$
f^3_{ab}	$[\delta_{a3},\delta_{b1}]+[\delta_{a6},\delta_{b2}]$
f^4_{ab}	$-[\delta_{a6},\delta_{b7}]+[\delta_{a1},\delta_{b4}]+[\delta_{a4},\delta_{b5}]$
f^5_{ab}	$-[\delta_{a3},\delta_{b7}]-2[\delta_{a6},\delta_{b8}]+[\delta_{a2},\delta_{b4}]$
f^6_{ab}	$[\delta_{a3},\delta_{b4}]+[\delta_{a6},\delta_{b5}]$
f^7_{ab}	$[\delta_{a1},\delta_{b7}]+[\delta_{a4},\delta_{b8}]$
f^8_{ab}	$[\delta_{a2},\delta_{b7}]+[\delta_{a5},\delta_{b8}]$

and

$$\begin{aligned}&[\eta_a,\theta_{bt}]+[\theta_a,\theta_{bx}]\\&\quad=u_2\{([\delta_{a3},\delta_{b4}]+[\delta_{a6},\delta_{b5}])\\&\qquad+([\delta_{a1},\delta_{b4}]-[\delta_{a6},\delta_{b7}]\\&\qquad+[\delta_{a4},\delta_{b5}])\hat{t}+([\delta_{a2},\delta_{b4}]\\&\qquad-[\delta_{a3},\delta_{b7}]-2[\delta_{a6},\delta_{b8}])\hat{x}\\&\qquad-([\delta_{a1},\delta_{b7}]+[\delta_{a4},\delta_{b8}])\hat{t}\hat{x}\\&\qquad-([\delta_{a2},\delta_{b7}]+[\delta_{a5},\delta_{b8}])\hat{x}^2\}\\&\qquad-u_2\hat{x}_t([\eta_a,\eta_{bt}]+[\theta_a,\eta_{bx}]).\end{aligned}\tag{2.13}$$

On the other hand, Eq. (2.7) yields

$$f^c_{ab}\eta_c=\dot{\hat{t}}^{-1}\{f^1_{ab}\hat{t}+f^2_{ab}\hat{x}+f^3_{ab}-f^7_{ab}\hat{t}^2-f^8_{ab}\hat{t}\hat{x}\},\tag{2.14}$$

and

$$\begin{aligned}f^c_{ab}\theta_c&=u_2\{f^4_{ab}\hat{t}+f^5_{ab}\hat{x}+f^6_{ab}-f^7_{ab}\hat{t}\hat{x}-f^8_{ab}\hat{x}^2\}\\&\quad-u_2\hat{x}_tf^c_{ab}\eta_c.\end{aligned}\tag{2.15}$$

Therefore, using the fact that $\hat{t}(t,x)$ and $\hat{x}(t,x)$ are independent functions [cf. Eq. (2.3)], we are in position to obtain all the *nonzero* structure constants of the group, if we equalize Eq. (2.12) with (2.14), and Eq. (2.13) with (2.15), and then separate the coefficients of the different powers of $\hat{t}$, $\hat{x}$, $\hat{t}\hat{x}$, and $\hat{x}_t$. We present our results in Table II. For the sake of completeness we also include herein the Lie algebra obeyed by the Z operators; cf. Table III.

Furthermore, one can also obtain the infinitesimal operators $\hat{Z}_a$, $a=1,...,8$, associated with the variables $(\hat{t},\hat{x})$ defined in Eq. (2.3). We next present this subject (in a rather sketchy way, for the sake of brevity) because these operators bring the Lie algebra of $\mathscr{P}_2$ to the fore and, therefore, throw light on the algebra obeyed by the Z_a's. As was shown in Ref. 1, if one introduces the column

$$\begin{aligned}\mathbf{v}&=(\hat{t},\hat{x},1)\,(\text{transposed})\\&=(u_1(t)/u_2(t),(x-u_P(t))/u_2(t),1)(\text{transposed}),\end{aligned}$$

then the diffeomorphisms (2.2) are consistent with the projective transformation

$$\mathbf{v}'=\phi(\mathbf{v};q)\mathbf{M}(q)\cdot\mathbf{v}=(\mathbf{M}(q)\cdot\mathbf{v})/(\mathbf{M}(q)\cdot\mathbf{v})_3,\tag{2.16}$$

where

$$\mathbf{M}(q)=\begin{pmatrix}q^1&q^2&q^3\\q^4&q^5&q^6\\q^7&q^8&1\end{pmatrix},\tag{2.17}$$

and where, clearly, $(\mathbf{M}(q)\cdot\mathbf{v})_3$ stands for the third row in $\mathbf{M}(q)\cdot\mathbf{v}$. Of course, by being "consistent" here we mean that $\mathbf{v}'$ corresponds precisely to

$$\begin{aligned}\mathbf{v}'&=(\hat{t}',\hat{x}',1)\ (\text{transposed})\\&=(u_1(t')/u_2(t'),(x'-u_P(t'))/u_2(t'),1)\ (\text{transposed})\end{aligned}$$

(cf. Ref. 1 for details). Hence the generators $\mathbf{\Omega}_a$ of the infinitesimal transformations

$$\mathbf{v}'=\mathbf{v}+\delta q^a\mathbf{\Omega}_a\tag{2.18}$$

are given by

$$\mathbf{\Omega}_a=\lim_{q\to e}\left(\frac{\partial}{\partial q^a}\right)\mathbf{v}'=\phi^{(e)}_{,a}\mathbf{v}+\mathbf{M}^{(e)}_{,a}\cdot\mathbf{v},\tag{2.19}$$

where

$$\phi^{(e)}_{,a}=\lim_{q\to e}\left(\frac{\partial}{\partial q^a}\right)\phi(\mathbf{v};q)=-v_1\delta_{a7}-v_2\delta_{a8},\tag{2.20}$$

and

$$\mathbf{M}^{(e)}_{,a}=\lim_{q\to e}\left(\frac{\partial}{\partial q^a}\right)\mathbf{M}(q)=\begin{pmatrix}\delta_{a1}&\delta_{a2}&\delta_{a3}\\\delta_{a4}&\delta_{a5}&\delta_{a6}\\\delta_{a7}&\delta_{a8}&0\end{pmatrix}.\tag{2.21}$$

TABLE III. The well-known Lie algebra of the projective group in the plane.

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
Z_1	0	$-Z_2$	$-Z_3$	Z_4	0	0	Z_7	0
Z_2	Z_2	0	0	$-Z_1+Z_5$	$-Z_2$	$-Z_3$	Z_8	0
Z_3	Z_3	0	0	Z_6	0	0	$-2Z_1-Z_5$	$-Z_2$
Z_4	$-Z_4$	Z_1-Z_5	$-Z_6$	0	Z_4	0	0	Z_7
Z_5	0	Z_2	0	$-Z_4$	0	$-Z_6$	0	Z_8
Z_6	0	Z_3	0	0	Z_6	0	$-Z_4$	$-Z_1-2Z_5$
Z_7	$-Z_7$	$-Z_8$	$2Z_1+Z_5$	0	0	Z_4	0	0
Z_8	0	0	Z_2	$-Z_7$	$-Z_8$	Z_1+2Z_5	0	0

Thus for the $\hat{Z}_a$ operators, i.e., for $\hat{Z}_a = \Omega_{aj}\,\partial_j = \Omega_{a1}\,\partial_{\hat{t}} + \Omega_{a2}\,\partial_{\hat{x}}$, one gets

$$\begin{aligned}
&\hat{Z}_1 = \hat{t}\,\partial_{\hat{t}}, \quad \hat{Z}_2 = \hat{x}\,\partial_{\hat{t}},\\
&\hat{Z}_3 = \partial_{\hat{t}}, \quad \hat{Z}_4 = \hat{t}\,\partial_{\hat{x}},\\
&\hat{Z}_5 = \hat{x}\,\partial_{\hat{x}}, \quad \hat{Z}_6 = \partial_{\hat{x}},\\
&\hat{Z}_7 = -\hat{t}^2\,\partial_{\hat{t}} - \hat{t}\hat{x}\,\partial_{\hat{x}}, \quad \hat{Z}_8 = -\hat{t}\hat{x}\,\partial_{\hat{t}} - \hat{x}^2\,\partial_{\hat{x}},
\end{aligned} \tag{2.22}$$

which are the well-known infinitesimal operators of the projective transformations in the plane. One may pursue the analysis one step further, and show that the structure constants associated with the $\hat{Z}$ operators are the same structure constants of the Z operators displayed in Table II (as indeed they should be). Hence the point symmetry Lie algebra corresponding to all second-order linear differential equations is isomorphic to the algebra of the projective group in the plane.

III. WORLD LINE TRANSFORMATIONS AND THEIR LIE ALGEBRA REALIZATIONS

It is our purpose in this section to examine the same subject from a different point of view.

It is well known (and rather obvious) that one may interpret an equation like (2.2) in two different ways (cf. Ref. 9, for instance). Either one adopts a *passive* viewpoint and interprets Eq. (2.2) as a transformation of space-time coordinates (that is, as a change of frame of reference), or else one adopts an *active* viewpoint and interprets (2.2) as a transformation of space-time points (i.e., as a mapping of events). Both standpoints are logically equivalent, and both are extremely powerful for the purposes of geometry and mechanics. The distinction, however, is not trivial at all; indeed, whether one uses one or the other viewpoint depends highly on the kind of problem one is willing to tackle.[9]

These relativistic features are not out of place, since here we are dealing with the *relativity theory* of all one-dimensional Newtonian linear systems. As a matter of fact, according to the passive point of view, the group of transformations (2.2) defines the most general set of preferred frames relative to which a moving particle performs a well-defined *kind* of motion. Accordingly, one may also interpret (2.2) as an active mapping of events that interconverts one world line of the system into another. Indeed, the active transformations of allowable world lines are worth considering for they bring some novelties into the picture. Technically, the change from the passive to the active standpoint means that, instead of looking at the similarity properties of the *differential equation*, one looks directly at the symmetries of the *primitive curves*, from which the differential equation appears as the eliminant.[10] This interpretation settles the problem tackled in this section.

According to these comments, once the explicit expressions for the automorphisms (2.2) have been obtained, one can visualize these transformations as realizations of the active symmetry group that changes one world line of the system into another. Thus the equivalence

$$\begin{aligned}
&x(t) = \alpha u_1(t) + \beta u_2(t) + u_P(t)\\
&\quad \Leftrightarrow x'(t') = \alpha' u_1(t') + \beta' u_2(t') + u_P(t')
\end{aligned} \tag{3.1}$$

holds upon the mapping of space-time points stated in Eq. (2.2). Of course, this means that if one substitutes from Eq. (2.2) into (3.1), then after some manipulations one arrives necessarily at expressions for (α',β') of the general form

$$\alpha' = A(\alpha,\beta;q), \quad \beta' = B(\alpha,\beta;q). \tag{3.2}$$

In other words, the curves $x(t)$ given by

$$\begin{aligned}
S(t,x;q) &= A(\alpha,\beta;q)u_1(T(t,x;q))\\
&\quad + B(\alpha,\beta;q)u_2(T(t,x;q)) + u_P(T(t,x;q)),
\end{aligned} \tag{3.3}$$

with $t' = T$ and $x' = S$ as given in Eq. (2.2), are completely *independent* of the parameters $q = (q^1,\ldots,q^8)$ of the group. [In fact, these curves correspond precisely to $x = \alpha u_1(t) + \beta u_2(t) + u_P(t)$.] Moreover, this also means that after performing two successive space-time mappings, say

$$(t,x) \overset{q}{\rightarrow} (t',x') \overset{q'}{\rightarrow} (t'',x''),$$

Eqs. (3.2) are such that they yield, of necessity,

$$\begin{aligned}
\alpha'' &= A(A(\alpha,\beta;q),B(\alpha,\beta;q);q') = A(\alpha,\beta;g(q';q)),\\
\beta'' &= B(A(\alpha,\beta;q),B(\alpha,\beta;q);q') = B(\alpha,\beta;g(q';q)),
\end{aligned} \tag{3.4}$$

where the group multiplication functions $g^a(q';q) = q''^a$, $a = 1,\ldots,8$, are the same functions obtained from two successive $\{T,S\}$ transformations.[6] Hence, in brief, the transformations stated in Eqs. (3.2) provide us with a *new* realization of the *same* group that leaves invariant the equation of motion. This new realization has a place in the *state-space* $\{\alpha,\beta\}$ of the classical system. In conclusion, in Eqs. (2.2) and (3.2) one has two groups of automorphisms, which act in the spaces $\{t,x\}$ and $\{\alpha,\beta\}$, respectively, and which are isomorphic indeed.

Our first problem, then, is to find explicitly expressions for $A(\alpha,\beta;q)$ and $B(\alpha,\beta;q)$. As we have done in our previous paper,[1] let us tackle this problem using a "compact" (i.e., matrix) notation. Thus we introduce formally the column vectors $\mathbf{u} = (\alpha, -1, \beta)$ (transposed), and we consider their scalar product with $\mathbf{v}$ (previously introduced); i.e., we set

$$\mathbf{u}^T\cdot\mathbf{v} = \alpha\,\frac{u_1(t)}{u_2(t)} - \frac{x - u_P(t)}{u_2(t)} + \beta. \tag{3.5}$$

In this fashion our problem reduces to finding the transformations

$$\mathbf{u}' = \mathbf{N}(\mathbf{u};q)\cdot\mathbf{u}, \tag{3.6}$$

where $\mathbf{N}(\mathbf{u};q)$ is a 3×3 matrix such that

$$\mathbf{u}^T\cdot\mathbf{v} = 0 \Leftrightarrow \mathbf{u}'^T\cdot\mathbf{v}' = 0. \tag{3.7}$$

Thus we are ready to solve the problem of finding A and B. In order to assure the equivalence stated in Eq. (3.1) it is enough to assume that the scalar product (3.5) is conformally invariant:

$$\mathbf{u}'^T\cdot\mathbf{v}' = \sigma(\mathbf{u},\mathbf{v};q)\mathbf{u}^T\cdot\mathbf{v}, \tag{3.8}$$

where $\sigma(\mathbf{u},\mathbf{v};q)$ remains at our disposal. Hence

$$\phi(\mathbf{v};q)\mathbf{u}^T\cdot(\mathbf{N}^T(\mathbf{u};q)\cdot\mathbf{M}(q) - \tau(\mathbf{u};q)\mathbf{I})\cdot\mathbf{v} = 0 \tag{3.9}$$

follows, where we have chosen

$$\sigma(\mathbf{u},\mathbf{v};q) = \tau(\mathbf{u};q)\phi(\mathbf{v};q), \tag{3.10}$$

and where $\mathbf{I}$ stands for the identity matrix. Since Eq. (3.9) holds for all $\mathbf{v}$, we conclude that the desired transformation (3.6) is given by

$$\mathbf{u}' = \tau(\mathbf{u};q)(\mathbf{M}^T(q))^{-1}\cdot\mathbf{u}, \tag{3.11}$$

where, according to the definition of $\mathbf{u}$, we have to require $(\alpha, -1, \beta) \to (\alpha', -1, \beta')$. This requirement fixes $\tau(\mathbf{u};q)$, and so we get the final answer,

$$\mathbf{u}' = -(\mathbf{P}(q)\cdot\mathbf{u})/(\mathbf{P}(q)\cdot\mathbf{u})_2, \tag{3.12}$$

where, clearly, we have written

$$\mathbf{P}(q) = (\mathbf{M}^T(q))^{-1}, \tag{3.13}$$

and wherefrom the explicit form of the transformation (3.1) follows. Of course, $(\mathbf{P}(q)\cdot\mathbf{u})_2$ denotes the second row of $\mathbf{P}(q)\cdot\mathbf{u}$. Since

$$(\mathbf{M}^T(q'))^{-1}\cdot(\mathbf{M}^T(q))^{-1} = ((\mathbf{M}(q')\cdot\mathbf{M}(q))^T)^{-1},$$

one has

$$\mathbf{P}(q')\cdot\mathbf{P}(q) = \mathbf{P}(g(q';q)). \tag{3.14}$$

Furthermore, this fact yields

$$\mathbf{u}'' = -\frac{\mathbf{P}(q')\cdot\mathbf{u}'}{(\mathbf{P}(q')\cdot\mathbf{u}')_2} = -\frac{\mathbf{P}(g(q';q))\cdot\mathbf{u}}{(\mathbf{P}(g(q';q))\cdot\mathbf{u})_2}, \tag{3.15}$$

as required. This result entails the fundamental property we had already stated in Eqs. (3.4) in a formal fashion. The matrix $\mathbf{P}(q)$ is given by

$$\mathbf{P}(q) = (\det(\mathbf{M}(q)))^{-1} \times \begin{pmatrix} q^5 - q^6q^8 & -q^4 + q^6q^7 & q^4q^8 - q^5q^7 \\ -q^2 + q^3q^8 & q^1 - q^3q^7 & q^2q^7 - q^1q^8 \\ q^2q^6 - q^3q^5 & q^3q^4 - q^1q^6 & q^1q^5 - q^2q^4 \end{pmatrix}. \tag{3.16}$$

Hence the infinitesimal operators W_a, $a = 1,...,8$, associated with the realization (3.12) of SL(3,R), in the $\{\alpha,\beta\}$ space, follow immediately. Indeed, let us write

$$u_j' = \psi(\mathbf{u};q)P_{jk}(q)u_k, \tag{3.17}$$

instead of Eq. (3.12), where

$$\psi(\mathbf{u};q) = -(P_{2k}(q)u_k)^{-1} = (P_{22} - P_{21}\alpha - P_{23}\beta)^{-1}.$$

Then one has

$$\begin{aligned} W_a = {} & (\psi^{(e)}_{,a} + P^{(e)}_{11,a})\alpha\,\partial_\alpha \\ & + (\psi^{(e)}_{,a} + P^{(e)}_{33,a})\beta\,\partial_\beta + P^{(e)}_{31,a}\alpha\,\partial_\beta \\ & + P^{(e)}_{13,a}\beta\,\partial_\alpha - P^{(e)}_{12,a}\,\partial_\alpha - P^{(e)}_{32,a}\,\partial_\beta. \end{aligned} \tag{3.18}$$

Since, finally,

$$\psi^{(e)}_{,a} = \delta_{a5} - \alpha\delta_{a2} - \beta\delta_{a8} \tag{3.19}$$

and

$$\mathbf{P}^{(e)}_{,a} = -\begin{pmatrix} \delta_{a1} & \delta_{a4} & \delta_{a7} \\ \delta_{a2} & \delta_{a5} & \delta_{a8} \\ \delta_{a3} & \delta_{a6} & 0 \end{pmatrix}, \tag{3.20}$$

one obtains the operators shown in Table IV. One calculates easily the structure constants of the algebra, which, of course, are the same already shown in Table II. Let us observe that the realization of the Lie algebra sl(3,R) in the classical state-space $\{\alpha,\beta\}$ is the same for all linear one-dimensional Newtonian systems. On the other hand, the space-time realizations of this algebra are specific and differ from one system to another.

TABLE IV. World line realization of the "active" sl(3,R) in the classical state space $\{\alpha,\beta\}$ of a linear system, i.e., $x(t) = \alpha u_1(t) + \beta u_2(t) + u_P(t) \to x'(t+\delta t) = (\alpha + \delta\alpha)u_1(t+\delta t) + (\beta + \delta\beta)u_2(t+\delta t) + u_P(t+\delta t)$.

W_a	$\eta_a(\alpha,\beta)\partial_\alpha + \theta_a(\alpha,\beta)\partial_\beta$
W_1	$-\alpha\,\partial_\alpha$
W_2	$-\alpha^2\,\partial_\alpha - \alpha\beta\,\partial_\beta$
W_3	$-\alpha\,\partial_\beta$
W_4	∂_α
W_5	$\alpha\,\partial_\alpha + \beta\,\partial_\beta$
W_6	∂_β
W_7	$-\beta\,\partial_\alpha$
W_8	$-\alpha\beta\,\partial_\alpha - \beta^2\,\partial_\beta$

IV. SOME MISCELLANEOUS EXAMPLES

In this section we present some interesting instances of the realization of the point symmetry Lie algebra associated with linear one-dimensional systems. The chosen examples correspond to the same systems already considered in Ref. 1, which are taken from elementary mechanics and analysis.

(a) *Free particle.* One has $\ddot{x} = 0$. Thus we take $u_1(t) = t$, $u_2(t) = 1$, and $u_P(t) = 0$. Then Table I yields immediately

$$\begin{aligned} & Z_1 = t\,\partial_t, \quad Z_2 = x\,\partial_t, \\ & Z_3 = \partial_t, \quad Z_4 = t\,\partial_x, \\ & Z_5 = x\,\partial_x, \quad Z_6 = \partial_x, \\ & Z_7 = -t^2\,\partial_t - tx\,\partial_x, \quad Z_8 = -tx\,\partial_t - x^2\,\partial_x. \end{aligned} \tag{4.1}$$

Clearly, this corresponds to the familiar realization of the algebra of the projective group in the plane, as a glance at Eq. (2.22) shows neatly. This case entails a trivial check of the formalism. Also, in this particular case, it is very easy to check the form of the operators $W_a(\alpha,\beta)$, shown in Table IV for the general case.

(b) *Free falling particle.* Now we set $\ddot{x} + g = 0$. Thus we have, for instance, $u_1(t) = t$, $u_2(t) = 1$, $u_P(t) = -\frac{1}{2}gt^2$, wherefrom Table I yields the following operators:

$$\begin{aligned} & Z_1 = t(\partial_t - gt\,\partial_x), \quad Z_2 = (x + \tfrac{1}{2}gt^2)(\partial_t - gt\,\partial_x), \\ & Z_3 = \partial_t - gt\,\partial_x, \quad Z_4 = t\,\partial_x, \\ & Z_5 = (x + \tfrac{1}{2}gt^2)\partial_x, \quad Z_6 = \partial_x, \\ & Z_7 = -t^2\,\partial_t - t(x - \tfrac{1}{2}gt^2)\partial_x, \\ & Z_8 = -(x + \tfrac{1}{2}gt^2)(t\,\partial_t + (x - \tfrac{1}{2}gt^2)\partial_x). \end{aligned} \tag{4.2}$$

Of course, any other admissible choice for the u's corresponds merely to a reparametrization of the projective group and, therefore, induces a new basis for the realization of the sl(3,R) algebra.

(c) *Simple harmonic oscillator.* For the equation $\ddot{x} + \omega^2 x = 0$, we take $u_1(t) = \sin \omega t$, $u_2(t) = \cos \omega t$, and $u_P(t) = 0$. Thus we have

$$\begin{aligned}
Z_1 &= (1/\omega)(\sin \omega t)\{(\cos \omega t)\partial_t - \omega(\sin \omega t)x\,\partial_x\},\\
Z_2 &= (1/\omega)x\{(\cos \omega t)\partial_t - \omega(\sin \omega t)x\,\partial_x\},\\
Z_3 &= (1/\omega)(\cos \omega t)\{(\cos \omega t)\partial_t - \omega(\sin \omega t)x\,\partial_x\},\\
Z_4 &= (\sin \omega t)\partial_x,\\
Z_5 &= x\,\partial_x,\\
Z_6 &= (\cos \omega t)\partial_x,\\
Z_7 &= -(1/\omega)(\sin \omega t)\{(\sin \omega t)\partial_t + \omega(\cos \omega t)x\,\partial_x\},\\
Z_8 &= -(1/\omega)x\{(\sin \omega t)\partial_t + \omega(\cos \omega t)x\,\partial_x\}.
\end{aligned} \tag{4.3}$$

These operators are equivalent to the infinitesimal operators obtained in Ref. 6, within a suitable reparametrization of the group.

(d) *A forced harmonic oscillator.* Let us consider the inhomogeneous equation of motion $\ddot{x} + \omega^2 x = f_0 \sin \Omega t$, where f_0 is a constant. We take $u_1(t) = \sin \omega t$, $u_2(t) = \cos \omega t$, and $u_P(t) = -(f_0/(\Omega^2 - \omega^2))\sin \Omega t$. In this fashion, we obtain from Table I the following infinitesimal operators:

$$\begin{aligned}
Z_1 &= \frac{1}{\omega}(\sin \omega t)\Big\{(\cos \omega t)\partial_t - \omega(\sin \omega t)x\,\partial_x\\
&\quad - \frac{f_0}{\Omega^2 - \omega^2}(\Omega \cos \Omega t \cos \omega t + \omega \sin \Omega t \sin \omega t)\partial_x\Big\},\\
Z_2 &= \frac{1}{\omega}\Big(x + \frac{f_0}{\Omega^2 - \omega^2}\sin \Omega t\Big)\\
&\quad \times\Big\{(\cos \omega t)\partial_t - \omega(\sin \omega t)x\,\partial_x\\
&\quad - \frac{f_0}{\Omega^2 - \omega^2}(\Omega \cos \Omega t \cos \omega t + \omega \sin \Omega t \sin \omega t)\partial_x\Big\},\\
Z_3 &= \frac{1}{\omega}(\cos \omega t)\Big\{(\cos \omega t)\partial_t - \omega(\sin \omega t)x\,\partial_x\\
&\quad - \frac{f_0}{\Omega^2 - \omega^2}(\Omega \cos \Omega t \cos \omega t + \omega \sin \Omega t \sin \omega t)\partial_x\Big\},\\
Z_4 &= (\sin \omega t)\partial_x,\\
Z_5 &= \Big(x + \frac{f_0}{\Omega^2 - \omega^2}\sin \Omega t\Big)\partial_x,\\
Z_6 &= (\cos \omega t)\partial_x,\\
Z_7 &= -\frac{1}{\omega}(\sin \omega t)\Big\{(\sin \omega t)\partial_t + \omega(\cos \omega t)x\,\partial_x\\
&\quad - \frac{f_0}{\Omega^2 - \omega^2}(\Omega \cos \Omega t \sin \omega t - \omega \sin \Omega t \cos \omega t)\partial_x\Big\},\\
Z_8 &= -\frac{1}{\omega}\Big(x + \frac{f_0}{\Omega^2 - \omega^2}\sin \Omega t\Big)\\
&\quad \times\Big\{(\sin \omega t)\partial_t + \omega(\cos \omega t)x\,\partial_x\\
&\quad - \frac{f_0}{\Omega^2 - \omega^2}(\Omega \cos \Omega t \sin \omega t - \omega \sin \Omega t \cos \omega t)\partial_x\Big\}.
\end{aligned} \tag{4.4}$$

Thus one has a rather formidable realization of sl(3,R).

(e) *Damped harmonic oscillator.* Now we consider the equation $\ddot{x} + 2\lambda\dot{x} + \omega^2 x = 0$, for which we set

$$u_1(t) = e^{-\lambda t}\sin \Omega t, \quad u_2(t) = e^{-\lambda t}\cos \Omega t,$$

with $\Omega = \sqrt{\omega^2 - \lambda^2}$, and $u_P(t) = 0$. So we get the operators

$$\begin{aligned}
Z_1 &= (1/\Omega)(\sin \Omega t)\{(\cos \Omega t)\partial_t\\
&\quad - (\lambda \cos \Omega t + \Omega \sin \Omega t)x\,\partial_x\},\\
Z_2 &= (1/\Omega)e^{\lambda t}x\{(\cos \Omega t)\partial_t\\
&\quad - (\lambda \cos \Omega t + \Omega \sin \Omega t)x\,\partial_x\},\\
Z_3 &= (1/\Omega)(\cos \Omega t)\{(\cos \Omega t)\partial_t\\
&\quad - (\lambda \cos \Omega t + \Omega \sin \omega t)x\,\partial_x\},\\
Z_4 &= e^{-\lambda t}(\sin \Omega t)\partial_x,\\
Z_5 &= x\,\partial_x,\\
Z_6 &= e^{-\lambda t}(\cos \Omega t)\partial_x,\\
Z_7 &= -(1/\Omega)(\sin \Omega t)\{(\sin \Omega t)\partial_t\\
&\quad - (\lambda \sin \Omega t - \Omega \cos \Omega t)x\,\partial_x\},\\
Z_8 &= -(1/\Omega)x\{(\sin \Omega t)\partial_t\\
&\quad - (\lambda \sin \Omega t - \Omega \cos \Omega t)x\,\partial_x\}.
\end{aligned} \tag{4.5}$$

(f) *Falling particle in a viscous media.* Now let the equation be $\ddot{x} + \lambda\dot{x} + g = 0$, with $u_1(t) = e^{-\lambda t}$, $u_2(t) = 1$, and $u_P(t) = -(g/\lambda)t$. Thus

$$\begin{aligned}
Z_1 &= -\frac{1}{\lambda}\Big(\partial_t - \frac{g}{\lambda}\partial_x\Big),\\
Z_2 &= -\frac{1}{\lambda}e^{\lambda t}\Big(x + \frac{g}{\lambda}t\Big)\Big(\partial_t - \frac{g}{\lambda}\partial_x\Big),\\
Z_3 &= -\frac{1}{\lambda}e^{\lambda t}\Big(\partial_t - \frac{g}{\lambda}\partial_x\Big), \quad Z_4 = e^{-\lambda t}\partial_x,\\
Z_5 &= \Big(x + \frac{g}{\lambda}t\Big)\partial_x, \quad Z_6 = \partial_x,\\
Z_7 &= e^{-\lambda t}\Big\{\frac{1}{\lambda}\partial_t - \Big(\frac{g}{\lambda^2} + x + \frac{g}{\lambda}t\Big)\partial_x\Big\},\\
Z_8 &= \Big(x + \frac{g}{\lambda}t\Big)\Big\{\frac{1}{\lambda}\partial_t - \Big(\frac{g}{\lambda^2} + x + \frac{g}{\lambda}t\Big)\partial_x\Big\}.
\end{aligned} \tag{4.6}$$

(g) *Inifinitesimal operators of* $\ddot{x} + t^{-1}\dot{x} - t^{-2}x = 0$. Although not very interesting from the point of view of mechanics, we did consider this equation in Ref. 1 as an example of a linear differential equation with time-dependent coefficients. We take $u_1(t) = t$, $u_2(t) = t^{-1}$, $u_P(t) = 0$ (i.e., $t = 0$ is a regular singular point). In this way we readily obtain the associated infinitesimal operators:

$$\begin{aligned}
&Z_1 = \tfrac{1}{2}(t\,\partial_t - x\,\partial_x), \quad Z_2 = \tfrac{1}{2}t^{-1}x(t\,\partial_t - x\,\partial_x),\\
&Z_3 = \tfrac{1}{2}t^{-2}(t\,\partial_t - x\,\partial_x), \quad Z_4 = t\,\partial_x,\\
&Z_5 = x\,\partial_x, \quad Z_6 = t^{-1}\partial_x,\\
&Z_7 = -\tfrac{1}{2}t^2(t\,\partial_t + x\,\partial_x), \quad Z_8 = -\tfrac{1}{2}tx(t\,\partial_t + x\,\partial_x).
\end{aligned} \tag{4.7}$$

All these examples correspond to realizations of sl(3,R), with structure constants as given in Table II (cf. also Table III).

ACKNOWLEDGMENTS

This work was supported in part by DIUC(PUCCH) Grant No. 050/87, by DGI(UCV) Grant No. 123.759/87, and in part by FONDECYT Grants No. 0670/87 and No. 0364/88.

[1]M. Aguirre and J. Krause, J. Math. Phys. **29**, 9 (1988).
[2]G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[3]L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[4]For instance, the rediscovery ten years ago [C. E. Wulfman and B. G. Wybourne, J. Phys. A: Math. Gen. **9**, 507 (1976)] that the Newtonian equation for the simple harmonic oscillator has point symmetry SL(3,R) was a surprise to physicists.
[5]S. Lie, *Vorlesungen über Differentialgleichungen mit bekannten infinitesimalen Transformationen* (Teubner, Leipzig, 1891; reprinted by Chelsea, New York, 1967).
[6]Cf. also M. Aguirre and J. Krause, J. Phys. A: Math. Gen. **20**, 3553 (1987), concerning the finite point realizations of SL(3,R) for the simple harmonic oscillator.
[7]J. Krause, J. Phys. A: Math. Gen. **18**, 1309 (1985); J. Math. Phys. **27**, 2922 (1986).
[8]By the way, in this sense the present article also closes a line of research that was initiated by us some years ago; cf. M. Aguirre and J. Krause, J. Math. Phys. **25**, 210 (1984).
[9]J. L. Synge, *Relativity: The Special Theory* (North-Holland, Amsterdam, 1965).
[10]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).

Branching rules for unitary representations of Virasoro and super-Virasoro algebras at $c=1$

Michael Baake
Physikalisches Institut der Universität Bonn, Nussallee 12, D-5300 Bonn 1, West Germany

(Received 3 January 1988; accepted for publication 23 March 1988)

The branching rules for unitary highest weight representations of Virasoro and super-Virasoro algebras that contain a subalgebra with central charge $c = 1$ are presented. These rules are useful for identifying points of higher symmetry in two-dimensional critical systems with $c = 1$ and for the analysis of a defected Ising chain.

I. INTRODUCTION

The investigation of two-dimensional critical systems with second-order phase transitions[1–3] has recently been stimulated by the discovery that their spectra at the critical point in the finite-size scaling limit can be described by unitary highest weight representations[4] of the Virasoro algebra $\mathscr{V}_c$, which is given by[5]

$$[L_m,L_n] = (m-n)\cdot L_{m+n} + (c/12)m(m^2-1)\delta_{m+n,0}\,, \tag{1.1}$$

where c is the so-called central charge. For the discrete series[2] with $c<1$ this is achieved by finitely many irreps, which, in general, is not possible for systems with $c\geqslant 1$. However, it may then happen that infinitely many irreps collapse to a single irrep of a larger algebra simultaneously indicating a higher symmetry.[2,6,7] From the mathematical point of view, this phenomenon corresponds to the branching rules between the irreps of the two algebras. Conversely, these branching rules are necessary to check a conjectured higher symmetry for validity.

Recently, several systems with $c = 1$ were analyzed that depend on a parameter, e.g., a defected Ising chain,[8,9] the Ashkin–Teller quantum chain,[6,10] and the *XXZ*–Heisenberg chain.[11] Here, several points of higher symmetry exist. For example, at the Ising decoupling point of the Ashkin–Teller model one has the direct product of two Virasoro algebras with $c=\frac{1}{2}$, and for certain values of the coupling constant the spectra are given by irreps of the super-Virasoro algebra.[6,12]

In what follows, we derive the corresponding branching rules for unitary irreps of several algebras that contain a Virasoro subalgebra with $c = 1$. We are only interested in unitary, highest weight irreps, which means $L_m^+ = L_{-m}$ and the existence of a unique (up to normalization) state $|\Delta\rangle$ with $L_m|\Delta\rangle = 0$, for $m>0$, and $L_0|\Delta\rangle = \Delta|\Delta\rangle$. This Δ is called the anomalous dimension.

The paper is organized as follows. In Sec. II the double-Ising scenario is investigated, where $\mathscr{V}_1$ is considered as a subalgebra of $\mathscr{V}_{1/2}\times\mathscr{V}_{1/2}$. Section III covers the U(1) Kac–Moody algebra together with the bosonic and fermionic oscillator representations of $\mathscr{V}_1$, which play a central role in $c = 1$ systems. The $N = 1$ superconformal algebra is discussed in Sec. IV while all other cases are briefly treated in the concluding section (Sec. V) in a simple, unified manner.

II. THE DOUBLE-ISING SCENARIO

Let us first consider the direct product $\mathscr{V}_c\times\mathscr{V}_c$ of two (commuting) Virasoro algebras with the same central charge c. Then, the sums of the generators, namely $L_m = L_m^{(1)} + L_m^{(2)}$, build a subalgebra that itself is a Virasoro algebra, however, with central charge $2c$. For the case $c=\frac{1}{2}$ we will now derive the corresponding branching rules. As is well known[1,5] irreps of $\mathscr{V}_{1/2}\times\mathscr{V}_{1/2}$ are labeled by a pair (Δ_1,Δ_2) of anomalous dimensions where each Δ_i separately can be 0, $\frac{1}{2}$, or $\frac{1}{16}$. The unitary irreps of $\mathscr{V}_1$ are labeled by a single anomalous dimension (Δ), which can be any non-negative real number.

To proceed, we need the characters $\chi^{\rm V}_{c,\Delta}(z)$, for $c=\frac{1}{2}$ and $c = 1$. With the abbreviation

$$\Pi_{\rm V}(z) = \prod_{m=1}^{\infty}\frac{1}{1-z^m}\,, \tag{2.1}$$

we have, for $c=\frac{1}{2}$ (see Refs. 4 and 5),

$$\chi^{\rm V}_{1/2,0}(z) = \sum_{n\in\mathbb{Z}}(z^{12n^2+n} - z^{12n^2+7n+1})\Pi_{\rm V}(z)\,,$$

$$\chi^{\rm V}_{1/2,1/2}(z) = z^{1/2}\sum_{n\in\mathbb{Z}}(z^{12n^2-5n} - z^{12n^2+13n+3})\Pi_{\rm V}(z)\,,$$

$$\chi^{\rm V}_{1/2,1/16}(z) = z^{1/16}\sum_{n\in\mathbb{Z}}(z^{12n^2-2n} - z^{12n^2+10n+2})\Pi_{\rm V}(z)\,. \tag{2.2}$$

For $c = 1$, the general formula reads[13]

$$\chi^{\rm V}_{1,\Delta}(z) = \begin{cases} z^{\Delta}\Pi_{\rm V}(z)\,, & \text{if } \Delta\neq m^2/4\,,\\ z^{m^2/4}(1-z^{m+1})\cdot\Pi_{\rm V}(z)\,, & \text{if } \Delta = m^2/4\,,\end{cases}$$
$$m \text{ integer},\quad m\geqslant 0\,. \tag{2.3}$$

By means of Jacobi's triple product identity,[14]

$$\prod_{m=1}^{\infty}(1-z^{2m})(1+xz^{2m-1})\Big(1+\frac{1}{x}z^{2m-1}\Big) = \sum_{n\in\mathbb{Z}}x^n z^{n^2},\quad x\neq 0,\quad |z|<1\,, \tag{2.4}$$

and Watson's quintuple product identity[15]

$$\prod_{m=1}^{\infty}(1-z^{2m})(1-xz^{2m})\Big(1-\frac{1}{x}z^{2m-2}\Big)\times(1-x^2z^{4m-2})\Big(1-\frac{1}{x^2}z^{4m-2}\Big) = \sum_{n\in\mathbb{Z}}z^{n(3n+1)}(x^{3n}-x^{-3n-1}),\quad x\neq 0,\quad |z|<1\,, \tag{2.5}$$

 0022-2488/88/081753-05$02.50

one finds for the characters of Eq. (2.2) the formulas

$$\chi^{V}_{1/2,0}(z) = \sum_{n\in\mathbb{Z}} z^{4n^2+n}\Pi_V(z^2)\,,$$
$$\chi^{V}_{1/2,1/2}(z) = \sum_{n\in\mathbb{Z}} z^{4n^2+3n+1/2}\Pi_V(z^2)\,, \tag{2.6}$$
$$\chi^{V}_{1/2,1/16}(z) = z^{1/16}\prod_{m=1}^{\infty}(1+z^m)\,,$$

and, furthermore, the useful identity

$$\chi^{V}_{1/2,0}(z) \pm \chi^{V}_{1/2,1/2}(z) = \prod_{m=1}^{\infty}(1 \pm z^{m-1/2})\,. \tag{2.7}$$

The character of the irrep (Δ_1,Δ_2) of $\mathscr{V}_{1/2}\times\mathscr{V}_{1/2}$ is nothing but $\chi^{V}_{1/2,\Delta_1}(z_1)\chi^{V}_{1/2,\Delta_2}(z_2)$. Identifying z_1 and z_2, i.e., $z=z_1=z_2$, the product can be transformed in order to determine the character of the corresponding reducible representation of $\mathscr{V}_1$. For $(\Delta_1,\Delta_2)\neq(0,0)$ or $(\frac{1}{2},\frac{1}{2})$, we have

$$\chi^{V}_{1/2,0}(z)\chi^{V}_{1/2,1/16}(z) = \sum_{n\in\mathbb{Z}} z^{(1/16)(8n+1)^2}\Pi_V(z)\,,$$
$$\chi^{V}_{1/2,0}(z)\chi^{V}_{1/2,1/2}(z) = \sum_{n=0}^{\infty} z^{(1/2)(2n+1)^2}\Pi_V(z)\,,$$
$$\chi^{V}_{1/2,1/2}(z)\chi^{V}_{1/2,1/16}(z) = \sum_{n\in\mathbb{Z}} z^{(1/16)(8n+3)^2}\Pi_V(z)\,, \tag{2.8}$$
$$\chi^{V}_{1/2,1/16}(z)\chi^{V}_{1/2,1/16}(z) = \sum_{n=0}^{\infty} z^{(1/8)(2n+1)^2}\Pi_V(z)\,.$$

The cases (0,0) and $(\frac{1}{2},\frac{1}{2})$ require a more subtle calculation because $\mathscr{V}_1$ irreps (Δ) with $\Delta = m^2/4$, m integer, arise in the decomposition. Here, it is advantageous to start from Eq. (2.2) and to use the identity ($|z|<1$)

$$\sum_{m,n\in\mathbb{Z}} z^{2r(m^2+n^2)+s(m+n)} = \frac{1}{2}\sum_{k,l\in\mathbb{Z}}\left(z^{r(k^2+l^2)+sk} + (-1)^{k+l}z^{r(k^2+l^2)+sk}\right) \tag{2.9}$$

before going to product form. This way one obtains

$$\chi^{V}_{1/2,0}(z)\chi^{V}_{1/2,0}(z) = \left\{\sum_{n=1}^{\infty} z^{2n^2} + \sum_{n=0}^{\infty} z^{4n^2}(1-z^{4n+1})\right\}\Pi_V(z)\,,$$
$$\chi^{V}_{1/2,1/2}(z)\chi^{V}_{1/2,1/2}(z) = \left\{\sum_{n=1}^{\infty} z^{2n^2} + \sum_{n=0}^{\infty} z^{(2n+1)^2}(1-z^{4n+3})\right\}\Pi_V(z)\,. \tag{2.10}$$

From Eqs. (2.8) and (2.9) one can directly identify the irreps of $\mathscr{V}_1$ by means of Eq. (2.3), which yields the branching rules

$$(0,0)\downarrow_{\mathscr{V}_1} = \bigoplus_{m=1}^{\infty}(2m^2)\oplus\bigoplus_{m=0}^{\infty}((2m)^2)\,,$$
$$(0,\tfrac{1}{16})\downarrow_{\mathscr{V}_1} = \bigoplus_{m\in\mathbb{Z}}(\tfrac{1}{16}(8m+1)^2)\,,$$
$$(0,\tfrac{1}{2})\downarrow_{\mathscr{V}_1} = \bigoplus_{m=0}^{\infty}(\tfrac{1}{2}(2m+1)^2)\,,$$
$$(\tfrac{1}{16},\tfrac{1}{16})\downarrow_{\mathscr{V}_1}\quad \bigoplus_{m=0}^{\infty}(\tfrac{1}{8}(2m+1)^2)\,, \tag{2.11}$$
$$(\tfrac{1}{16},\tfrac{1}{2})\downarrow_{\mathscr{V}_1} = \bigoplus_{m\in\mathbb{Z}}(\tfrac{1}{16}(8m+3)^2)\,,$$
$$(\tfrac{1}{2},\tfrac{1}{2})\downarrow_{\mathscr{V}_1}\quad \bigoplus_{m=1}^{\infty}(2m^2)\oplus\bigoplus_{m=0}^{\infty}((2m+1)^2)\,.$$

Note that these results are unique as a result of the convergence of the power series involved.

The relation to the bosonic and fermionic oscillator representations of the Virasoro algebra are discussed in Sec. III. Let us, at this point, briefly discuss $\mathscr{V}_1$ as a subalgebra of $\mathscr{V}_{1/2}$, a case that occurs in the treatment of a defected Ising chain.[8,9] Let L_m, $m\in\mathbb{Z}$, generate $\mathscr{V}_{1/2}$. Then the new generators

$$R_m = \tfrac{1}{2}L_{2m} + \tfrac{1}{32}\delta_{m,0} \tag{2.12}$$

define a Virasoro algebra with $c=1$, $\mathscr{V}_1$, that is a subalgebra of $\mathscr{V}_{1/2}$. Substituting $q=z^2$ in Eq. (2.6) one gets the branching rules

$$(0)\downarrow_{\mathscr{V}_1} = \bigoplus_{n\in\mathbb{Z}}(2n^2+\tfrac{1}{2}n+\tfrac{1}{32})\,,$$
$$(\tfrac{1}{2})\downarrow_{\mathscr{V}_1} = \bigoplus_{n\in\mathbb{Z}}(2n^2+\tfrac{3}{2}n+\tfrac{9}{32})\,, \tag{2.13}$$
$$(\tfrac{1}{16})\downarrow_{\mathscr{V}_1} = \bigoplus_{n>0}((n/4)(n+1)+\tfrac{1}{16}) = \bigoplus_{m\in\mathbb{Z}}((4m+1)^2/16)\,.$$

The shift $\frac{1}{32}$ results from R_0 and is explicitly seen in the spectrum of the defected Ising chain. Since on the right-hand side of Eq. (2.13) no irrep with $\Delta = m^2/4$, m integer, occurs the conformal tower always has the standard degeneracy given by $\Pi_V(q)$.

III. OSCILLATOR REPRESENTATIONS OF $\mathscr{V}_1$ AND THE U(1) KAC–MOODY ALGEBRA

Let us begin with a formula that immediately follows from Eq. (2.3):

$$\sum_{k=0}^{\infty}\chi^{V}_{1,(1/4)(m+2k)^2}(z) = z^{(1/4)m^2}\Pi_V(z)\,. \tag{3.1}$$

This is related to the bosonic realization of the Virasoro algebra with $c=1$. There, one has the Sugawara structure[7,16]

$$L_m = \frac{1}{2}\sum_{r\in\mathbb{Z}} :T_{m-r}T_r:\,, \tag{3.2}$$

where : : denotes normal ordering, together with

$$[T_m,L_n] = mT_{m+n}\,,\quad [T_m,T_n] = m\delta_{m+n,0}\,. \tag{3.3}$$

In fact, this defines a U(1) Kac–Moody algebra, where the Hermiticity condition is assumed to be $T^{+}_{m} = T_{-m}$ (and hence $L^{+}_{m} = L_{-m}$). Since

$$L_0 = \frac{1}{2}T_0^2 + \sum_{k=1}^{\infty}T_{-k}T_k\,, \tag{3.4}$$

one obtains $\Delta = \varphi^2/2$ for the irreps where the charge φ is the real eigenvalue of $T_0 = T_0^{+}$. Here φ labels the irreps of the

U(1) Kac–Moody algebra, the corresponding character is given by

$$\chi_{\varphi}^{U(1)}(z,y) = \mathrm{tr}(z^{L_0}y^{T_0}) = z^{\varphi^2/2}y^{\varphi}\Pi_V(z) . \tag{3.5}$$

Taking $y = 1$, a comparison of Eq. (3.5) with Eqs. (2.3) and (3.1) yields the decomposition into $\mathscr{V}_1$ irreps,

$$(\varphi)^{U(1)}\downarrow_{\mathscr{V}_1} = \begin{cases} (\Delta), & \text{if } \Delta \neq m^2/4, \\ \bigoplus_{l=0}^{\infty} ((m+2l)^2/4), & \text{if } \Delta = m^2/4, \end{cases}$$
$$m \in \mathbb{N}_0, \quad \Delta = \tfrac{1}{2}\varphi^2 . \tag{3.6}$$

This formula has been derived previously[17] without reference to the simplifying Kac–Moody structure.

At this point, we will shortly discuss the properties of the twisted U(1) Kac–Moody algebra,[7,18] which is obtained from the untwisted one [Eq. (3.3)] by taking T_μ ($\mu \in \mathbb{Z} + \frac{1}{2}$) instead of T_m ($m \in \mathbb{Z}$). The Sugawara construction then gives

$$L_m = \frac{1}{2} \sum_{\mu \in \mathbb{Z} + 1/2} :T_{m-\mu} T_\mu: + \frac{1}{16}\delta_{m,0} . \tag{3.7}$$

The contribution of $\frac{1}{16}$ to L_0 is necessary in order to match the commutation rules. As a consequence, one obtains

$$L_0 = \frac{1}{16} + \sum_{\mu \in \mathbb{N} - 1/2} T_{-\mu} T_\mu . \tag{3.8}$$

Because no zero mode is present, we have only one irrep with the fixed anomalous dimension $\Delta = \frac{1}{16}$, labeled $(\frac{1}{16})_T^{U(1)}$. Its character (see also Refs. 7 and 18) is given by

$$\chi_T^{U(1)}(z) = \mathrm{tr}(z^{L_0}) = z^{1/16} \sum_{m=1}^{\infty} \pi_{\mathrm{odd}}(m) z^{m/2} , \tag{3.9}$$

where $\pi_{\mathrm{odd}}(m)$ is the number of partitions of m into odd integers. By means of the corresponding generating function[19] and of the triple product formula (2.4) one finds

$$\begin{aligned} \chi_T^{U(1)}(z) &= z^{1/16} \prod_{m=1}^{\infty} \frac{1}{1 - z^{m-1/2}} \\ &= z^{1/16}\Pi_V(z) \prod_{m=1}^{\infty} (1 + z^{m-1/2})(1 - z^{2m}) \\ &= \sum_{n \in \mathbb{Z}} z^{(1/16)(4n+1)^2} \Pi_V(z) , \end{aligned} \tag{3.10}$$

which simultaneously yields the branching rule into irreps of the Virasoro algebra [please note the coincidence with Eq. (2.13)]:

$$\begin{aligned} (\tfrac{1}{16})_T^{U(1)} \downarrow_{\mathscr{V}_1} &= \bigoplus_{n \in \mathbb{Z}} (\tfrac{1}{16}(4n+1)^2) \\ &= \bigoplus_{m \in \mathbb{Z}} (\tfrac{1}{16}(8m+1)^2) \oplus \bigoplus_{m \in \mathbb{Z}} (\tfrac{1}{16}(8m+3)^2) . \end{aligned} \tag{3.11}$$

The twisted U(1) Kac–Moody algebra (3.7) leaves us with a curious situation. Although L_0 has a very similar structure in the untwisted (3.4) and the twisted (3.8) case, respectively, the irrep $(\frac{1}{16})_T^{U(1)}$ decomposes into infinitely many Virasoro irreps with the full degeneracy [given by $\Pi_V(z)$]—a phenomenon that never occurs in the untwisted case [cf. (3.6)]! For the occurrence of U(1) Kac–Moody structures in statistical systems with $c = 1$, the reader is referred to Refs. 7, 18, and 19.

Before we turn to the case of higher symmetries, let us shortly comment on the fermionic oscillator representations. For $c = \frac{1}{2}$, they are well-known (e.g., Ref. 16 and references therein) to be

$$L'_m = \frac{1}{2} \sum_{\mu \in \mathbb{Z} + 1/2} \mu :a_{m-\mu} a_\mu: , \tag{3.12}$$

for the Neveu–Schwarz sector ($a_\mu^+ = a_{-\mu}$, $\{a_\mu, a_\nu\} = \delta_{\mu+\nu,0}$), and

$$L''_m = \frac{1}{2} \sum_{r \in \mathbb{Z}} r :a_{m-r} a_r: + \frac{1}{16}\delta_{m,0} , \tag{3.13}$$

for the Ramond sector ($a_r^+ = a_{-r}$, $\{a_r, a_s\} = \delta_{r+s,0}$). Equation (3.12) contains the irreps (0) and $(\frac{1}{2})$ while (3.13) contains two copies of the irrep $(\frac{1}{16})$ due to the Z_2-zero mode a_0 (since a_0 is not in the Cartan subalgebra, we have, strictly speaking, a degenerate highest weight representation). By the sum of two generators, e.g.,

$$L_m = \frac{1}{2} \sum_{\mu \in \mathbb{Z} + 1/2} \mu (:a_{m-\mu} a_\mu: + :b_{m-\mu} b_\mu:) \tag{3.14}$$

(where we assume $\{a_\mu, b_\nu\} = 0$) one obtains a Virasoro representation with $c = 1$ and a content of irreps that can be read from Eq. (2.11) adding the contributions of (0,0), $(0,\frac{1}{2})$, $(\frac{1}{2},0)$, and $(\frac{1}{2},\frac{1}{2})$. In fact, we can build a complete U(1) Kac–Moody algebra by means of

$$T_m = i \sum_{\mu \in \mathbb{Z} + 1/2} a_{m-\mu} b_\mu \tag{3.15}$$

with the property that the L's of Eq. (3.2)—though quartic in a and b—coincide with those of Eq. (3.14) because they have the same matrix elements. Completely analogous arguments are valid for the other combinations of Eqs. (3.12) and (3.13). Two copies of the Ramond sector result in an irrep content corresponding to $4 \cdot (\frac{1}{16}, \frac{1}{16})$ while a combination of one Ramond with one Neveu–Schwarz generator builds twice the irrep $(\frac{1}{16})_T^{U(1)}$ of the twisted U(1) Kac–Moody algebra [cf. Eqs. (3.8)–(3.11)].

IV. THE $N=1$ SUPERCONFORMAL ALGEBRA

Let us now consider the Ramond and Neveu–Schwarz algebras defined by Eq. (1.1) together with (cf. Refs. 2, 4, and 5 and references therein)

$$\begin{aligned} [L_m, G_r] &= (m/2 - r) G_{m+r} , \\ \{G_r, G_s\} &= 2L_{r+s} + (c/3)(r^2 - \tfrac{1}{4})\delta_{r+s,0} . \end{aligned} \tag{4.1}$$

This is a supersymmetric extension of the Virasoro algebra where $m \in \mathbb{Z}$ and either $r,s \in \mathbb{Z}$ (Ramond case) or $r,s \in \mathbb{Z} + \frac{1}{2}$ (Neveu–Schwarz case).

Obviously, the generators L_m, $m \in \mathbb{Z}$, build an ordinary Virasoro algebra with central charge c with respect to which the unitary irreps of the super-Virasoro algebra ($L_m^+ = L_{-m}$, $G_r^+ = G_{-r}$) decompose completely. Of special interest are the cases where c takes one of the discrete values $c < \frac{3}{2}$ for the super-Virasoro algebra[2] since this again results in a quantization of the possible anomalous dimensions Δ. In what follows, we focus again on $c=1$ ($c=0$ is trivial and $c=\frac{7}{10}$ is treated completely in Ref. 20).

For a derivation of the branching rules, we need again the character formulas of Goddard *et al.*[5] Let us first investigate the Ramond case. With the abbreviation

$$\Pi_R(z)=\prod_{m=1}^{\infty}\frac{1+z^m}{1-z^m}=\prod_{m=1}^{\infty}(1+z^m)\cdot\Pi_V(z)\,, \tag{4.2}$$

we obtain by means of Eqs. (2.4) and (2.5)

$$\chi^R_{1,1/24}(z)=z^{1/24}\sum_{n\in\mathbf{Z}}(z^{12n^2}-z^{12n^2+12n+3})\Pi_R(z) = \sum_{n\in\mathbf{Z}}z^{(1/24)(6n+1)^2}\Pi_V(z)\,, \tag{4.3}$$

$$\chi^R_{1,1/16}(z)=z^{1/16}\sum_{n\in\mathbf{Z}}(z^{12n^2+n}-z^{12n^2+7n+1})\Pi_R(z) = \chi^V_{1/2,0}(z)\chi^V_{1/2,1/16}(z) = \sum_{n\in\mathbf{Z}}z^{(1/16)(8n+1)^2}\Pi_V(z)\,, \tag{4.4}$$

$$\chi^R_{1,9/16}(z)=z^{9/16}\sum_{n\in\mathbf{Z}}(z^{12n^2+5n}-z^{12n^2+13n+3})\Pi_R(z) = \chi^V_{1/2,1/2}(z)\chi^V_{1/2,1/16}(z) = \sum_{n\in\mathbf{Z}}z^{(1/16)(8n+3)^2}\Pi_V(z)\,, \tag{4.5}$$

$$\chi^R_{1,3/8}(z)=z^{3/8}\sum_{n\in\mathbf{Z}}(z^{12n^2-4n}-z^{12n^2+8n+1})\Pi_R(z) = \sum_{m=0}^{\infty}z^{(3/8)(2m+1)^2}\Pi_V(z)\,. \tag{4.6}$$

From these formulas one can directly read the following branching rules:

$$\begin{aligned}(\tfrac{1}{24})^R\downarrow_{\mathscr{V}_1}&=\bigoplus_{n\in\mathbf{Z}}(\tfrac{1}{24}(6n+1)^2)\,,\\ (\tfrac{1}{16})^R\downarrow_{\mathscr{V}_1}&=\bigoplus_{n\in\mathbf{Z}}(\tfrac{1}{16}(8n+1)^2)\,,\\ (\tfrac{9}{16})^R\downarrow_{\mathscr{V}_1}&=\bigoplus_{n\in\mathbf{Z}}(\tfrac{1}{16}(8n+3)^2)\,,\\ (\tfrac{3}{8})^R\downarrow_{\mathscr{V}_1}&=\bigoplus_{m=0}^{\infty}(\tfrac{3}{8}(2m+1)^2)\,.\end{aligned} \tag{4.7}$$

Let us now turn to the Neveu–Schwarz case where we define

$$\Pi_{NS}(z)=\prod_{m=1}^{\infty}\frac{1+z^{m-1/2}}{1-z^m}=\prod_{m=1}^{\infty}(1+z^{m-1/2})\cdot\Pi_V(z)\,. \tag{4.8}$$

From Eq. (2.7) and the characters of Goddard *et al.*[5] it is easy to calculate

$$\chi^{NS}_{1,1/16}(z)=z^{1/16}\sum_{n\in\mathbf{Z}}(z^{12n^2-2n}-z^{12n^2+10n+2})\Pi_{NS}(z) = \chi^V_{1/2,1/16}(z)\{\chi^V_{1/2,0}(z)+\chi^V_{1/2,1/2}(z)\} = \sum_{n\in\mathbf{Z}}z^{(1/16)(4n+1)^2}\Pi_V(z)\,. \tag{4.9}$$

Simultaneously, this proves the formula

$$\chi^{NS}_{1,1/16}(z)=\chi^R_{1,1/16}(z)+\chi^R_{1,9/16}(z) = \chi^V_{1/2,0}(z)\chi^V_{1/2,1/16}(z)+\chi^V_{1/2,1/2}(z)\chi^V_{1/2,1/16}(z)\,. \tag{4.10}$$

For $\Delta=\frac{1}{6}$, one has—by means of Eqs. (2.4) and (2.5)—

$$\chi^{NS}_{1,1/6}(z)=z^{1/6}\sum_{n\in\mathbf{Z}}(z^{12n^2+3n}-z^{12n^2+15n+4+1/2})\Pi_{NS}(z) = \sum_{n\in\mathbf{Z}}(z^{(1/6)(6n+1)^2}+z^{(2/3)(3n+1)^2})\Pi_V(z). \tag{4.11}$$

The remaining characters ($\Delta=0$ and $\Delta=1$) read

$$\begin{aligned}\chi^{NS}_{1,0}(z)&=\sum_{n\in\mathbf{Z}}(z^{12n^2-n}-z^{12n^2+5n+1/2})\Pi_{NS}(z)\,,\\ \chi^{NS}_{1,1}(z)&=\sum_{n\in\mathbf{Z}}(z^{12n^2-7n+1}-z^{12n^2+11n+2+1/2})\Pi_{NS}(z)\,.\end{aligned} \tag{4.12}$$

Instead of a direct calculation, it is advantageous to determine the sum and the difference of these characters. From Eqs. (4.12), (2.2), and (2.7) one gets

$$\chi^{NS}_{1,0}(z)-\chi^{NS}_{1,1}(z)=(\chi^V_{1/2,0}(z))^2-(\chi^V_{1/2,1/2}(z))^2 = \sum_{n\in\mathbf{Z}}(z^{4n^2}-z^{(2n+1)^2})\Pi_V(z)\,, \tag{4.13}$$

while for the sum one finds, with the substituting $q=i\,z^{1/4}$,

$$\chi^{NS}_{1,0}(z)+\chi^{NS}_{1,1}(z)=\sum_{n\in\mathbf{Z}}q^{n(3n-1)}\Pi_{NS}(z) = \sum_{n\in\mathbf{Z}}(z^{6n^2}+z^{6n^2+6n+3/2})\Pi_V(z)\,. \tag{4.14}$$

Combining the last two equations we obtain the formulas

$$\begin{aligned}\chi^{NS}_{1,0}(z)&=\Big\{\sum_{n=1}^{\infty}z^{6n^2}+\sum_{n=0}^{\infty}z^{(3/2)(2n+1)^2} + \sum_{n=0}^{\infty}z^{4n^2}(1-z^{4n+1})\Big\}\Pi_V(z)\,,\\ \chi^{NS}_{1,1}(z)&=\Big\{\sum_{n=1}^{\infty}z^{6n^2}+\sum_{n=0}^{\infty}z^{(3/2)(2n+1)^2} + \sum_{n=0}^{\infty}z^{(2n+1)^2}(1-z^{4n+3})\Big\}\Pi_V(z)\,.\end{aligned} \tag{4.15}$$

From Eqs. (4.9), (4.11), (4.15), and (2.3) we can extract the branching rules for the Neveu–Schwarz algebra

$$\begin{aligned}(0)^{NS}\downarrow_{\mathscr{V}_1}&=\bigoplus_{k=1}^{\infty}(6k^2)\oplus\bigoplus_{k=0}^{\infty}(\tfrac{3}{2}(2k+1)^2)\oplus\bigoplus_{k=0}^{\infty}(4k^2)\,,\\ (1)^{NS}\downarrow_{\mathscr{V}_1}&=\bigoplus_{k=1}^{\infty}(6k^2)\oplus\bigoplus_{k=0}^{\infty}(\tfrac{3}{2}(2k+1)^2)\oplus\bigoplus_{k=0}^{\infty}((2k+1)^2)\,,\\ (\tfrac{1}{16})^{NS}\downarrow_{\mathscr{V}_1}&=\bigoplus_{n\in\mathbf{Z}}(\tfrac{1}{16}(8n+1)^2)\oplus\bigoplus_{n\in\mathbf{Z}}(\tfrac{1}{16}(8n+3)^2)\,,\\ (\tfrac{1}{6})^{NS}\downarrow_{\mathscr{V}_1}&=\bigoplus_{n\in\mathbf{Z}}(\tfrac{1}{6}(6n+1)^2)\oplus\bigoplus_{n\in\mathbf{Z}}(\tfrac{2}{3}(3n+1)^2)\,.\end{aligned} \tag{4.16}$$

V. CONCLUDING REMARKS

In the treatment of quantum spin chains with $c = 1$ one finds the $N = 2$ rather than the $N = 1$ superconformal algebra,[6,11,18,21] which stems from the underlaying bosonic structure. Hence one also needs the decomposition of $N = 2$ irreps into $N = 1$ irreps and into U(1) Kac–Moody irreps, which is implicitly given in several publications (e.g., Refs. 22 and 23). Since this is also related to the representation theory of shifted and twisted Kac–Moody algebras[24,25] we can refer the reader to another publication[7] where the occurrence of higher symmetries in systems with $c = 1$ was discussed in terms of irreps of in general still unknown algebras, labeled by the quotient p/q of two coprime integers. It was claimed that for every positive rational number such an algebra exists [in fact, $p/q = 1$ is the SU(2) Kac–Moody algebra, $p/q = \frac{3}{2}$ the $N = 2$ superconformal algebra, $p/q = \frac{1}{2}$ the double-Ising algebra, and $p/q = 3$ corresponds to the Zamolodchikov–Fateev invariance.[26]

The vacuum representation $\langle 0\rangle^{p/q}$ of the (conjectured) p/q algebra was given explicitly in terms of irreps of the U(1) Kac–Moody algebra,[7] which is contained as a subalgebra,

$$\langle 0\rangle^{p/q} = \bigoplus_{m\in\mathbf{Z}} (\sqrt{2p/q}\,m)^{\mathrm{U}(1)} \tag{5.1}$$

[here, T_0 is always taken with the normalization of Eq. (3.3)]. All other irreps are considered either as a *shift* of (5.1), namely,

$$\langle \rho\rangle^{p/q} = \bigoplus_{m\in\mathbf{Z}} (\sqrt{2p/q}\,(m+\rho))^{\mathrm{U}(1)} \tag{5.2}$$

(wherefrom the branching rules are obvious), or as a representation of the *twisted* version of that algebra. This irrep can only be $(\frac{1}{16})_T^{p/q}$ [cf. (3.8)] but it contains two copies of $(\frac{1}{16})_T^{\mathrm{U}(1)}$ because of the existence of a zero mode not contained in the Cartan subalgebra.

At this point, the treatment of algebras with $c = 1$ seems to be rather complete. However, one should resolve the restriction to $c = 1$. Unfortunately, the branching rules for $c > 1$ become much more complicated because there the multiplicities of the $\mathscr{V}_c$-irreps arising explode. One then needs more complicated techniques to face the problem of missing labels, which will be interesting in itself.

ACKNOWLEDGMENTS

The author would like to thank P. D. Jarvis and P. Reicke for helpful comments and the DFG for financial support.

[1]A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, J. Stat. Phys. **34**, 763 (1984); Nucl. Phys. B **241**, 333 (1984).
[2]D. Friedan, Z. Qiu, and S. Shenker, Phys. Rev. Lett. **52**, 1575 (1984); Phys. Lett. B **151**, 37 (1985).
[3]J. L. Cardy, Nucl. Phys. B **240**, 514 (1984); **270**, 186 (1986); G. v. Gehlen, V. Rittenberg, and H. Ruegg, J. Phys. A: Math. Gen. **19**, 107 (1985).
[4]A. Rocha-Caridi, in *Vertex Operators in Mathematics and Physics*, edited by J. Lepowsky, S. Mandelstam, and I. M. Singer (Springer, New York, 1985), p. 451; in *Infinite Lie Algebras and Conformal Invariance in Condensed Matter and Particle Physics*, edited by K. Dietz and V. Rittenberg (World Scientific, Singapore, 1987), p. 59.
[5]P. Goddard, A. Kent, and D. Olive, Commun. Math. Phys. **103**, 105 (1986).
[6]M. Baake, G. v. Gehlen, and V. Rittenberg, J. Phys. A: Math. Gen. **20**, L479, L487 (1987).
[7]M. Baake, Ph. Christe, and V. Rittenberg, to be published in Nucl. Phys. B.
[8]M. Henkel and A. Patkos, Nucl. Phys. B **285**, 29 (1987).
[9]M. Baake, P. Chaselon, and M. Schlottmann, "SO(2n) Kac–Moody structures in the Ising quantum chain with equidistant defects," in preparation.
[10]M. Baake, Ph.D. thesis, Bonn-IR-87-09, 1987.
[11]F. Alacaraz, M. Baake, U. Grimm, and V. Rittenberg, J. Phys. A: Math. Gen. **21**, L117 (1988).
[12]G. v. Gehlen and V. Rittenberg, J. Phys. A: Math. Gen. **19**, L1039 (1986); S.-K. Yang, Nucl. Phys. B **285** [FS19], 183 (1987).
[13]V. Kac, Lect. Notes Phys. **94**, 444 (1979).
[14]G. E. Andrews, Proc. Am. Math. Soc. **16**, 333 (1965).
[15]L. Carlitz and M. V. Subbarao, Proc. Am. Math. Soc. **32**, 42 (1972).
[16]E. Corrigan, Phys. Lett. B **169**, 259 (1986).
[17]M. Wakimoto and H. Yamada, Hiroshima Math. J. **16**, 427 (1986).
[18]G. v. Gehlen, V. Rittenberg, and G. Schütz, to be published in J. Phys. A: Math. Gen.
[19]R. Dijkgraaf, E. Verlinde, and H. Verlinde, Utrecht preprint THU-87/17; P. Ginsparg, Harvard preprint HUTP-87/A068.
[20]Y. Matsuo and S. Yahikosawa, Phys. Lett. B **178**, 211 (1986).
[21]G. Waterson, Phys. Lett. B **171**, 77 (1986); S.-K. Yang and H. B. Zheng, Nucl. Phys. B **285**, 410 (1987).
[22]Y. Matsuo, preprint Tokyo-UT-494, 1986.
[23]F. Ravanini and S.-K. Yang, Phys. Lett. B **193**, 202 (1987).
[24]P. Goddard and D. Olive, Int. J. Mod. Phys. A **1**, 303 (1987).
[25]A. Schwimmer and N. Seiberg, Phys. Lett. B **184**, 191 (1987).
[26]A. B. Zamolodchikov and V. A. Fateev, Sov. Phys. JETP **62**, 215 (1985).

On the representation of Slater-type densities by Gaussian densities

G. Braunss and E. Teubner

Mathematisches Institut und Institut für Theoretische Physik, Justus-Liebig-Universität, Giessen, West Germany

(Received 16 March 1988; accepted for publication 23 March 1988)

It is proved that for arbitrary $m,n\in\mathbb{N}_0$ and $\alpha > 0$, $\beta > 0$, there exists an integral representation $|x|^m\,|y|^n \exp(-\alpha|x| - \beta|y|) = \int_{\mathbb{R}_+\times\mathbb{R}} K_{mn}(s,t)\exp[-(s+it)|x|^2 - (s-it)|y|^2]d(s,t)$, $x,y\in\mathbb{R}^3$, where $K_{mn}(s,t)$ is a singular distribution in $\mathscr{D}'(\mathbb{R})\otimes\mathscr{Z}'(\mathbb{R})$.

I. INTRODUCTION AND LEMMA

While trying to construct solutions of the time-dependent Hartree–Fock equation[1] the following problem occured: Let $x,y\in\mathbb{R}^3$, let $|x| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$, etc., let $\zeta = s + it\in\mathbb{C}$, $\bar{\zeta} = s - it$, and let $\rho(|x|,|y|;s,t) = \exp(-\zeta|x|^2 - \bar{\zeta}|y|^2)$. Does there exist then a distribution K_{mn} with support in $\mathbb{R}_+\times\mathbb{R}$ so that for $m,n\in\mathbb{N}_0$ (= natural numbers including zero) and $\alpha > 0, \beta > 0$,

$$|x|^m|y|^n e^{-\alpha|x| - \beta|y|} = \int_{\mathbb{R}^2} K_{mn}(s,t)\rho(|x|,|y|;s,t)d(s,t)?$$

By using Fourier transformations it can be shown that this problem is equivalent to the following one: Let $p,q\in\mathbb{R}$, and let

$$f(p^2,q^2) = [(\lambda + p^2)(\mu + q^2)]^{-1},$$
$$\lambda = \alpha^2,\quad \mu = \beta^2.$$

Does there exist then a distribution G with support in $\mathbb{R}_+\times\mathbb{R}$ so that

$$f(p^2,q^2) = \int_{\mathbb{R}^2} G(s,t)\rho(p,q;s,t)d(s,t)?$$

The lemma below will answer this question in the affirmative.

First let us define our notation: $\mathscr{D}(\mathbb{R})$ and $\mathscr{Z}(\mathbb{R})$ will denote the (locally convex) spaces of test functions with compact support which are infinitely often differentiable and of entire analytic functions, respectively; $\mathscr{D}'(\mathbb{R})$ and $\mathscr{Z}'(\mathbb{R})$ denote the corresponding dual spaces (cf. Refs. 2 and 3). By Θ we shall denote the heaviside function: $\Theta(t) = 0$ for $t<0$, and $\Theta(t) = 1$ for $t>0$. Thus $\delta^{(k)}(t) \equiv d^{k+1}\Theta(t)/dt^{k+1}$, $k\in\mathbb{N}_0$, denotes the kth derivative of the Dirac distribution concentrated at the origin.

Lemma: For $\lambda > 0$, $\mu > 0$, and $(p,q)\in\mathbb{R}^2$, let

$$f(p^2,q^2) = [(\lambda + p^2)(\mu + q^2)]^{-1}. \tag{1}$$

Then there exists a distribution $G\in\mathscr{D}'(\mathbb{R})\otimes\mathscr{Z}'(\mathbb{R})$ with support in $\mathbb{R}_+\times\mathbb{R}$ so that

$$f(p^2,q^2) = \int_{\mathbb{R}^2} G(s,t) \times\exp[-(p^2+q^2)s - i(p^2-q^2)t]d(s,t). \tag{2}$$

The distribution G is explicitly given by

$$G(s,t) = (2/\pi)\Theta(s)e^{-as-ibt}\int_{\mathbb{R}} \tau^{-1}\sinh(s\tau)e^{it\tau}\,d\tau \tag{3a}$$

$$= 4\Theta(s)e^{-as-ibt} \times\sum_{0\leqslant k}(-1)^k s^{2k+1}\delta^{(2k)}(t)/(2k+1)!, \tag{3b}$$

where $a = \lambda + \mu$, $b = \lambda - \mu$.

II. PROOF OF THE LEMMA

We shall construct $G(s,t)$ in a straightforward manner and then verify (2). Writing $x_1\equiv p^2$, $x_2\equiv q^2$, $z_1 = x_1 + iy_1$, and $z_2 = x_2 + iy_2$, the function $f(x_1,x_2)$ has an analytic continuation $f(z_1,z_2)$ that is analytic for all y_1, y_2 and all $x_1 > -\lambda$, $x_2 > -\mu$. Let us make the following substitutions:

$$u_1 = x_1 + x_2,\quad v_1 = y_1 + y_2,\quad w_1 = u_1 + iv_1;$$
$$u_2 = x_1 - x_2,\quad v_2 = y_1 - y_2,\quad w_2 = u_2 + iv_2.$$

Thus for real p,q,

$$u_1 = p^2 + q^2\geqslant 0,\quad u_1 + u_2 = 2p^2\geqslant 0,\quad u_1 - u_2\geqslant 2q^2.$$

Assume now a distribution G with support in $\mathbb{R}_+\times\mathbb{R}$ to exist so that

$$f(z_1,z_2) = 4[(2\lambda + w_1 + w_2)(2\mu + w_1 - w_2)]^{-1}$$
$$= \int_{\mathbb{R}^2} G(s,t)e^{-w_1 s - iw_2 t}\,d(s,t)$$
$$= \int_{\mathbb{R}^2} G(s,t)e^{-u_1 s + v_2 t}e^{-i(v_1 s + u_2 t)}\,d(s,t).$$

Inverting (formally) the Fourier transformation one obtains

$$(2\pi)^2 G(s,t)e^{-u_1 + v_2 t} = 4\int_{\mathbb{R}^2}\frac{e^{i(v_1 s + u_2 t)}}{(2\lambda + w_1 + w_2)(2\mu + w_1 - w_2)}\,d(v_1,u_2). \tag{4}$$

Since in the end we shall verify our result and, besides, do not claim uniqueness, this procedure is justified. Let

$$b_1\equiv 2\lambda + u_1 + u_2 = 2(\lambda + p^2) > 0,$$
$$b_2\equiv 2\mu + u_1 - u_2 = 2(\mu + q^2) > 0.$$

By the residue formula,

$$I_1 = \int_{\mathbf{R}} \frac{e^{iv_1 s}}{2\lambda + w_1 + w_2}\, dv_1 = 2\pi\Theta(s) e^{-(b_1 + iv_2)s},$$

$$I_2 = \int_{\mathbf{R}} \frac{e^{iv_1 s}}{2\mu + w_1 - w_2}\, dv_1 = 2\pi\Theta(s) e^{-(b_2 - iv_2)s}.$$

Partial fraction decomposition of the integrand in (4) yields

$$\begin{aligned}(2\pi)^2 G(s,t) e^{-u_1 s + v_2 t} &= 2\int_{\mathbf{R}} (I_1 - I_2)\frac{e^{itu_2}}{\mu - \lambda - w_2}\, du_2\\ &= 8\pi\Theta(s) e^{-(\lambda + \mu + u_1)s}\\ &\quad \times \int_{\mathbf{R}} \frac{\sinh[(w_2 + \lambda - \mu)s]}{\lambda - \mu + w_2} e^{itu_2}\, du_2\\ &= 8\pi\Theta(s) e^{-(a + u_1)s} e^{-i(b + iv_2)t}\\ &\quad \times \int_{-\infty + iv_2}^{\infty + iv_2} \tau^{-1}\sinh(s\tau) e^{it\tau}\, d\tau, \qquad (5)\end{aligned}$$

where $a = \lambda + \mu$, $b = \lambda - \mu$. Consider a regularization such that the last integral in (5) is replaced by

$$\begin{aligned}I_\varepsilon(s,t) &= \int_{-\infty + iv_2}^{\infty + iv_2} e^{-\varepsilon\tau^2}\tau^{-1}\sinh(s\tau) e^{it\tau}\, d\tau\\ &= \lim_{R\to\infty}\left\{\left(\int_{-R}^{R} + \int_{-R + iv_2}^{-R} + \int_{R}^{R + iv_2}\right)\right.\\ &\quad \left.\times\, e^{-\varepsilon\tau^2}\tau^{-1}\sinh(s\tau) e^{it\tau}\, d\tau\right\}, \quad \varepsilon > 0.\end{aligned}$$

One has, for $R, v \in \mathbf{R}$,

$$\begin{aligned}&\left|\int_R^{R + iv} e^{-\varepsilon\tau^2}\tau^{-1}\sinh(s\tau) e^{it\tau}\, d\tau\right|\\ &= \left|\int_0^v e^{-\varepsilon(R + i\tau)^2}(R + i\tau)^{-1}\right.\\ &\quad \left.\times \sinh(s(R + i\tau)) e^{it(R + i\tau)}\, d\tau\right|\\ &\leqslant e^{-\varepsilon R^2}\cosh(sR)\int_0^{|v|}(R^2 + \tau^2)^{-1/2} e^{\varepsilon\tau^2 + |t|\tau}\, d\tau\\ &\leqslant |v| R^{-1} e^{-\varepsilon R^2}\cosh(sR) e^{\varepsilon v^2 + |tv|} \to 0, \quad \text{for } |R| \to \infty.\end{aligned}$$

Thus in the distribution sense

$$\begin{aligned}G(s,t) &= (2/\pi)\Theta(s) e^{-as - ibt}\lim_{\varepsilon\to 0} I_\varepsilon(s,t)\\ &= \frac{2}{\pi}\Theta(s) e^{-as - ibt}\int_{\mathbf{R}} \tau^{-1}\sinh(s\tau) e^{it\tau}\, d\tau. \qquad (6)\end{aligned}$$

The function $h(z) \equiv z^{-1}\sinh(sz)$ is, for each $s \in \mathbf{R}$, entire analytic, hence we can expand it in a Taylor series and integrate termwise. The result is

$$G(s,t) = 4\Theta(s) e^{-as - ibt}\sum_{0\leqslant k}\frac{(-1)^k s^{2k+1}\delta^{(2k)}(t)}{(2k + 1)!}. \qquad (7)$$

Before verifying (2) we shall prove that $G \in \mathscr{D}'(\mathbf{R}) \otimes \mathscr{Z}'(\mathbf{R})$, thereby showing that the infinite sum (7) makes sense. Let $\psi \in \mathscr{Z}(\mathbf{R})$, $\alpha(s) \equiv 4\Theta(s) e^{-as}$, and

$$G_N(s,t) = \alpha(s) e^{-ibt}\sum_{0\leqslant k\leqslant N}\frac{(-1)^k s^{2k+1}\delta^{(2k)}(t)}{(2k + 1)!}.$$

A short calculation yields

$$\begin{aligned}(G_N(s,.),\psi) &\equiv \int_{\mathbf{R}} G_N(s,t)\psi(t)\, dt\\ &= \alpha(s)\sum_{0\leqslant j\leqslant 2N} C_{N,j}(s)\psi^{(j)}(0),\end{aligned}$$

where $\psi^{(j)} \equiv d^j\psi/dt^j$ and

$$\begin{aligned}C_{N,j}(s) &= \sum_{0\leqslant k\leqslant N}(-1)^k\binom{2K}{j}\frac{(-ib)^{2k-j}s^{2k+1}}{(2k + 1)!}\\ &= -i(-ib)^{-j-1}(j!)^{-1}\\ &\quad \times \sum_{j/2\leqslant k\leqslant N}(sb)^{2k+1}((2k - j)!(2k + 1))^{-1}\\ &= -i(-ib)^{-j-1}(j!)^{-1}\int_0^{sb}\tau^j\\ &\quad \times \sum_{j/2\leqslant k\leqslant N}\tau^{2k-j}((2k - j)!)^{-1}\, d\tau\end{aligned}$$

[using the convention $\binom{2k}{j} = 0$ for $j > 2k$].
From this it follows that

$$\begin{aligned}C_j(s) &= \lim_{N\to\infty} C_{N,j}(s)\\ &= -i(-ib)^{-j-1}(j!)^{-1}\int_0^{sb}\tau^j h_j(\tau)\, d\tau,\end{aligned}$$

where

$$h_j(\tau) = \begin{cases}\cosh\tau, & \text{for } j \text{ even},\\ \sinh\tau, & \text{for } j \text{ odd}.\end{cases}$$

Partial integration yields ($h_j' \equiv dh_j/d\tau$)

$$\begin{aligned}|C_j(s)| &\leqslant [|b|^{j+1}(j + 1)!]^{-1}\left|(sb)^{j+1}h_j(sb)\right.\\ &\quad \left. - \int_0^{sb}\tau^{j+1}h_j'(\tau)\, d\tau\right|\\ &\leqslant 2s^{j+1}e^{s|b|}/(j + 1)!, \quad s \geqslant 0.\end{aligned}$$

Consequently,

$$\begin{aligned}|(G(s,.),\psi)| &= \lim_{N\to\infty}|(G_N(s,.),\psi)|\\ &\leqslant 8\Theta(s) e^{-(a - |b|)s}\sum_{0\leqslant j}\frac{s^{j+1}|\psi^{(j)}(0)|}{(j + 1)!}.\end{aligned}$$

Since ψ is entire analytic there exists a finite positive constant A_ψ such that

$$\sup_{0\leqslant j}|\psi^{(j)}(0)| \leqslant A_\psi.$$

Further, if $c = 2\min(\lambda,\mu)$ then

$$0 < a - |b| = \lambda + \mu - |\lambda - \mu| = c.$$

Thus for arbitrary $\phi \in \mathscr{D}(\mathbf{R})$,

$$\begin{aligned}|(G,\phi\otimes\psi)| &\leqslant 8\int_0^\infty e^{-(a - |b|)s}|\phi(s)|\sum_{0\leqslant j}\frac{s^{j+1}|\psi^{(j)}(0)|\, ds}{(j + 1)!}\\ &\leqslant 8A_\psi\int_0^\infty e^{-cs}(e^s - 1)\,|\phi(s)|\, ds < \infty.\end{aligned}$$

Hence $G \in \mathscr{D}'(\mathbf{R}) \otimes \mathscr{Z}'(\mathbf{R})$ and $\operatorname{supp}(G) \subset \mathbf{R}_+ \times \mathbf{R}$. To verify (2) let

$$g_k(s,t) = 4(-1)^k[(2k+1)!]^{-1} \times \Theta(s) e^{-(a+u_1)s} s^{2k+1} e^{-i(u_2+b)t} \delta^{(2k)}(t).$$

Then $g_k(s,t)$ is, for each $k \in \mathbf{N}_0$, an integrable distribution, and

$$\int_{\mathbf{R}^2} g_k(s,t) d(s,t) = 4(u_1+a)^{-2}[(u_2+b)/(u_1+a)]^{2k}.$$

Since

$$u_2 + b = p^2 - \mu - (q^2 - \lambda) < p^2 + \mu + q^2 + \lambda = u_1 + a,$$

one has $|u_2 + b| < u_1 + a$. Thus

$$\int_{\mathbf{R}^2} G(s,t) e^{-u_1 s - i u_2 t} d(s,t)$$

$$= \lim_{N \to \infty} \sum_{0 \leqslant k \leqslant N} \int_{\mathbf{R}^2} g_k(s,t)\, d(s,t)$$

$$= 4(u_1+a)^{-2} \sum_{0 \leqslant k} \left[\frac{u_2+b}{u_1+a}\right]^{2k}$$

$$= 4(u_1+a)^{-2}\{1 - [(u_2+b)/(u_1+a)]^2\}^{-1}$$

$$= 4[(u_1+a+u_2+b)(u_1+a-u_2-b)]^{-1}$$

$$= 4[(2\lambda + u_1 + u_2)(2\mu + u_1 - u_2)]^{-1}$$

$$= [(\lambda + p^2)(\mu + q^2)]^{-1} = f(p^2, q^2).$$

This proves the Lemma.

[1]E. Teubner, N. Grün, and W. Scheid, Phys. Lett. A **126**, 103 (1987).
[2]I. M. Gelfand and G. E. Shilov, *Generalized Functions I* (Academic, New York, 1964).
[3]I. M. Gelfand and G. E. Shilov, *Generalized Fucntions II* (Academic, New York, 1964).

Nonperturbative solutions of nonlinear differential equations using continued fractions[a)]

William L. Ditto and Thomas J. Pickett
Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634-1911

(Received 9 September 1987; accepted for publication 30 March 1988)

A new approach to the solution of nonlinear differential equations of mathematical physics is presented. Continued fractions are exploited to convert a power series of a solution into a closed form expression that yields an excellent approximation to the exact solution. These solutions contain the appropriate number of arbitrary constants to accommodate boundary conditions. The method is also shown to be capable of generating certain exact solutions. Evidence is provided for the conjecture that the known exact solutions are members of families of exact solutions.

I. INTRODUCTION

Nonlinear differential equations play a central role in modern theoretical physics[1]; yet the solution of these equations is one of the most difficult problems for the mathematical physicist. Numerical methods are difficult to validate and at best provide numerical solutions, containing no arbitrary constants. One is faced with a multitude of *ad hoc* methods for constructing solutions in closed form. One would like to systematically generate in closed form solutions that contain an appropriate number of arbitrary constants. Various techniques such as WBK or saddle-point methods are useful for first approximations but successive approximations are extremely complex. Recent work[2,3] has been pursued using continued fractions in a nonperturbative iterative technique for generating approximate solutions of nonlinear field equations. It has been shown that in at least one case[3] the continued fraction can be summed to obtain an exact solution. The purpose of this paper is to extend this technique and present some new results on physically interesting nonlinear differential equations.

One can construct a solution in the form of a power series with little difficulty, but power series in general represent a solution only in a small neighborhood. A possible way to extend the neighborhood is to use Padé approximants.[4] In this paper, we introduce an alternative to the Padé approximant method that extends the domain over which the function adequately approximates a solution. This alternative exploits continued fractions to convert a power series into a function that represents the solution faithfully over a larger domain. The new function is found to be a very good approximation to the solution of the differential equation. In addition, the function contains one or more arbitrary constants which can then be fixed by boundary conditions.

In Sec. II, we describe the method by which continued fractions convert power series into a function with a larger radius of convergence. In Secs. III and IV, we construct approximations to two known exact solutions of a nonlinear differential equation using the continued fraction technique. We discuss the accuracy of the method and the relation of the boundary conditions to the arbitrary constants. In particular, we show that the currently known exact solutions are members of families of solutions that differ from one another by the value of an arbitrary constant. In Sec. V, we use this technique to construct an exact particular solution to a nonlinear differential equation.

II. THE METHOD

In this section, we show how to use the information provided by a power series to construct a function with a larger domain of convergence. For this we rely heavily on Khovanskii.[5] Consider the power series $Y=\Sigma_{n=0}^{\infty}a_nx^n$. Then

$$\frac{d}{dx}\ln Y=\frac{Y'}{Y}=\sum_{n=0}^{\infty}na_nx^{n-1}\left(\sum_{n=0}^{\infty}a_nx^n\right)^{-1}, \quad (2.1)$$

which is just the ratio of two power series. Thus, formally, $Y=\exp[\int dx\, Y'/Y]$. One could expand the integrand in powers of x to integrate, but this would only reproduce the original series representation for Y. However, we can truncate the series for Y and expand Y'/Y as a continued fraction, which when summed and integrated yields an approximation for Y that is more accurate than the original truncated power series. Let

$$f(x)=\frac{a_{10}+a_{11}x+a_{12}x^2+a_{13}x^3+\cdots}{a_{00}+a_{01}x+a_{02}x^2+a_{03}x^3+\cdots}=\cfrac{1}{\dfrac{a_{00}}{a_{10}}+\dfrac{a_{00}+a_{01}x+a_{02}x^2+\cdots}{a_{10}+a_{11}x+a_{12}x^2+\cdots}-\dfrac{a_{00}}{a_{10}}} \quad (2.2)$$

[a)] A portion of this work was presented at the Lepton–Photon conference, Hamburg, West Germany, 1987.

 0022-2488/88/081761-10$02.50

$$= \cfrac{a_{10}}{a_{00} + x\cfrac{(a_{10}a_{01} - a_{00}a_{11}) + (a_{10}a_{02} - a_{00}a_{12})x + \cdots}{a_{10} + a_{11}x + a_{12}x^2 + \cdots}}$$

$$= \cfrac{a_{10}}{a_{00} + x\cfrac{a_{20} + a_{21}x + a_{22}x^2 + \cdots}{a_{10} + a_{11}x + a_{12}x^2 + \cdots}}$$

$$= \cfrac{a_{10}}{a_{00} + x\cfrac{a_{20}}{a_{10} + x\cfrac{(a_{20}a_{11} - a_{10}a_{21}) + (a_{20}a_{21} - a_{10}a_{22})x + \cdots}{a_{20} + a_{21}x + a_{22}x^2 + \cdots}}}$$

$$= \cfrac{a_{10}}{a_{00} + \cfrac{a_{20}\,x}{a_{10} + \cfrac{a_{30}\,x}{a_{20} + \cdots}}} \equiv \frac{a_{10}}{a_{00}\,+}\;\frac{a_{20}\,x}{a_{10}\,+}\;\frac{a_{30}\,x}{a_{20}\,+}\;\cdots, \tag{2.3}$$

where we are now using the lowered plus sign to denote a continued fraction. The computations are conveniently set out in the following scheme:

$$\begin{matrix} a_{00} & a_{01} & a_{02} & \cdots \\ a_{10} & a_{11} & a_{12} & \cdots \\ a_{20} & a_{21} & a_{22} & \cdots \\ a_{30} & a_{31} & a_{32} & \cdots \\ \vdots & \vdots & \vdots & \end{matrix} \;. \tag{2.4}$$

Here $a_{mn} = a_{m-1,0}a_{m-2,n+1} - a_{m-2,0}a_{m-1,n+1}$. Thus a particular entry in a given row is obtained by cross multiplying the two previous entries in the column to the immediate right with the two previous entries in the first column. As an example consider

$$f(x) = \frac{e^x}{e^x - 1} = \frac{1 + x + \frac{1}{2}x^2 + \cdots}{x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots}$$

$$= \frac{x^{-1}[1 + x + \frac{1}{2}x^2 + \cdots]}{1 + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots}. \tag{2.5}$$

The table for $xf(x)$ is

$$\begin{matrix} 1 & \frac{1}{2} & \frac{1}{6} & \frac{1}{24} & \cdots \\ 1 & 1 & \frac{1}{2} & \frac{1}{6} & \cdots \\ -\frac{1}{2} & -\frac{1}{3} & -\frac{1}{8} & \cdots & \\ -\frac{1}{6} & -\frac{1}{8} & \cdots & & \\ -\frac{1}{144} & \cdots & & & \end{matrix} \;. \tag{2.6}$$

Therefore

$$f(x) = x^{-1}\left[\frac{1}{1\,+}\;\frac{-\frac{1}{2}x}{1\,+}\;\frac{-\frac{1}{6}x}{-\frac{1}{2}\,+}\;\frac{-\frac{1}{144}x}{-\frac{1}{6}}\cdots\right]$$

$$= \frac{x^{-1}}{1\,+}\;\frac{-\frac{1}{2}x}{1\,+}\;\frac{+\frac{1}{3}x}{1\,+}\;\frac{-\frac{1}{12}x}{1\,+}\;\cdots. \tag{2.7}$$

If $a_{k,0} = 0$ for $k > 1$, then this scheme gives division by zero. In this case, a modification of the method is required. This is illustrated by the following example. Let

$$f(x) = \frac{\sin(x)}{\cos(x)} = \frac{x[1 - \frac{1}{6}x^2 + \frac{1}{120}x^4 + \cdots]}{1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 + \cdots}. \tag{2.8}$$

The table for $x^{-1}f(x)$ begins

$$\begin{matrix} 1 & 0 & -\frac{1}{2} & 0 & \frac{1}{24} & 0 & \cdots \\ 1 & 0 & -\frac{1}{6} & 0 & \frac{1}{120} & 0 & \cdots \\ 0 & -\frac{1}{3} & 0 & \frac{1}{30} & 0 & \cdots & \end{matrix} \;. \tag{2.9}$$

We modify the scheme by deleting the offending zero and shifting the entire row to the left. The table becomes

$$\begin{matrix} 1 & 0 & -\frac{1}{2} & 0 & \cdots \\ 1 & 0 & -\frac{1}{6} & 0 & \cdots \\ -\frac{1}{3} & 0 & \frac{1}{30} & 0 & \cdots \end{matrix} \;. \tag{2.10}$$

This shift is accompanied by an increase in the power of x for the coefficient in that row. Hence the continued fraction for $f(x)$ will begin

$$x\left[\frac{1}{1\,+}\;\frac{-\frac{1}{3}x^2}{1\,+}\;\cdots\right].$$

If there are two or more zeros in the leftmost columns they are all deleted, the row is shifted appropriately, and the power of x is increased by 1 for each zero. In the present example the table becomes

$$\begin{matrix} 1 & 0 & -\frac{1}{2} & 0 & \cdots \\ 1 & 0 & -\frac{1}{6} & 0 & \cdots \\ -\frac{1}{3} & 0 & \frac{1}{30} & 0 & \cdots \\ \frac{1}{45} & 0 & \cdots & & \\ \vdots & \vdots & & & \end{matrix} \;, \tag{2.11}$$

where the last row has also been shifted. Therefore

$$f(x) = \tan(x) = \frac{x}{1\,+}\;\frac{-\frac{1}{3}x^2}{1\,+}\;\frac{\frac{1}{45}x^2}{-\frac{1}{3}\,+}\;\cdots$$

$$= \frac{x}{1\,+}\;\frac{-\frac{1}{3}x^2}{1\,+}\;\frac{-\frac{1}{15}x^2}{1\,+}\;\cdots. \tag{2.12}$$

Another possibility is when all the entries of a row vanish. Then the continued fraction terminates. For example, let

$$f(x) = \frac{\lambda + 2\lambda^2 x + 3\lambda^3 x^2 + \cdots}{1 + \lambda x + \lambda^2 x^2 + \lambda^3 x^3 + \cdots}. \qquad (2.13)$$

Note that $f(x) = g'(x)/g(x)$, with

$$g(x) = \sum_{n=0}^{\infty} (\lambda x)^n = \frac{1}{1-\lambda x}.$$

Constructing the table for $f(x)$ gives

$$\begin{array}{ccccccc} 1 & \lambda & \lambda^2 & \lambda^3 & \cdots & & \lambda^n \\ \lambda & 2\lambda^2 & 3\lambda^3 & 4\lambda^4 & \cdots & & (n+1)\lambda^{n+1} \\ -\lambda^2 & -2\lambda^3 & -3\lambda^4 & \cdots & & & -(n+1)\lambda^{n+2} \\ 0 & 0 & 0 & \cdots & & & [(n+1)-(n+1)]\lambda^{n+4} \equiv 0 \end{array}. \qquad (2.14)$$

So

$$f(x) = \frac{\lambda}{1} + \frac{-\lambda^2 x}{\lambda} = \frac{\lambda}{1-\lambda x}. \qquad (2.15)$$

That this is the correct value for $f(x)$ can be seen by taking the derivative of $g(x)$:

$$g'(x) = \frac{\lambda}{(1-\lambda x)^2}$$

and

$$\frac{g'(x)}{g(x)} = \frac{\lambda}{(1-\lambda x)^2} \frac{(1-\lambda x)}{1} = \frac{\lambda}{1-\lambda x}.$$

III. EXAMPLE

We now show how to use these ideas to construct approximate solutions to nonlinear differential equations. As an example, consider a nonlinear generalization[1] of the Klein–Gordon equation,

$$\check{\partial}^2 \phi + m^2 \phi - \lambda \phi^3 = 0. \qquad (3.1)$$

Assume $\phi = \phi(x)$, $x = \pm i\check{k}\cdot\check{x}$, where $\check{k} = (k^\circ, \vec{k})$ and $\check{k}^2 = m^2$. Here an inverted caret denotes a four-vector. This assumption leads to the following equation for ϕ:

$$\phi'' - \phi + (\lambda/m^2)\phi^3 = 0, \qquad (3.2)$$

where the primes denote differentiation with respect to x. Rescaling the dependent variable $Y = (\lambda/m^2)^{1/2}\phi$ leads to the following equation for Y:

$$Y'' - Y + Y^3 = 0. \qquad (3.3)$$

Now let $W = Y'/Y$. Then $Y = \exp[\int W\,dx]$ and

$$Y(W' + W^2) - Y + Y^3 = 0$$

or

$$W' + W^2 - 1 + \exp\left[2\int W\,dx\right] = 0. \qquad (3.4)$$

This is a generalization of the Ricatti equation and is the starting point of the investigation.[2,3,6] This equation can be solved iteratively by choosing W in the integrand, resulting in a first-order differential equation. This produces approximate (and possibly exact) solutions.[3] The method presented here consists of inserting into the integrand a form for W containing undetermined constants that will be fixed by the resulting linear equation. The choice for the integrand is arbitrary; we have found power series used in conjunction with continued fractions to be useful. Calculationally there are two ways to proceed. We shall illustrate these two approaches in the following two examples. In the first example, we substitute for W a continued fraction expansion with undetermined coefficients. In the second, we use a power series to accomplish the same result and only introduce the continued fraction at the end of the calculation. The power series has the advantage of calculational simplicity; the continued fraction has the advantage that there exist exact solutions for which W is a finite continued fraction, but as a power series it contains an infinite number of terms. An example of this is shown in Sec. V.

We will construct a solution to Eq. (3.3) that approximates the known exact solution $Y = (\sqrt{2})\text{sech}(x)$. Let

$$W = \frac{a_0 x}{1} + \frac{a_1 x^2}{1} + \frac{a_2 x^2}{1} + \frac{a_3 x^2}{1} + \frac{a_4 x^2}{1} \qquad (3.5)$$

$$= \frac{a_0 x[1 + (a_2 + a_3 + a_4)x^2 + a_2 a_4 x^4]}{1 + (a_1 + a_2 + a_3 + a_4)x^2 + (a_2 a_4 + a_1 a_4 + a_1 a_3)x^4}, \qquad (3.6)$$

with the a_i's as yet undetermined.

We choose this form for W since we know $Y = (\sqrt{2})\text{sech}(x)$ is an even function of x, so $W = Y'/Y$ is odd. This integrates to

$$\exp\left[2\int \frac{a_0 x[1 + (a_2 + a_3 + a_4)x^2 + a_2 a_4 x^4]}{1 + (a_1 + a_2 + a_3 + a_4)x^2 + (a_2 a_4 + a_1 a_4 + a_1 a_3)x^4}\,dx\right]$$

$$= C_0 \left[\frac{[2A/(B-E)]x^2 + 1}{[2A/(B+E)]x^2 + 1}\right]^{(a_0[A(2A - BC) + D(B^2 - 2A)])/2A^2E} \cdot [Ax^4 + Bx^2 + 1]^{[(AC - BD)/2A^2]a_0} \exp((a_0 D/A)x^2), \qquad (3.7)$$

where

$$A = a_1 a_3 + a_1 a_4 + a_2 a_4, \quad B = a_1 + a_2 + a_3 + a_4, \quad C = B - a_1, \quad D = a_2 a_4, \quad E = (B^2 - 4A)^{1/2},$$

and C_0 is a constant of integration. Expanding this function in powers of x and inserting it into the differential equation (3.3) while taking Y to be of the form $\Sigma_{n \text{ even}} b_n x^n$ gives

$$0 = \sum_{n \text{ even}}^{\infty} b_n \{ n(n-1)x^{n-2} - x^n [1 - C_0 \{ 1 + a_0 x^2 + \tfrac{1}{2} a_0 (a_0 - a_1) x^4 + (\tfrac{1}{6} a_0^3 - \tfrac{1}{2} a_0 a_1 + \tfrac{1}{3} a_0 a_1 (a_1 + a_2)) x^6$$
$$+ (\tfrac{1}{24} a_0^4 - \tfrac{1}{4} a_0^3 a_1 + \tfrac{1}{8} a_0^2 a_1^2 + \tfrac{1}{3} a_0^2 a_1 (a_1 + a_2) - \tfrac{1}{4} a_0 a_1 [(a_1 + a_2)^2 + a_2 a_3]) x^8 + \cdots \}] \} .$$

Solving for the b_n's in the usual way gives

$$b_0 \text{ is arbitrary}, \quad b_2 = \tfrac{1}{2} b_0 (1 - C_0) \cdots .$$

At this point it will be convenient to fix the constant C_0 in terms of the a_i's in the continued fraction. We do this with the help of the algorithm in Sec. II. We have at this point

$$Y = b_0 + \tfrac{1}{2} b_0 (1 - C_0) x^2 + \cdots$$

and

$$y' = b_0 (1 - C_0) x + \cdots ,$$

so that

$$W = \frac{Y'}{Y} = x \cdot \frac{b_0(1 - C_0) + \cdots}{b_0 + \tfrac{1}{2} b_0 (1 - C_0) x^2 + \cdots} .$$

Constructing the table, we have

$$\begin{array}{llll} b_0 & 0 & \tfrac{1}{2} b_0 (1 - C_0) & \cdots \\ b_0 (1 - C_0) & 0 & \cdots & \end{array} ,$$

so that

$$W = x \cdot \frac{b_0 (1 - C_0)/b_0}{1 \quad + \cdots} = \frac{(1 - C_0) x}{1 \quad + \cdots} .$$

Thus we see that $1 - C_0 = a_0$ is the choice that makes the series solution of the differential equation consistent with W as in Eq. (3.5). We can now write the results of Eq. (3.7) as

b_0 is arbitrary—for simplicity take $b_0 = 1$,

$$b_2 = \tfrac{1}{2} a_0 ,$$
$$b_4 = (a_0/24)(3a_0 - 2) ,$$
$$b_6 = (a_0/6!) [27 a_0^2 - 30 a_0 + 4] ,$$
$$b_8 = (a_0/8!) [441 a_0^3 - 684 a_0^2 + 252 a_0 - 8] ,$$
$$b_{10} = \frac{a_0}{90} \Big[b_8 + (a_0 - 1) \Big\{ b_6 + \tfrac{1}{2} (a_0 - a_1) b_4 + (\tfrac{1}{6} a_0^2 - \tfrac{1}{2} a_0 a_1 + \tfrac{1}{3} a_1 (a_1 + a_2)) b_2$$
$$+ \frac{a_0^3}{24} - \frac{a_0^2 a_1}{4} + \frac{a_0 a_1}{8} - \frac{a_1}{4} ((a_1 + a_2)^2 + a_2 a_3) \Big\} \Big] + (a_0 - 1) \frac{a_0^2 a_1}{270} (a_1 + a_2) .$$

From this power series we construct the continued fraction for Y'/Y as in Sec. II. We begin by creating the table (2.4):

$$\begin{array}{lllllllllll}
1 & 0 & \frac{a_0}{2} & 0 & \frac{a_0}{24}(3a_0 - 2) & 0 & \frac{a_0}{720}(27a_0^2 - 30a_0 + 4) & 0 & \frac{a_0}{40320}(R_1) & 0 & b_{10} \quad \cdots \\
a_0 & 0 & \frac{a_0}{6}(3a_0 - 2) & 0 & \frac{a_0}{120}(27a_0^2 - 30a_0 + 4) & 0 & \frac{a_0}{5040}(R_1) & 0 & P_1 & & \cdots \\
\frac{a_0}{3} & 0 & \frac{a_0}{60}(-6a_0^2 + 10a_0 - 2) & 0 & \frac{a_0}{2520}(R_2) & 0 & P_2 & & & & \\
\frac{a_0^2}{90}(9a_0^2 - 7) & 0 & \frac{a_0^2}{1260}(R_3) & 0 & P_3 & & & & & & \\
\frac{a_0^3}{18900}(R_4) & 0 & P_4 & & \cdots & & & & & & \\
P_5 & \cdots & & & & & & & & &
\end{array} ,$$

where the P_i's and R_i's have been abbreviated for convenience. They are

$$P_1 = 10 b_{10} ,$$
$$P_2 = (a_0^2/8!) [441 a_0^3 - 684 a_0^2 + 252 a_0 - 8] - P_1 ,$$
$$P_3 = (a_0^2/3 \cdot 7!) [441 a_0^3 - 684 a_0^2 + 252 a_0 - 8] - a_0 P_2 ,$$

$$P_4 = (a_0^3/7! \cdot 45)[9a_0^2 - 7][-126a_0^3 + 237a_0^2 - 112a_0 + 4] - (a_0/3)P_3,$$

$$P_5 = (a_0^5/1260 \cdot 18\,900)[-189a_0^4 + 204a_0^2 - 11][63a_0^3 - 24a_0^2 - 49a_0 + 12] - (a_0^2/90)(9a_0^2 - 7)P_4,$$

$$R_1 = (441a_0^3 - 684a_0^2 + 252a_0 - 8),$$

$$R_2 = (-126a_0^3 + 237a_0^2 - 112a_0 + 4),$$

$$R_3 = (63a_0^3 - 24a_0^2 - 49a_0 + 12),$$

$$R_4 = (-189a_0^4 + 204a_0^2 - 11).$$

In this table all rows except the first two begin with a zero; we have already deleted the zero and shifted the columns for these rows. Thus the continued fraction for Y'/Y is

$$\frac{Y'}{Y} = \frac{a_0 x}{1} + \frac{(a_0/3)x^2}{a_0} + \frac{(a_0/90)(9a_0^2 - 7)x^2}{(a_0/3)} + \frac{(a_0^3/18\,900)(-189a_0^4 + 204a_0^2 - 11)x^2}{(a_0^2/90)(9a_0^2 - 7)} + \frac{P_5 x^2}{(a_0^3/18\,900)(-189a_0^4 + 204a_0^2 - 11) + \cdots},$$

$$\frac{Y'}{Y} = \frac{a_0 x}{1} + \frac{\frac{1}{3}x^2}{1} + \frac{\frac{1}{30}(9a_0^2 - 7)x^2}{1} + \frac{\frac{1}{70}[(-189a_0^4 + 204a_0^2 - 11)/(9a_0^2 - 7)]x^2}{1} + \frac{[(90)(18\,900)P_5/a_0^5(9a_0^2 - 7)(-189a_0^4 + 204a_0 - 11)]x^2}{1} + \cdots.$$

We have now fixed the constants a_1 through a_4 in Eq. (3.5) as functions of a_0. This a_0 is an arbitrary constant that may be used to meet boundary conditions [$a_0 \neq 0$ and a_0 must be chosen such that E in Eq. (3.7) does not become imaginary]. To find the solution to the differential equation we must evaluate $Y = \exp \int Y'/Y\, dx$. This has already been done in Eq. (3.7). Therefore we have, using $C_0 = 1 - a_0$,

$$Y = (1 - a_0)^{1/2} \left[\frac{[2A/(B-E)]x^2 + 1}{[2A/(B+E)]x^2 + 1} \right]^{(a_0[A(2A - BC) + D(B^2 - 2A)])/4A^2E} (Ax^4 + Bx^2 + 1)^{[(AC - BD)/4A^2]a_0} \exp(a_0 Dx^2/2A), \tag{3.8}$$

with the same identifications for A through E.

For some values of a_0 this is an approximate solution to (3.3). This region includes $a_0 = -1$, for at this value we recognize Y'/Y as the first five terms of the continued fraction expansion of $-\tanh x$, which gives the solution $Y = (\sqrt{2})\mathrm{sech}(x)$. This is an exact solution to Eq. (3.3). For $a_0 = -1$ the various abbreviations become

$$A = \tfrac{1}{63}, \quad B = \tfrac{4}{9}, \quad C = \tfrac{1}{5},$$
$$D = \tfrac{1}{945}, \quad E = \tfrac{2}{63}(133)^{1/2}.$$

With these values the approximate solution is

$$Y = (\sqrt{2}) \left[\frac{x^2/[14 - (133)^{1/2}] + 1}{x^2/[14 + (133)^{1/2}] + 1} \right]^{49/15(133)^{1/2}} \times \left[\frac{x^4}{63} + \frac{4x^2}{9} + 1 \right]^{-77/60} e^{-x^2/30}. \tag{3.9}$$

This function agrees remarkably well with $(\sqrt{2})\mathrm{sech}(x)$. Figure 1 compares the exact solution with the approximation while Fig. 2 shows the absolute error $|\sqrt{2}\,\mathrm{sech}(x) - Y|$. We see that Y is a very good approximation to the exact solution. For other values of a_0 Eq. (3.8) yields approximate solutions to Eq. (3.3), even though no such exact solutions are yet known. In the next section we will give justification for the conjecture that such exact solutions exist and discuss more fully the role of the arbitrary constant(s) appearing in these solutions.

IV. ANOTHER EXAMPLE

We now construct another approximate solution to Eq. (3.3) using an equivalent, but calculationally simpler, approach. The continued fraction technique as described in Sec. II is employed towards the end of the calculation after determining the series coefficients simply by substituting the power series into Eq. (3.3). Consider again Eq. (3.3),

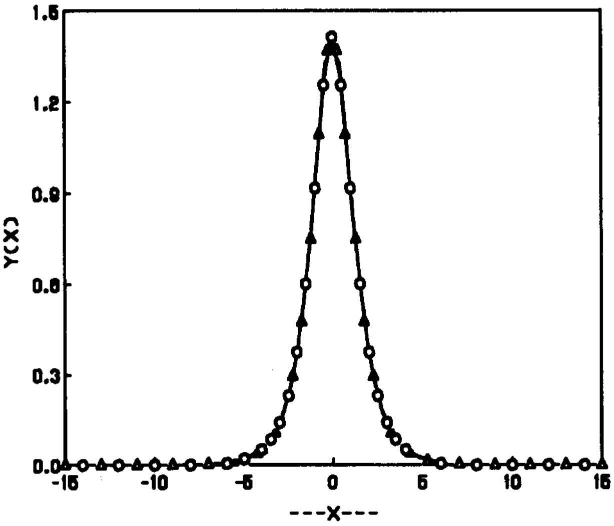


FIG. 1. Graph of the exact $[Y(x) = (2)^{1/2}\,\mathrm{sech}(x)]$ and approximate (3.9) solutions to $Y'' - Y(1 - Y^2) = 0$. Δ: exact solutions; O: approximate solution.

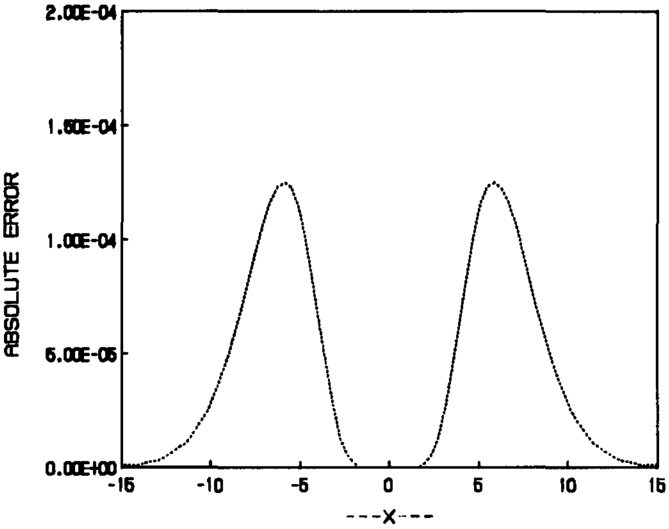


FIG. 2. Graph of the absolute error between the exact and approximate solutions in Fig. 1 versus x.

$$0 = Y'' - Y(1 - Y^2)\,.$$

Assume

$$Y = \sum_{\substack{n=0 \\ n \text{ odd}}}^{\infty} a_n x^n \simeq a_1 x + a_3 x^3 + a_5 x^5 + a_7 x^7\,. \tag{4.1}$$

Then

$$Y^3 \simeq a_1^3 x^3 + 3a_1^2 a_3 x^5 + (3a_1^2 a_5 + 3a_1 a_3^2) x^7\,. \tag{4.2}$$

Equation (3.3) then becomes

$$\begin{aligned} 0 = {} & (6a_3 x + 20a_5 x^3 + 42a_7 x^5 + 72a_9 x^7 + \cdots) \\ & - (a_1 x + a_3 x^3 + a_5 x^5 + a_7 x^7 + \cdots) \\ & + (a_1^3 x^3 + 3a_1^2 a_3 x^5 + (3a_1^2 a_5 + 3a_1 a_3^2) x^7) \\ & + (3a_1^2 a_7 + a_3^3 + 6a_1 a_3 a_5)\,. \end{aligned}$$

Grouping together like powers of x one finds

$$\begin{aligned} & (6a_3 - a_1)x + (20a_5 - a_3 + a_1^3)x^3 \\ & \quad + (42a_7 - a_5 + 3a_1^2 a_3)x^5 \\ & \quad + (72a_9 - a_7 + (3a_1^2 a_5 + 3a_1 a_3^2))x^7 = 0\,. \end{aligned}$$

Equating like powers of x to zero gives

$$\begin{aligned} 0 &= 6a_3 - a_1\,, \\ 0 &= 20a_5 - a_3 + a_1^3\,, \\ 0 &= 42a_7 - a_5 + 3a_1^2 a_3\,, \\ 0 &= 72a_9 - a_7 + (3a_1^2 a_5 + 3a_1 a_3^2)\,, \end{aligned}$$

yielding

$$\begin{aligned} a_3 &= (1/3!)a_1\,, \\ a_5 &= (a_1/5!)[1 - 6a_1^2]\,, \\ a_7 &= (a_1/7!)[1 - 66a_1^2]\,, \\ a_9 &= (a_1/9!)[1 - 612a_1^2 + 756a_1^4]\,. \end{aligned} \tag{4.3}$$

We now have the first five terms of the power series representation of a particular solution to Eq. (3.3). Note that it contains one arbitrary constant a_1. We now employ the method of Sec. II to improve the convergence properties of Y. One could use Padé approximants to do this if one could determine which approximant is appropriate. Our method has the advantage of providing a unique approximant for each y, and successive approximations converge to the actual solution.

From (4.1),

$$Y' = a_1 + 3a_3 x^2 + 5a_5 x^4 + 7a_7 x^6 + 9a_9 x^8\,.$$

Therefore

$$\frac{Y'}{Y} = x^{-1}\left[\frac{1 + 3a_3 x^2 + 5a_5 x^4 + 7a_7 x^6 + 9a_9 x^8}{1 + a_3 x^2 + a_5 x^4 + a_7 x^6 + a_9 x^8}\right], \tag{4.4}$$

where all the a_n's have been redefined such that $a_n \to a_n/a_1$. Forming the table as in Sec. II we have

1	0	a_3	0	a_5	0	a_7	0	a_9
1	0	$3a_3$	0	$5a_5$	0	$7a_7$	0	$9a_9$
$-2a_1$	0	$-4a_5$	0	$-6a_7$	0	$-8a_9$	—	—
$-6a_3^2 + 4a_5$	0	$6a_7 - 10a_3 a_5$	0	$8a_9 - 14a_3 a_7$	—	—	—	— ,
M_1	0	M_2	0	—	—	—	—	—
M_3	—	—	—	—	—	—	—	—

where

$$\begin{aligned} M_1 &= 4[a_3^2 a_5 - 4a_5^2 + 3a_3 a_7]\,, \\ M_2 &= 8[a_3^2 a_7 - 3a_5 a_7 + 2a_3 a_9]\,, \\ M_3 &= 48a_3^4 a_7 + 160a_3 a_5^3 - 40a_3^3 a_5^2 + 96a_3^3 a_9 - 64a_3 a_5 a_9 + 72a_3 a_7^2 - 272a_3^2 a_5 a_7\,. \end{aligned}$$

All rows in the table, except the first two, have the leading zero deleted and have been shifted one column to the left. Thus the coefficients provided by these rows will multiply x^2 rather than x in the continued fraction. From Eqs. (4.3) it is seen that all of these expressions are actually expressions involving a_1 only; we have written them in this form for convenience. Thus

$$\frac{Y'}{Y} = \frac{x^{-1}}{1\,+}\;\frac{-2a_3 x^2}{1\,+}\;\frac{(4a_5 - 6a_3^2)x^2}{-2a_3\,+}\;\frac{M_1 x^2}{(4a_5 - 6a_3^2)\,+}\;\frac{M_3 x^2}{M_1\,+}\;\cdots\,. \tag{4.5}$$

We define

$$A\equiv\frac{6a_3^2-4a_5}{2a_3}+\frac{M_1}{2a_3(6a_3^2-4a_5)}+\frac{M_3}{M_1(4a_5-6a_3^2)}$$

$$B\equiv -M_3/2a_3M_1$$

$$C\equiv A-2a_3$$

$$D\equiv -\frac{M_3}{2a_3M_1}-2a_3\left(\frac{M_1}{2a_3(6a_3^2-4a_5)}+\frac{M_3}{M_1(4a_5-6a_3^2)}\right).$$

Thus contracting Eq. (4.5) we see

$$Y'/Y=x^{-1}[1+Ax^2+Bx^4]/(1+Cx^2+Dx^4),$$

$$Y=\exp\int\frac{Y'}{Y}=C_0\,x[1+Cx^2+Dx^4]^{(B-D)/4D}\left[\frac{2Dx^2/[C-(C^2-4D)^{1/2}]+1}{2Dx^2/[C+(C^2-4D)^{1/2}]+1}\right]^{[2AD-C(B+D)]/4D(C^2-4D)^{1/2}}, \tag{4.6}$$

for $C^2-4D>0$.

Here C_0 is a constant of integration, but is not arbitrary; it must be chosen so that the above expression for Y is consistent with (4.1). To do this we note that both expressions are expansions about $x=0$. We see that, in Eq. (4.1),

$$\left.\frac{Y}{x}\right|_{x=0}=a_1,$$

while in Eq. (4.6),

$$\left.\frac{Y}{x}\right|_{x=0}=C_0.$$

Therefore we must choose $C_0=a_1$. [This is equivalent to demanding that Eq. (4.6) and Eq. (4.1) have the same value for the derivative at $x=0$.] The numerical value of a_1 (and hence A, B, C, D) will be determined by a boundary condition fixing $Y'(0)$. For example, suppose we choose $a_1=i/\sqrt{2}$. Then Eq. (4.5) becomes

$$\frac{Y'}{Y}=\frac{x^{-1}}{1\,+}\,\frac{-x^2}{3\,+}\,\frac{3x^2}{10\,+}\,\frac{-11x^2}{21\,+}\,\frac{25x^2}{66}, \tag{4.7}$$

and

$$A=\tfrac{13}{198},\quad B=\tfrac{5}{2772},\quad C=\tfrac{-53}{198},\quad D=\tfrac{551}{41580}.$$

With these values Eq. (4.6) becomes a very good approximation to $Y=i\tan(x/\sqrt{2})$, which is an exact solution to Eq. (3.3):

$$Y=\frac{i}{\sqrt{2}}x\left(1-\tfrac{53}{198}x^2+\tfrac{551}{41580}x^4\right)^{-119/551}\left[\frac{-551x^2/20790\left[\tfrac{53}{198}+\left(\tfrac{25583}{1372140}\right)^{1/2}\right]+1}{551x^2/20790\left[-\tfrac{53}{198}+\left(\tfrac{25583}{1372140}\right)^{1/2}\right]+1}\right]^{\frac{5938}{54549}\left(\frac{1372140}{25583}\right)^{1/2}}. \tag{4.8}$$

Figure 3 shows a comparison of the exact solution $i\tan(x/\sqrt{2})$ to Eq. (4.8), as well as the absolute error, $|\tan(x/\sqrt{2})-Y/i|$. Again, the agreement is remarkable. The function in Eq. (4.8) can be differentiated twice and substituted into Eq. (3.3) to see how closely the left-hand side approximates zero. This function, $Y''-Y+Y^3$, we call the error function for Eq. (4.8), and it is plotted in Fig. 4. This function is difficult to interpret, since one might expect the error function to be of the same order of magnitude as the absolute error. This is generally not true, even for very good approximations. Here, for example, at the point $x=1$, the absolute error is $\sim 2\times10^{-7}$, while the error function is ~ 3. The error function is important in investigating solutions that cannot be expressed in terms of elementary functions. In Eq. (4.6), if we let $a_1=\sqrt{-0.4}$ rather than $\sqrt{-0.5}$ we obtain another approximate solution to Eq. (3.3). This function, as well as its error function, is shown in Fig. 5. Comparison with Fig. 4 shows that although the qualitative behavior is the same for each, the function in Fig. 5 approximates a solution to Eq. (3.3) even better than the function in Fig. 4 approximates the solution $Y=i\tan(x/\sqrt{2})$. Thus we believe Fig. 5 shows a plot of another exact solution of Eq. (3.3). Figures 6–9 show similar results for various values of a_1.

Equation (3.3) has no solutions of the tangent type other than $Y=i\tan(x/\sqrt{2})$. Thus the solutions in Figs. 5–9 cannot be of the form $Y=A\tan(BX)$, A,B constants, yet they are members of a family of solutions with a continuous parameter a_1 that includes the tangent for $a_1=i/\sqrt{2}$.

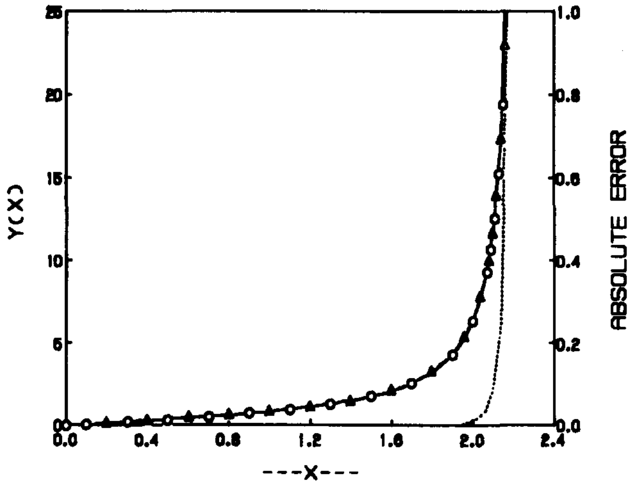


FIG. 3. Graph of the exact $[Y(x)=i\tan(x/(2)^{1/2})]$ and approximate (4.8) solutions (solid lines) to $Y''-Y(1-Y^2)=0$ and the absolute error (dotted line). Δ: exact solution; O: approximate solution.

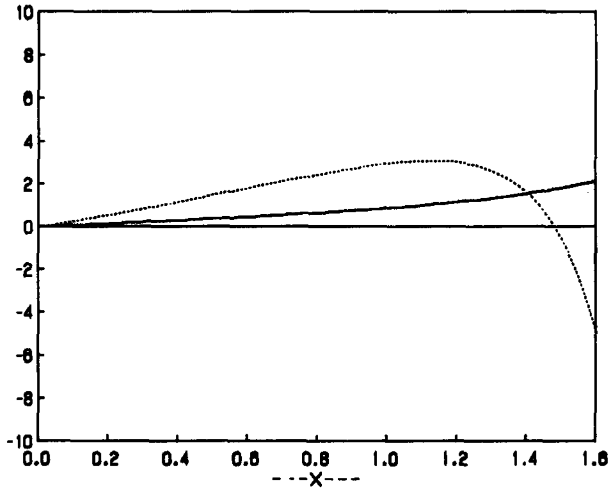


FIG. 4. Graph of $Y(x)/i$ and the error function $(Y'' - Y(1 - Y^2))/i$ for (4.6) $(a_1 = \sqrt{-0.5})$. Dotted line: error function; solid line: Y/i.

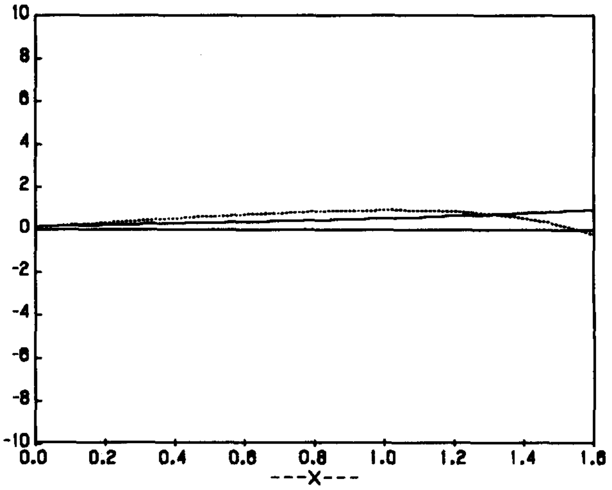


FIG. 6. Graph of $Y(x)/i$ and the error function $(Y'' - Y(1 - Y^2))/i$ for (4.6) $(a_1 = \sqrt{-0.1})$. Dotted line: error function; solid line: Y/i.

V. EXACT SOLUTION

In this section we show how to apply our method to construct an exact solution (no approximation) to a particular nonlinear differential equation. This is possible because the continued fraction representing Y'/Y terminates for this solution. Consider the nonlinear generalization of the Klein–Gordon equation,

$$\partial_\mu \partial^\mu \eta + m^2 \eta + \lambda \eta^3 = 0 . \tag{5.1}$$

Let $\eta = \eta(x)$, $x = e^{\pm i\check{p}\cdot\check{x}}$, and $\check{p}^2 = m^2$. Thus

$$x^2 \frac{d^2\eta}{dx^2} + x \frac{d\eta}{dx} - \eta - \frac{\lambda}{m^2} \eta^3 = 0 . \tag{5.2}$$

Rescaling, let $Y = (\lambda / m^2)^{1/2} \eta$; then Eq. (5.2) becomes

$$Y'' + (1/x) Y' - Y/x^2 [1 + Y^2] = 0 . \tag{5.3}$$

We now try to find an odd particular solution to Eq. (5.3). Hence we represent Y by the power series

$$\sum_{\substack{n\ \text{even} \\ n=0}}^{\infty} b_n x^{n+1}$$

in Eq. (5.3). Thus we get

$$\sum b_n \{ n(n+2) x^{n-1} - x^{n+1} [b_0^2 + 2 b_2 b_0 x^2 + (2 b_0 b_4 + b_2^2) x^4 + O(x^6)] \} = 0 . \tag{5.4}$$

Equating like powers of x we obtain

b_0 is arbitrary,

$b_2 = (1/2^3) b_0^3$,

$b_4 = (1/2^6) b_0^5$,

$b_6 = (1/2^9) b_0^7$,

$b_8 = (1/2^{12}) b_0^9$,

and by inspection,

$b_n = b_0^{n+1} / 2^{3n/2}$.

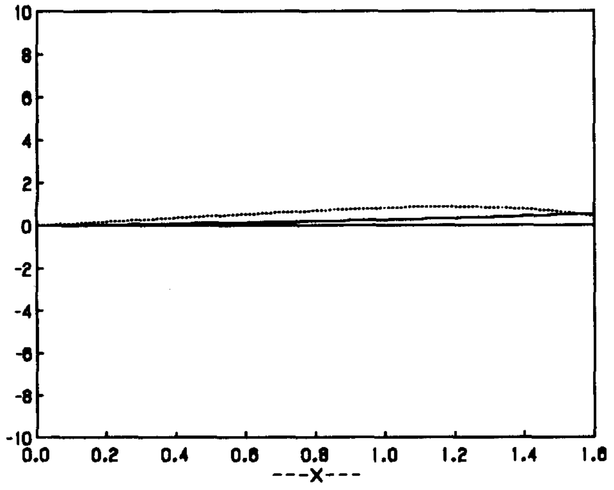


FIG. 5. Graph of $Y(x)/i$ and the error function $(Y'' - Y(1 - Y^2))/i$ for (4.6) $(a_1 = \sqrt{-0.05})$. Dotted line: error function; solid line: Y/i.

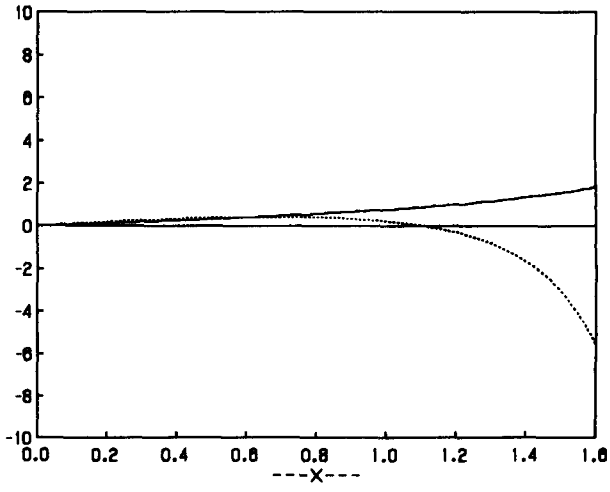


FIG. 7. Graph of $Y(x)/i$ and the error function $(Y'' - Y(1 - Y^2))/i$ for (4.6) $(a_1 = \sqrt{-0.4})$. Dotted line: error function; solid line: Y/i.

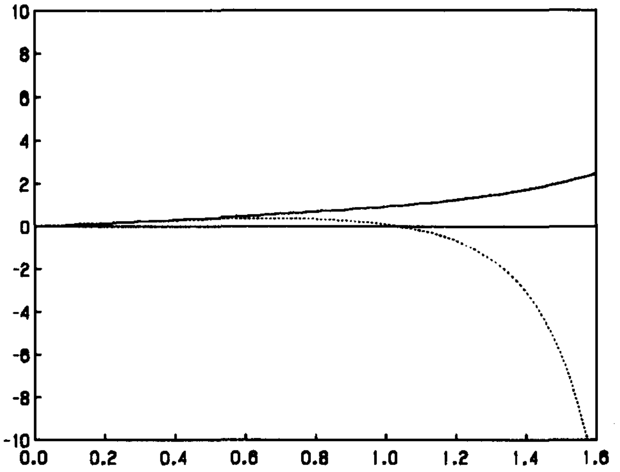


FIG. 8. Graph of $Y(x)/i$ and the error function $(Y'' - Y(1 - Y^2))/i$ for (4.6) $(a_1 = \sqrt{-0.6})$. Dotted line: error function; solid line: Y/i.

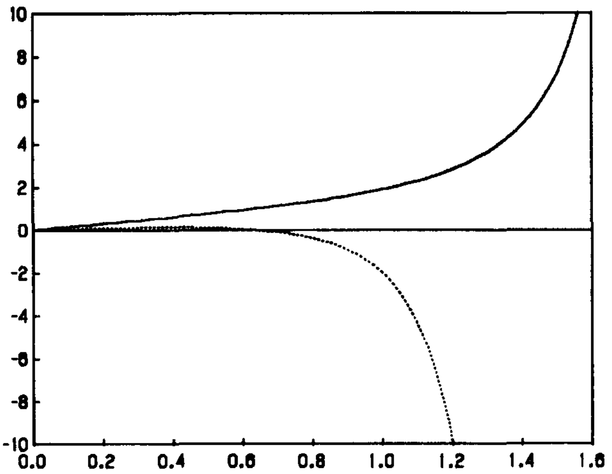


FIG. 9. Graph of $Y(x)/i$ and the error function $(Y'' - Y(1 - Y^2))/i$ for (4.6) $(a_1 = \sqrt{-2})$. Dotted line: error function; solid line: Y/i.

Thus

$$\frac{Y'}{Y} = \sum_{\substack{n=0 \\ n \text{ even}}}^{\infty} (n+1) b_0^{n+1} 2^{-3n/2} x^n \left(\sum_{\substack{n=0 \\ n \text{ even}}}^{\infty} b_0^{n+1} 2^{-3n/2} x^{n+1} \right)^{-1}. \tag{5.5}$$

Constructing the table as described in Sec. II we get

b_0	0	$\frac{b_0^3}{2^3}$	0	$\frac{b_0^5}{2^6}$	0	$\frac{b_0^7}{2^9}$	0	$\cdots$	$\frac{b_0^{2j-1}}{2^{3(j-1)}}$	0	$\frac{b_0^{2j+1}}{2^{3j}}$	0	$\frac{b_0^{2j+3}}{2^{3(j+1)}}$
b_0	0	$\frac{3b_0^3}{2^3}$	0	$\frac{5b_0^5}{2^6}$	0	$\frac{7b_0^7}{2^9}$	0	$\cdots$	$\frac{(2j-1)b_0^{2j-1}}{2^{3(j-1)}}$	0	$\frac{(2j-1)b_0^{2j+1}}{2^{3j}}$	$\cdots$	
$\frac{-2b_0^4}{2^3}$	0	$\frac{-4b_0^6}{2^6}$	0	$\frac{-6b_0^8}{2^9}$	0	$\frac{-8b_0^{10}}{2^{12}}$	0	$\cdots$	$\frac{(-2j)b_0^{2(j+1)}}{2^{3j}}$	0	$\cdots$		.
$\frac{-2b_0^7}{2^6}$	0	$\frac{-4b_0^9}{2^9}$	0	$\frac{-6b_0^{11}}{2^{12}}$	0	$\frac{-8b_0^{13}}{2^{15}}$	0	$\cdots$	$\frac{(-2j)b_0^{2j+5}}{2^{3(j+1)}}$	0	$\cdots$		
$\frac{b_0^{13}}{2^{12}}[8-8]=0$	0	$\frac{b_0^{15}}{2^{15}}[12-12]=0$	0	$\cdots$		$\frac{4b_0^{2j+11}}{2^{3(j+3)}}[(j+1)-(j+1)]\equiv 0$							

Again, each of the last three rows has had the leading zero deleted and has been shifted one column to the left. Therefore from the table we obtain

$$\frac{Y'}{Y} = \frac{b_0 x^{-1}}{b_0} + \frac{-(2b_0^4/8)x^2}{b_0} + \frac{-2(b_0^7/2^6)x^2}{-2(b_0^4/8)}$$
$$= \frac{8 + b_0^2 x}{x(8 - b_0^2 x)}. \tag{5.6}$$

Therefore

$$Y = \exp \int \frac{Y'}{Y} dx = \frac{8 b_0 x}{b_0^2 x^2 - 8}, \tag{5.7}$$

which is an exact particular solution to Eq. (5.3). Therefore

$$\eta = \left(\frac{\lambda}{m^2}\right)^{-1/2} Y = \left(\frac{\lambda}{m^2}\right)^{-1/2} \frac{8 b_0 e^{\pm i \cdot p \cdot x}}{b_0^2 e^{\pm 2ip \cdot x} - 8},$$

which is an exact particular solution[7] to Eq. (5.1).

It is possible to obtain Eq. (5.7) directly by summing the series for Y. However, the continued fraction method should give a closed form expression for Y in all cases, whether the series can be summed or not.

VI. CONCLUDING REMARKS

The continued fraction method outlined above represents a powerful tool for obtaining excellent approximate (and exact) solutions to nonlinear differential equations. In this paper we have produced accurate approximations to solutions of a nonlinear differential equation. In addition, we have provided evidence that the known solutions are members of a family of exact solutions. We have produced approximations to some of these solutions. We have also been able to obtain an exact solution using our method. The strengths of this method are that the calculations are relatively simple, the solutions are expressed in closed form, and the solutions contain arbitrary constants that can accommodate boundary conditions.

ACKNOWLEDGMENTS

We would like to express our appreciation to Dr. Philip Burt for many enlightening discussions.

This work was partially supported by the South Carolina Energy Research and Development Corporation.

[1]P. B. Burt, *Quantum Mechanics and Nonlinear Waves* (Harwood, New York, 1981).
[2]P. B. Burt, Phys. Lett. A **109**, 133 (1985).
[3]P. B. Burt and T. J. Pickett, Lett. Nuovo Cimento **44**, 473 (1985).
[4]G. Baker, Jr. and P. Graves Morris, *Padé Approximants, Part I* (Addison-Wesley, Reading, MA, 1981).
[5]A. N. Khovanskii, *The Applications of Continued Fractions* (Noordhoff, Groningen, 1963).
[6]H. T. Davis, *Introduction to Nonlinear and Integral Equations* (Dover, New York, 1962).
[7]P. B. Burt, Lett. Nuovo Cimento **38**, 103 (1983).

Solutions of the wave equations with boundary conditions on the hyperplane $z-ct=0$

P. Hillion
Institut Henri Poincaré, 75231 Paris, France

(Received 14 January 1988; accepted for publication 13 April 1988)

The boundary value problem for the wave equation is discussed with data $\varphi(r)$ on the hyperplane $z-ct=0$. We give the general form of the solution when $\varphi(r)$ is an entire function of r^2 of order 1 and type $0<\tau<\infty$. As a particular case the focus wave mode solutions of the wave equation are obtained.

I. INTRODUCTION

In recent years it has been shown that the wave equation has pulse wave solutions with Gaussian transverse structure (the so-called focus wave modes). These solutions leave at least two issues unanswered: (i) how is it possible to launch such pulses, and (ii) what is their physical usefulness?

In order to answer the first question (at least partially) we investigated the possibility that they are solutions of a boundary value problem. In fact, we could show that the boundary value problem for the wave equation with data on the hyperplane $z-ct=0$ leads to the focus wave modes when the boundary condition is Gaussian. For other boundary conditions we found new wave solutions. We prove this result here.

Using the variables $\xi=z-ct$, $\bar{\xi}=z+ct$, the wave equation becomes

$$\Delta_\perp \psi + 4\frac{\partial^2\psi}{\partial\xi\,\partial\bar{\xi}}=0,\quad \Delta_\perp\psi=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\psi}{\partial\theta^2},\tag{1}$$

where r, θ are the transverse coordinates. Let ψ be a solution of (1) independent of θ. Then

$$\psi_m=r^m\frac{\partial^m}{(r\,\partial r)^m}(\psi e^{im\theta}),$$
$$i=\sqrt{-1},\quad m \text{ positive integer},\tag{2}$$

is also a solution of Eq. (1). Consequently one just has to consider the solutions that are independent of θ, which we take in the form

$$\psi=\psi(r,\xi)e^{ik\bar{\xi}},\quad k \text{ real scalar},\tag{3}$$

corresponding to cylindrical harmonic waves propagating along 0z. Substituting (3) into (1) gives

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right)+\alpha\frac{\partial\psi}{\partial\xi}=0,\quad \alpha=4ik.\tag{4}$$

We are looking for the solutions of Eq. (4) such that $\psi(r)$ is given on the hyperplane $\xi=0$.

Let us consider the Laplace equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\varphi}{\partial r}\right)+\frac{\partial^2\varphi}{\partial\xi^2}=0,\tag{5}$$

and let $\varphi(r,\xi)$ be a solution of (5) such that $\partial\varphi(r,\xi)/\partial\xi=0$ for $\xi=0$. Then

$$\psi(r,\xi)=\sqrt{\frac{\alpha}{\pi\xi}}\int_0^\infty e^{-\alpha s^2/4\xi}\varphi(r,s)\,ds\tag{6}$$

is a solution of Eq. (4) provided that the integral exists and that differentiation under the sign of integration is permissible.

To prove this last result one just has to take the Laplace transform of both Eqs. (4) and (5) with respect to the variable ξ. The solutions become $\tilde{\varphi}(r,p)$ and $\tilde{\psi}(r,p)$, where p is the symbolic variable and between these solutions one has the relation $\tilde{\psi}(r,p)=\tilde{\varphi}(r,\sqrt{\alpha p})$. Then (6) follows from the classical relation[1,2] between the inverse transforms of functions of p and $\sqrt{p}$.

A simple calculation also allows us to check (6) since

$$\Delta_\perp\psi=\sqrt{\frac{\alpha}{\pi\xi}}\int_0^\infty e^{-\alpha s^2/4\xi}\Delta_\perp\varphi\,ds$$
$$=-\sqrt{\frac{\alpha}{\pi\xi}}\int_0^\infty e^{-\alpha s^2/4\xi}\frac{\partial^2\varphi}{\partial s^2}\,ds.$$

Integrating by parts with the condition $\partial\varphi/\partial s=0$ at $s=0$ supplies the result.

Relation (6) is formal since α is a pure imaginary number. Then we introduce $\alpha_\epsilon=\epsilon+4ik$, where ϵ is a positive arbitrary quantity, and we define (6) as

$$\psi(r,\xi)=\lim_{\epsilon\to0}\sqrt{\frac{\alpha_\epsilon}{\pi\xi}}\int_0^\infty e^{-\alpha_\epsilon s^2/4\xi}\varphi(r,s)\,ds.\tag{6'}$$

From now on we shall use (6) and similar expressions with this last meaning but without writing $\lim_{\epsilon\to0}$ and α_ϵ.

We may remark that $\varphi(r,-\xi)$ is also a solution of (5) and that the derivative of $\varphi(r,\xi)+\varphi(r,-\xi)$ with respect to ξ is zero at $\xi=0$. Consequently one may replace (6) by

$$\psi(r,\xi)=\frac{1}{2}\sqrt{\frac{\alpha}{\pi\xi}}\int_{-\infty}^{+\infty}e^{-\alpha s^2/4\xi}\varphi(r,s)\,ds,\tag{7}$$

without any restriction on the derivative of φ.

As a consequence of (7) we get

$$\lim_{\xi\to0}\psi(r,\xi)=\varphi(r,0);\tag{8}$$

that is, to solve the boundary value problem for Eq. (4) we are led to look for the solutions of Eq. (5) with boundary conditions on the hyperplane $\xi=0$.

 0022-2488/88/081771-05$02.50

Remark 1: A somewhat more general solution of Eq. (4) may be obtained since multiplying $\tilde{\varphi}(r,p)$ by some arbitrary function of p does not affect its being a solution of the Laplace transform equation. Applying the convolution product rule leads to

$$\psi(r,\xi) = \frac{1}{2}\sqrt{\frac{\alpha}{\pi}} \int_0^{\xi} d\zeta \frac{\chi(\xi-\zeta)}{\sqrt{\zeta}} \int_{-\infty}^{+\infty} e^{-\alpha s^2/4\xi} \varphi(r,s) ds. \tag{9}$$

Remark 2: The paraxial approximation of the wave equation (still assuming cylindrical symmetry) is

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \hat{\psi}}{\partial r}\right) + 2ik\frac{\partial \hat{\psi}}{\partial r} = 0. \tag{10}$$

The comparison of (4) and (10) shows that if $\psi(r,\xi)$ is a solution of (4), then

$$\hat{\psi}(r,z) = \psi(r,2z) \tag{11}$$

is a solution of (10) corresponding to the boundary condition $\psi(r)$ on the plane $z = 0$.

We only consider here the solutions $\psi(r,\xi)$ but the results we obtain are valid *mutatis mutandis* for the solutions $\hat{\psi}(r,z)$. We start with a discussion of the general solution of Eq. (5).

II. GENERAL SOLUTION OF EQ. (5)

According to Whittaker[3,4] the general solution of Eq. (5) is

$$\varphi(r,\xi) = \frac{1}{\pi}\int_0^{\pi} f(\xi + ir\cos\theta) d\theta, \tag{12}$$

where f is a summable function such that differentiation under the sign of integration is permissible.

In its holomorphy domain, f has the Taylor series expansion

$$f(\xi + ir\cos\theta) = \sum_{n=0}^{\infty} (i)^n r^n \frac{\cos^n\theta}{n!} f^{(n)}(\xi); \tag{13}$$

then, using the relations

$$\frac{1}{\pi}\int_0^{\pi} \cos^{2n}\theta\, d\theta = \frac{1}{2^{2n}}\binom{2n}{n}, \qquad \frac{1}{\pi}\int_0^{\pi} \cos^{2n+1}\theta\, d\theta = 0, \tag{14}$$

$\varphi(r,\xi)$ becomes

$$\varphi(r,\xi) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2^{2n}} \frac{r^{2n}}{(2n)!}\binom{2n}{n} f^{(2n)}(\xi). \tag{15}$$

This relation shows that $\varphi(r,\xi)$ is a function of r^2, a remark useful later. Moreover one has the inversion formula[4]

$$f(\xi) = \frac{1}{\pi}\int_0^{\infty} \frac{dl}{I_0(lr)} \int_{-\infty}^{+\infty} e^{il(s-\xi)} \varphi(r,s) ds, \tag{16}$$

which suggests that $\varphi(r,\xi)$ is the potential of a cylindrical sheet of charges with strength $f(\xi)$.[5] Here I_0 is the modified Bessel function of the first kind of order zero.

As an illustration of the previous relations, one has, for instance,

$$f(\xi + ir\cos\theta) = e^{-i\lambda(\xi + ir\cos\theta)} \Rightarrow \varphi(r,\xi) = e^{-i\lambda\xi} I_0(\lambda r) \Rightarrow f(\xi) = e^{-i\lambda\xi}. \tag{17}$$

The following relations[4] will be also useful:

$$\frac{1}{\pi}\int_0^{\pi} (\xi + ir\cos\theta)^n d\theta = \rho^n P_n\left(\frac{\xi}{\rho}\right), \tag{18a}$$

$$\frac{1}{\pi}\int_0^{\pi} (\xi + ir\cos\theta)^{-n-1} d\theta = \rho^{-n-1} P_n\left(\frac{\xi}{\rho}\right), \tag{18b}$$

where P_n is the Legendre polynomial of order n and $\rho = (\xi^2 + r^2)^{1/2}$.

Before discussing the boundary value problem for Eqs. (4) and (5) we consider the case when the function f is given.

III. APPLICATION OF RELATIONS (7) AND (12)

Starting with $f(\xi + ir\cos\theta) = e^{-i\lambda(\xi + ir\cos\theta)}$ one has according to (17) $\varphi(r,\xi) = e^{-i\lambda\xi} I_0(\lambda r)$. Substituting this expression into (7) gives

$$\psi(r,\xi) = e^{-\lambda^2\xi/\alpha} I_0(\lambda r). \tag{19}$$

In the same way according to (18a) the function $(\xi + ir\cos\theta)^{-1}$ leads to $\varphi(r,\xi) = (\xi^2 + r^2)^{-1/2}$, which gives

$$\begin{aligned}\psi(r,\xi) &= \sqrt{\frac{\alpha}{\pi\xi}} \int_0^{\infty} \frac{e^{-\alpha s^2/4\xi}}{(r^2+s^2)^{1/2}} ds \\ &= \sqrt{\frac{\alpha}{\pi\xi}} \int_0^{\infty} e^{-(\alpha r^2/4\xi)\sinh^2\theta} d\theta \\ &= \frac{1}{2}\sqrt{\frac{\alpha}{\pi\xi}}\, e^{-\alpha r^2/8\xi} K_0\left(\frac{\alpha r^2}{8\xi}\right),\end{aligned} \tag{20}$$

where K_0 is the modified Bessel function of the second kind of order zero. Let us now consider the function

$$f(\xi + ir\cos\theta) = e^{-\lambda^2(\xi + ir\cos\theta)^2}. \tag{21}$$

Substituting (21) into (12), we get, after interchanging integrations in (7),

$$\begin{aligned}\psi(r,\xi) &= \frac{1}{2\pi}\sqrt{\frac{\alpha}{\pi\xi}} \int_0^{\pi} d\theta \\ &\quad\times \int_{-\infty}^{+\infty} \exp\left\{-\frac{\alpha s^2}{4\xi} - \lambda^2(s + r\cos\theta)^2\right\} ds \\ &= \frac{1}{2\pi}\sqrt{\frac{\alpha}{\pi\xi}} \int_0^{\pi} d\theta\, e^{-u^2\cos^2\theta} \\ &\quad\times \int_{-\infty}^{+\infty} \exp\left\{-\left(\frac{\alpha}{4\xi} + \lambda^2\right)\right. \\ &\quad\times \left.\left(\frac{s + i\lambda^2 r\cos\theta}{\alpha/4\xi + \lambda^2}\right)^2\right\} ds \\ &= \sqrt{\frac{\alpha}{\alpha + 4\xi\lambda^2}}\, \frac{1}{\pi}\int_0^{\pi} e^{-u^2\cos^2\theta} d\theta,\end{aligned}$$

that is,

$$\psi(r,\xi) = \sqrt{\frac{\alpha}{\alpha + 4\xi\lambda^2}}\, e^{-u^2/2} I_0\left(\frac{u^2}{2}\right), \quad u^2 = \frac{\alpha\lambda^2 r^2}{\alpha + 4\xi\lambda^2}. \tag{22}$$

The last example is supplied by the function

$$f(\xi + ir\cos\theta) = J_n(\lambda\xi + i\lambda r\cos\theta), \tag{23}$$

where J_n is the Bessel function of the first kind of order n. One has[6]

$$J_n(\lambda\xi + i\lambda r\cos\theta) = \sum_{m=-\infty}^{+\infty} J_m(\lambda\xi) J_{n-m}(i\lambda r\cos\theta). \tag{23'}$$

Substituting (23') into (12) we get, after interchanging integration and summation in (7),

$$\psi_n(r,\xi) = \frac{1}{2}\sqrt{\frac{\alpha}{\pi\xi}} \sum_{m=-\infty}^{+\infty} \int_{-\infty}^{+\infty} ds\, e^{-\alpha s^2/4\xi} \times J_m(\lambda s) \frac{1}{\pi}\int_0^{\pi} d\theta\, J_{n-m}(i\lambda r\cos\theta), \tag{24}$$

but one has[6,7]

$$\int_{-\infty}^{+\infty} e^{-\alpha s^2/4\xi} J_{2\nu}(\lambda s) ds = \sqrt{\frac{4\pi\xi}{\alpha}}\, e^{-\lambda^2\xi/2\alpha} I_\nu\left(\frac{\lambda^2\xi}{2\alpha}\right), \tag{25}$$

the integral being zero for $m = 2\nu + 1$, while, for $n - m = 2p$,

$$\frac{1}{\pi}\int_0^{\pi} J_{2p}(i\lambda r\cos\theta) d\theta = J_p^2\left(\frac{i\lambda r}{2}\right). \tag{25'}$$

These last two relations lead to

$$\psi_{2n}(r,\xi) = e^{-\lambda^2\xi/2\alpha} \sum_{2p=0}^{\infty} I_p\left(\frac{\lambda^2\xi}{2\alpha}\right) J_{n-p}^2\left(\frac{i\lambda r}{2}\right), \tag{26}$$

where I_p is the modified Bessel function of the first kind of order p.

Tagging the solutions by their registration number, their behavior on the hyperplane $\xi = 0$ is as follows:

$$\psi_{19}(r,0) = I_0(\lambda r), \quad \psi_{20}(r,0) = \frac{1}{r},$$
$$\psi_{22}(r,0) = e^{-\lambda^2 r^2/2} I_0\left(\frac{\lambda^2 r^2}{2}\right), \quad \psi_{26}(r,0) = I_n^2\left(\frac{\lambda r}{2}\right). \tag{27}$$

To obtain bounded solutions for $r \mapsto \infty$, λ must be a pure imaginary scalar in ψ_{19} and ψ_{26}.

IV. SOLUTIONS WITH BOUNDARY CONDITIONS ON $\xi = 0$

We first assume that in its holomorphy domain the function f has the Taylor series expansion

$$f(\xi + ir\cos\theta) = \sum_{n=0}^{\infty} (\xi + ir\cos\theta)^{2n} \frac{f^{(2n)}(0)}{(2n)!},$$
$$f^{(2n)}(0) = \left(\frac{\partial^{2n} f(u)}{\partial u^{2n}}\right)_{u=0}, \tag{28}$$

which implies $f^{(2n+1)}(0) = 0$.

Then using (18a) we get from (12)

$$\varphi(r,\xi) = \sum_{n=0}^{\infty} \rho^{2n} \frac{1}{(2n)!} P_{2n}\left(\frac{\xi}{\rho}\right) f^{(2n)}(0), \tag{29}$$

which leads to

$$\varphi(r,0) = \sum_{n=0}^{\infty} \frac{r^{2n}}{(2n)!} P_{2n}(0) f^{(2n)}(0) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2}\left(\frac{r}{2}\right)^{2n} f^{(2n)}(0), \tag{30}$$

since one has

$$P_{2n}(0) = (-1)^n \frac{1}{2^{2n}}\binom{2n}{n}. \tag{30'}$$

One may also obtained (30) by making $\xi = 0$ in (15).

Let $\varphi(r)$ be given on $\xi = 0$. We assume that $\varphi(r)$ is a function of r^2 with the power series expansion

$$\varphi(r) = \sum_{n=0}^{\infty} \frac{\varphi_n}{n!} r^{2n}. \tag{31}$$

The comparison of (30) and (31) gives

$$f^{(2n)}(0) = (-1)^n 2^{2n} n! \varphi_n. \tag{32}$$

Taking (32) into account, (29) becomes

$$\varphi(r,\xi) = \sum_{n=0}^{\infty} (-1)^n \frac{n!\varphi_n}{(2n)!} (2\rho)^{2n} P_{2n}\left(\frac{\xi}{\rho}\right), \tag{33}$$

which leads to

$$\varphi(0,\xi) = \sum_{n=0}^{\infty} (-1)^n \frac{n!\varphi_n}{(2n)!} (2\xi)^{2n}, \tag{33'}$$

since $P_{2n}(1) = 1$.

Now for $r = 0$, relation (16) reduces to

$$f(\xi) = \frac{1}{\pi}\int_0^{\infty} dl\, e^{-il\xi} \int_{-\infty}^{+\infty} ds\, e^{ils} \varphi(0,s). \tag{34}$$

Substituting (33') into (34) we get, in terms of derivatives of the Dirac distribution,

$$f(\xi) = \sum_{n=0}^{\infty} \frac{n!}{(2n)!} 2^{2n} \varphi_n \int_0^{\infty} \delta^{2n}(l) e^{-il\xi} dl = \sum_{n=0}^{\infty} (-1)^n \frac{n!}{(2n)!} \varphi_n (2\xi)^n.$$

Then using the Borel transform[8] we get finally

$$f(\xi + ir\cos\theta) = \int_0^{\infty} e^{-t} \sum_{n=0}^{\infty} (-1)^n \frac{t^n}{(2n)!} \times \varphi_n (2\xi + ir\cos\theta)^{2n} dt, \tag{35}$$

which brings us back to the situation discussed in the previous section where the function f was given.

Let us consider, for instance, the boundary conditions $\varphi(r) = e^{-\lambda^2 r^2}$; one has $\varphi_n = (-1)^n \lambda^{2n}$ and (35) leads to $f(\xi + ir\cos\theta)$

$$= \int_0^{\infty} e^{-t} \sum_{n=0}^{\infty} \frac{(\lambda^2 t)^n}{(2n)!} (2\xi + ir\cos\theta)^{2n} dt$$
$$= \int_0^{\infty} e^{-t} \cosh(2\lambda\sqrt{t}(\xi + ir\cos\theta)) dt. \tag{36}$$

Substituting (36) into (12) and the result into (7), we get

$$\psi(r,\xi) = \frac{1}{2}\sqrt{\frac{\alpha}{\pi\xi}}\int_{-\infty}^{+\infty} ds\, e^{-\alpha s^2/4\xi}$$
$$\times \frac{1}{\pi}\int_0^{\pi} d\theta \int_0^{\infty} dt \cosh(2\lambda\sqrt{t}\,(s + ir\cos\theta)). \tag{37}$$

Interchanging the integrals one has, after integration on s and θ,

$$\psi(r,\xi) = \frac{1}{2}\int_0^{\infty} e^{-t(1-4\lambda^2\xi/\alpha)} J_0(\lambda r\sqrt{t})dt.$$

Using the variable $t = u^2$, this integral becomes a Weber's integral,[6,7] leading to

$$\psi(r,\xi) = [1/(1-4\lambda^2\xi/\alpha)]\, e^{-\lambda^2 r^2/(1-4\lambda^2\xi/\alpha)}, \tag{38}$$

with $\alpha = 4ik$ the solution (38) multiplied by $e^{ik\bar{\xi}}$, in agreement with (3), becomes a focus wave mode solution of the wave equation.[9]

Using the dimensionless variable $v = -\lambda^2 r^2$, the power series expansion (31) becomes

$$\varphi(v) = \sum_{n=0}^{\infty} \frac{\hat{\varphi}_n}{n!} v^n. \tag{39}$$

This implies that $\varphi(v)$ is an entire function of order 1 and type $0 < \tau < \infty$ (see Ref. 8) with

$$\tau = \lim_{n\to\infty} \sup {}^n\sqrt{|\hat{\varphi}_n|}\,. \tag{40}$$

Consequently, when the boundary condition $\varphi(v)$ on the hyperplane $\xi = 0$ is an entire function of order 1 and type τ, the relations (35), (12), and (7) supply the solution of the wave equation.

The confluent hypergeometric functions are one of the most important classes of this type of entire functions. Many special functions such as the error function, Fresnel integrals, Bessel functions, and Whittaker functions can be expressed in terms of confluent hypergeometric functions.

The exponential function is of type 1 while the Bessel function $I_0(i\sqrt{v})$ is of type $\frac{1}{4}$. Using this result[8] (that is, that the order or type of product of two entire functions is at most the larger of the respective orders or types) we see that the type of $\psi_{22}(r,0) = e^{v/2} I_0(v/2)$ is at most $\frac{1}{2}$.

Remark: The previous results may be generalized to the spinor wave equation[10] with the coordinates $(r,\theta,\xi,\bar{\xi})$:

$$2\frac{\partial}{\partial\bar{\xi}}\psi_1 + e^{-i\theta}\left(\frac{\partial}{\partial r} - \frac{i}{r}\frac{\partial}{\partial\theta}\right)\psi_2 = 0,$$
$$e^{i\theta}\left(\frac{\partial}{\partial r} + \frac{i}{r}\frac{\partial}{\partial\theta}\right)\psi_1 - 2\frac{\partial\psi_2}{\partial\xi} = 0. \tag{41}$$

It is easy to see that each component ψ_1, ψ_2 satisfies the wave equation (1). Looking for the solutions $\psi_{1,2} \equiv \psi_{1,2}(r,\theta,\xi)e^{ik\bar{\xi}}$, we get from (41)

$$2ik\psi_1 + e^{-i\theta}\left(\frac{\partial}{\partial r} - \frac{i}{r}\frac{\partial}{\partial\theta}\right)\psi_2 = 0,$$
$$e^{i\theta}\left(\frac{\partial}{\partial r} + \frac{i}{r}\frac{\partial}{\partial\theta}\right)\psi_1 - 2\frac{\partial\psi_2}{\partial\xi} = 0, \tag{41'}$$

which leads to

$$\psi_1 = \frac{i}{2k} e^{-i\theta}\left(\frac{\partial}{\partial r} - \frac{i}{r}\frac{\partial}{\partial\theta}\right)\psi_2, \tag{42}$$

while ψ_2 is a solution of the wave equation.

Consequently, taking for ψ_2 any of the solutions previously discussed and using (42), one can obtain a solution of the spinor wave equation.

Let us note that, for the spinor field, the boundary conditions on the hyperplane $\xi = 0$ have to satisfy condition (42).

V. DISCUSSION

We have only considered here the solutions with cylindrical symmetry but, using (2), it is easy to generalize the previous results to boundary conditions on the hyperplane $\xi = 0$ of the type $\varphi(r)e^{im\theta}$.

According to the second remark at the end of the Introduction, we may consider this work as the relativistic generalization of the boundary value problem for the paraxial equation with data on the plane $z = 0$. Let us notice that for an observer traveling with the wave the solutions $\psi(r,\xi)$ have the look $\psi(r)$, a remark already made by Belanger[11] for the focus wave modes.

It is also interesting to consider the relativistic behavior of the solutions under a Lorentz transformation. Indeed for an observer moving along 0z with the uniform velocity v, the transverse coordinates r,θ are invariant while one has

$$(\xi,k)\mapsto(\xi',k') = \sqrt{(1-\beta)/(1+\beta)}\,(\xi,k),$$
$$\beta = v/c, \quad \bar{\xi}\mapsto\bar{\xi}' = \sqrt{(1+\beta)/(1-\beta)}\,\bar{\xi}. \tag{43}$$

Using (43) it is easy to prove that the solutions that we previously obtained behave as scalar under the Lorentz group.

The physical usefulness[12] of these solutions, an issue previously stated in the Introduction, will be discussed later.

Let us call the previous solutions, solutions of class a and note $\psi_a(r,\xi,\lambda)$, λ being the parameter of the dimensionless variable v. One can obtain two other classes of solutions. First we define the solutions of class b by the relation

$$\psi_b(r,\xi',\lambda) = \int_{-\infty}^{+\infty} f(\xi - s)\psi_a(r,s',\lambda)ds, \tag{44}$$

where f is a differentiable function such that $f\psi_a$ is zero at infinity. An interesting case is when f is the Gaussian function shifted in direct and Fourier transform spaces:

$$f(s,\xi;w) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{(s-\xi)^2}{4\xi^2} + iw\left(s - \frac{\xi}{2}\right)\right\}, \tag{45}$$

where σ and w are positive scalars. Then ψ_b is the Gabor transform of ψ_a (see Refs. 13 and 14). The solutions of class c are obtained by a weighted integration on λ:

$$\psi_c(r,\xi;\mu) = \int_0^{\infty} F\left(\frac{\lambda}{\mu}\right)\psi_a(r,\xi;\lambda)d\lambda. \tag{46}$$

Taking the focus wave mode solutions, for ψ_a relation (46) supplies the splash wave mode solutions.[15]

[1] B. Van des Pol and H. Bremmer, *Operational Calculus* (Cambridge U.P., Cambridge, 1959).
[2] G. Doetsch, *Introduction to the Theory and Application of the Laplace Transform* (Springer, Berlin, 1974).
[3] E. T. Whittaker, Math. Ann. LVII, 333 (1902).
[4] H. Bateman, *Partial Differential Equations of Mathematical Physics* (Cambridge U.P., Cambridge, 1959).
[5] P. M. Morse and H. Feschbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[6] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1953).
[7] M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1968).
[8] R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[9] A. Sezginer, J. Appl. Phys. **57**, 678 (1958).
[10] P. Hillion, J. Math. Phys. **28**, 1743 (1987).
[11] P. A. Belanger, J. Opt. Soc. Am. A **3**, 541 (1986).
[12] T. T. Wu and H. Lehmann, J. Appl. Phys. **58**, 2064 (1985).
[13] D. Gabor, J. Inst. Elec. Eng. **93**, 429 (1946).
[14] C. W. Helstrom, IEEE Inf. **12**, 81 (1966).
[15] R. W. Ziolkowski, J. Math. Phys. **26**, 861 (1985).

Solutions of reaction-diffusion integral equations describing explosive evolution of densities for localized structures

Hans Wilhelmsson[a),b)]

Laboratoire de Physique Théorique et Mathématique, Université Paris VII, Tour centrale, 3 e étage, 2 place Jussieu, Paris Cedex 05, France and Laboratoire Physique des Milieux Ionisés, Ecole Polytechnique, F-91128 Palaiseau, France

(Received 24 June 1987; accepted for publication 30 March 1988)

Reaction-diffusion integral equations are studied with special regard to explosive-type solutions. Certain localized structures can grow explosively in time without change in spatial structure. The characteristic shapes can be described in terms of elliptic functions.

I. INTRODUCTION

Modern science exhibits a rich variety of phenomena of nonlinear nature. Their theoretical description often poses extremely challenging problems. Examples of such nonlinear phenomena are found in plasma and laser physics, classical mechanics, chemistry, and population biology.

One type of phenomenon that may occur in plasma physics as well as in other fields is the so-called explosive instabilities,[1,2] characterized by the fact that amplitudes of interacting waves, or population densities, may grow to infinite values in a finite time. In plasmas, such explosive instabilities are caused by positive–negative energy wave interaction, which may occur, e.g., in systems that possess free energy, such as electron beams in plasmas.[3,4] The plasmon density may then grow towards infinite values in a finite period of time.

It is the purpose of this paper to study a generalized form of the reaction-diffusion equation,[5] which includes also a time-integral part and to look for solutions of explosive type.

II. THE REACTION-DIFFUSION INTEGRAL EQUATION

Consider the equation

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial n}{\partial x}\right) + cn^2 + dn\int_{-\infty}^{t} n^2\,dt'. \qquad (1)$$

Here we assume that the diffusion coefficient $D = an$, where a is a constant, i.e., we have a linear dependence of the diffusion coefficient on the density n. Equation (1) is clearly nonlocal in time but local in space.

The assumption that the contribution from the integral term is formally local in space means that it is assumed that the process related to the integrand occurs, for all times, at or close to the point x, or that, as a result of spatial homogeneity, the contributions from processes at a distance could be considered equivalent to nearby processes. It remains to be verified separately to what extent the assumption about locality in space is justified for each particular situation that the model equation will be applied to.

It should be emphasized that spatial homogeneity might be justified as an approximation when representing correlation effects in the integrand of the integral term including the reaction processes, whereas inhomogeneity effects could be more important for the diffusion term, because of the derivatives. It should also be emphasized in this connection that the solutions we consider here are solutions that are localized in space. In ionized media spatial localization of processes is favored by low temperatures and the presence of magnetic fields.

One might imagine various situations where the integral term would correspond to physical situations of annihilating or creative processes. For such cases, where in Eq. (1) the constant d is negative, i.e., when annihilating processes are considered, the integral term could as an example correspond to an attachment, e.g., a loss of a free electron, to a neutral particle, which was formed at an earlier stage by a process of recombination of an electron and an ion, the rate of the early process being proportional to $n_e^2(t')$, assuming in the earlier stage $n_e(t') = n_i(t')$, where n_e and n_i represent the electron and ion densities. In Eq. (1) (the second term on the rhs) the coefficient c could, however, be positive as a result of creation of new free electrons by collisional ionization of ions by electrons, again assuming $n_e = n_i$.

Alternative cases, where the constant d is positive, would, for example, correspond to plasma–plasma interaction for situations where free energy is available, which are accordingly explosive, or to creative interaction of specia in population biology.

III. SOLUTIONS OF THE REACTION-DIFFUSION INTEGRAL EQUATION

Let us introduce the following renormalization of space and time, namely

$$(c/a)^{1/2}x \to x, \quad ct \to t$$

and furthermore the quantity $\epsilon = d/c^2$.

Then Eq. (1) takes the form

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x}\left(n\frac{\partial n}{\partial x}\right) + n^2 + \epsilon n\int_{-\infty}^{t} n^2\,dt'. \qquad (2)$$

We look for separable solutions in the form

$$n(x,t) = T(t)X(x). \qquad (3)$$

Inserting expression (3) into Eq. (2) and dividing through by T^2X leads to

---

[a)] On leave of absence from the Institute for Electromagnetic Theory and Plasma Physics, Chalmers University of Technology, S-412 96 Göteborg, Sweden.

[b)] Titulaire d'une chaire de physique des plasmas de la Fondation de France.

$$\frac{\dot{T}}{T^2} = \frac{1}{X}\frac{\partial}{\partial x}\left(X\frac{\partial X}{\partial x}\right) + X + \epsilon X^2 \frac{1}{T}\int_{-\infty}^{t} T^2(t')dt'. \quad (4)$$

Requiring that the right-hand side of Eq. (4) be independent of t leads to the differential equation $T^2 = \alpha\dot{T}$, where α is an arbitrary constant. This equation has the solution $T(t) = -\alpha/(t+\alpha\beta)$, where $\alpha\beta$ must be negative for the solution to be valid for large negative t. Thus T can be written as $T = T(0)\cdot t_0/(t_0 - t)$, where

$$t_0 = -\alpha\beta = \left\{\frac{d}{dt}[\ln T(t)]\right\}_{t=0} = \left\{\frac{\partial}{\partial t}[\ln n(x,t)]\right\}_{t=0}$$

is the time of explosion which for separable solutions, as expressed by relation (3), does not depend on x.

Furthermore let us introduce the convenient notation $\phi = \alpha X$. From Eq. (2) we then obtain the following ordinary differential equation:

$$\frac{d}{dx}\left(\phi\frac{d\phi}{dx}\right) = \phi - \phi^2 - \epsilon\phi^3. \quad (5)$$

By multiplying both sides of Eq. (5) by $\phi(d\phi/dx)$ and integrating twice we can express

$$x - x_0 = \pm\frac{1}{\sqrt{2}}\int_{\phi_0}^{\phi}\frac{\phi\, d\phi}{[\phi^3/3 - \phi^4/4 - \epsilon(\phi^5/5) + C]^{1/2}}, \quad (6)$$

where C is a constant of integration. In the following we here chose $C = 0$.

By making the convenient substitution $y = \sqrt{\phi}$ we then obtain

$$x = \sqrt{6}\int_{\sqrt{\phi}}^{\sqrt{\phi_0}}\frac{dy}{[1 - \frac{3}{4}y^2 - \epsilon\frac{3}{5}y^4]^{1/2}}. \quad (7)$$

First, consider the case where $\epsilon > 0$, i.e., $d > 0$, $c > 0$ in Eq. (1). We express the relation (7) in the form

$$x = \left(\frac{10}{\epsilon}\right)^{1/2}\int_{\sqrt{\phi}}^{\sqrt{\phi_0}}\frac{dy}{[(a^2+y^2)(b^2-y^2)]^{1/2}}, \quad (8)$$

where

$$a^2 = \frac{5}{8\epsilon} + \left[\left(\frac{5}{8\epsilon}\right)^2 + \frac{5}{3\epsilon}\right]^{1/2},$$
$$b^2 = -\frac{5}{8\epsilon} + \left[\left(\frac{5}{8\epsilon}\right)^2 + \frac{5}{3\epsilon}\right]^{1/2}, \quad (9)$$

and

$$\phi_0 = b^2, \quad a^2 > b^2.$$

The integral in relation (8) can then be expressed in terms of an elliptic integral $F(k,\varphi)$ as

$$\int_{\sqrt{\phi}}^{\sqrt{\phi_0}}\frac{dy}{[(a^2+y^2)(b^2-y^2)]^{1/2}} = [a^2+b^2]^{-1/2}F(k,\varphi), \quad (10)$$

where

$$k = b/[a^2+b^2]^{1/2}, \quad (11)$$

$$\cos\varphi = \sqrt{\phi}/b = (\phi/\phi_0)^{1/2}. \quad (12)$$

In relations (10) and (11)

$$[a^2+b^2]^{1/2} = \sqrt{2}[(5/8\epsilon)^2 + 5/3\epsilon]^{1/4}$$

so that we obtain from (8)

$$x = \sqrt{5}\left[\left(\frac{5}{8}\right)^2 + \frac{5}{3}\epsilon\right]^{-1/4} F\left\{\frac{b}{(a^2+b^2)^{1/2}}, \cos^{-1}\frac{\sqrt{\phi}}{b}\right\}, \quad (13)$$

which defines x as a function of ϕ in terms of the elliptic integral F, with a and b expressed by (9). We notice that when $\epsilon \to 0$, i.e., $d \to 0$ in Eq. (1), $k \to 0$ in (11), and since $F(0,\varphi) = \varphi$, from (13)

$$\lim_{\epsilon\to 0} x = 2\sqrt{2}\cos^{-1}((\sqrt{3}/2)\sqrt{\phi}),$$

or

$$\phi = \tfrac{2}{3}[1 + \cos(x/\sqrt{2})]. \quad (14)$$

We now, furthermore, consider the case where $\epsilon < 0$, i.e., $d < 0$, $c > 0$. In this case there is a partial counterbalance in between the integral term and the quadratic term in Eq. (1). Here we make use of the elliptic integral form

$$\int_{\sqrt{\phi}}^{\sqrt{\phi_0}}\frac{dy}{[(A^2-y^2)(B^2-y^2)]^{1/2}} = B^{-1}F(k,\varphi), \quad (15)$$

where

$$k = A/B, \quad (16)$$

$$\cos\varphi = \left[\frac{\phi}{A^2}\frac{B^2-A^2}{B^2-\phi}\right]^{1/2}, \quad (17)$$

where

$$A^2 = \frac{5}{8|\epsilon|} - \left[\left(\frac{5}{8|\epsilon|}\right)^2 - \frac{5}{3|\epsilon|}\right]^{1/2},$$
$$B^2 = \frac{5}{8|\epsilon|} + \left[\left(\frac{5}{8|\epsilon|}\right)^2 - \frac{5}{3|\epsilon|}\right]^{1/2}. \quad (18)$$

For the expressions in the square roots of (18) to be positive we have to assume $|\epsilon| < \frac{15}{64}$. In (15)–(18) we have $B > A > y$, $\phi_0 = A^2$. For $\epsilon < 0$ we then obtain

$$x = \sqrt{10}\,[\tfrac{5}{8} + ((\tfrac{5}{8})^2 - \tfrac{5}{3}|\epsilon|)^{1/2}]^{-1/2} \cdot F\left\{\frac{A}{B}, \cos^{-1}\left[\frac{\phi}{A^2}\frac{B^2-A^2}{B^2-\phi}\right]^{1/2}\right\} \quad (|\epsilon| < \tfrac{15}{64}), \quad (19)$$

which for this case defines x as a function of ϕ in terms of the elliptic integral F, with A and B expressed by (18).

In the limit $|\epsilon| = \frac{15}{64}$ we have $k = 1$, $\varphi = \pi/2$ or 0 and $F(1,\varphi) = \ln[\tan(\pi/4 + \varphi/2)]$. From (17) and (18) we then obtain $F = \infty$ and $F = 0$ in the two points $\varphi = \pi/2$ and $\varphi = 0$, and correspondingly $x = \infty$ and $x = 0$. In these points $\phi = \phi_0 = \frac{8}{3}$ from (18), when $|\epsilon| = \frac{15}{64}$.

From the relation (19) we find that in the limit $|\epsilon| \to 0$ we again recover the simple result (14), as we should.

In Fig. 1 ϕ is plotted as a function of x from relations (13), (14), and (19) for $\epsilon > 0$, $\epsilon = 0$, and $\epsilon < 0$, respectively.

The related solutions to Eq. (2), expressed in renormalized variables, therefore become

$$n(x,t) = (t_0 - t)^{-1}\phi(x), \quad (20)$$

where for the different cases ϕ is related to x as given by relations (13) and (19) or by (14) for $\epsilon = 0$.

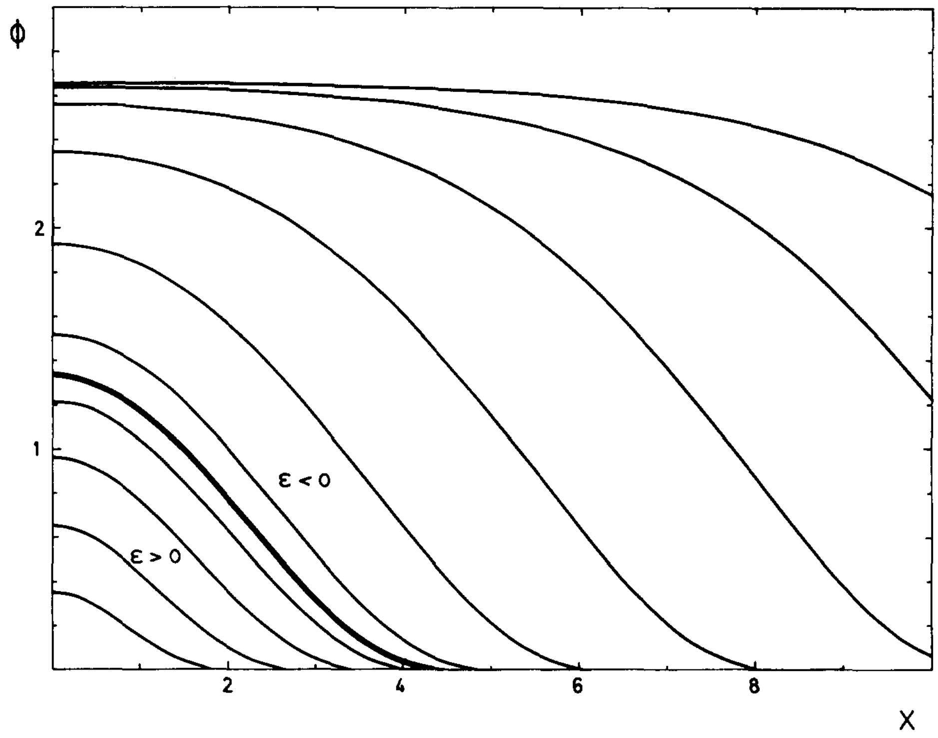


FIG. 1. The function $\phi(x)$ in relations (13), (14), and (19) for $\epsilon > 0$, $\epsilon = 0$, and $\epsilon < 0$, respectively. The more strongly marked curve corresponds to $\epsilon = 0$. The ϵ-values chosen for $\epsilon > 0$ are 0.1, 0.5, 2, and 10. For $\epsilon < 0$ the plotted curves correspond to $|\epsilon| = 0.1$, 0.2, 0.23, 0.234, 0.234 35, 0.234 373, approaching the horizontal line $\phi = \phi_0 = \frac{3}{4}$ in the limit where $|\epsilon| = \frac{15}{64} = 0.234\,375$.

IV. INITIAL AND BOUNDARY CONDITIONS

It would be appropriate to address the question of suitable initial conditions and proper boundary conditions for Eq. (2). For comparison it would be interesting to note that if one solves the corresponding equation where the diffusion term as well as the integral term is absent one obtains a solution $n(x,t) = n(x,0)/[1 - n(x,0)\cdot t]$, which is not in a separable form since the time of explosion $t_0 = n^{-1}(x,0)$ depends on x. Only if $n(x,0) = n_0$ is a constant, which corresponds to $X(x) = \alpha^{-1}$ and to $\phi = 1$, do we obtain a "separable solution" in the limit for what remains of Eq. (4) if the diffusion term and the integral term is omitted. In the expression for $t_0 = \{(\partial/\partial t)[\ln(x,t)]\}_{t=0}$ it is, in fact, necessary that for t_0 constant when considering the full Eq. (4) that $n(x,t)$ is either independent of x or that it is of the separable form (3). For the latter case different values of the arbitrary constant α, or of t_0, would correspond to different initial stages of the shape-preserving evolution of $n(x,t)$.

It should be mentioned in this connection that proper boundary conditions could be either localized or periodic as allowed by the structure of the solutions of Eq. (2). Here we assume that the extension of the localized structure of the solution is smaller than that of any external boundary.

The next interesting question to address would then be: What happens if initially the space dependence differs from the form defined by $\phi(x)$ as a solution of Eq. (5). For cases where the integral term in Eq. (2) is omitted the answer is that the new different form of the space dependence will adjust itself in the course of time and approach the form $\phi(x)$ asymptotically. Analytic theory to be published elsewhere by the author on these issues has been confirmed by computer investigations. These results confirm the interesting role that the simple localized solution (15) for $\epsilon = 0$ (see also Fig. 1) could play in more complex situations. They, furthermore, elucidate the properties of dynamic stability of the shape of the localized solutions. The interesting issue of the interaction of two or several localized solutions has also been addressed by the author in separate forthcoming publications considering the case $\epsilon = 0$. Such localized solutions might, as it turns out for the case $\epsilon = 0$, even be considered as units of superposition for complex situations and generalized systems.

V. SOLUTIONS OF COUPLED REACTION-DIFFUSION INTEGRAL EQUATIONS

Next let us consider the coupled set of reaction-diffusion integral equations

$$\frac{\partial n_1}{\partial t} = \frac{\partial}{\partial x}\left(D_1 \frac{\partial n_1}{\partial x}\right) + c_1 n_1^2 + g_1 n_1 n_2 + d_1 n_1 \int_{-\infty}^{t} n_1 n_2 \, dt', \tag{21}$$

$$\frac{\partial n_2}{\partial t} = \frac{\partial}{\partial x}\left(D_2 \frac{\partial n_2}{\partial x}\right) + c_2 n_2^2 + g_2 n_1 n_2 + d_2 n_2 \int_{-\infty}^{t} n_1 n_2 \, dt', \tag{22}$$

where we assume

$$D_1 = D_2 = A(n_1 + n_2), \quad d_1 = d_2 = d.$$

Solutions to the coupled Eqs. (21) and (22) can then be obtained in the form $n_1 = a_1 n$, $n_2 = a_2 n$, where a_1 and a_2 are determined by

$$a_1 = \frac{c_2 - g_1}{c_1 c_2 - g_1 g_2} c, \quad a_2 = \frac{c_1 - g_2}{c_1 c_2 - g_1 g_2} c, \tag{23}$$

and n satisfies the equation

$$\frac{\partial n}{\partial t} = A(a_1 + a_2) \frac{\partial}{\partial x}\left(n \frac{\partial n}{\partial x}\right) + cn^2 + da_1 a_2 n \int_{-\infty}^{t} n^2 \, dt', \tag{24}$$

which is of the type solved above.

VI. CONCLUDING REMARKS

Exact solutions have been obtained for the reaction-diffusion integral equation (2) for positive and negative values of the coefficient ϵ.

An evolution according to these solutions would require particular initial conditions in space to be fulfilled, which conform with the obtained solutions. One might expect, however, that properties similar to those described in Sec. IV for the case $\epsilon = 0$, but for generalized initial conditions, would prevail also for cases where $|\epsilon| > 0$. Confirmation of such a prediction would undoubtedly attach some extra weight to the importance of the solutions obtained here. Detailed considerations of related questions are, however, beyond the scope of the present paper.

The solutions found may be regarded as examples of particular solutions of reaction-diffusion integral equations. They may offer a challenge for further investigations.

ACKNOWLEDGMENT

It is a pleasure for me to extend thanks to Hans-George Gustavsson who arranged the computer plotting of the results expressed by relations (13) and (19) as depicted in Fig. 1.

[1]B. B. Kadomtsev, A. B. Mikhailovskii, and A. V. Timofeev, Sov. Phys. JETP **20**, 1517 (1965).

[2]B. Coppi, M. N. Rosenbluth, and R. N. Sudan, Ann. Phys. (NY) **55**, 207 (1969).

[3]H. Wilhelmsson, L. Stenflo, and F. Engelmann, J. Math. Phys. **11**, 1738 (1970).

[4]J. Weiland and H. Wilhelmsson, *Coherent Non-Linear Interaction of Waves in Plasmas* (Pergamon, Oxford, 1977), Vol. 88.

[5]H. Wilhelmsson, Phys. Rev. A **36**, 965 (1987).

Introducing division by an "a" number and a new "b" number in particle physics

Quan Xiang-Lin
Department of Physics, Dalian Institute of Technology, Dalian, People's Republic of China
Wang Lian-Ji
Dalian Marine College, Dalian, People's Republic of China
Song He-Shan
Department of Physics, Dalian Institute of Technology, Dalian, People's Republic of China

(Received 19 February 1986; accepted for publication 27 January 1988)

In this paper a new "b" number ($b = a_1 a_2 \cdots a_n$, where n is even and each a is an anticommutative number or Grassmann number) is introduced, and the specialities of this "b" number are also expounded. Moreover, division by an "a" number is defined. At the same time a derivative concept of the functional containing an "a" number is also given in the traditional approach, and a new formula for the derivative of this functional is advanced. The new formula not only encompasses Berezin's formula but also solves the problems that cannot be solved by Berezin's formula. For "b" numbers, a unique infinitesimal calculus will be proposed.

I. INTRODUCTION

In 1844 Grassmann's algebra was put forward.[1,2] In 1965 Berezin gave the rules of functional derivation and integration for Grassmann algebra elements.[3,4] In the 1970's the supersymmetry theory was formulated. The theory uses Grassmann's algebraic elements θ_α as parameters. These are anticommutative numbers, i.e., "a" numbers for the sake of brevity.

In this paper we introduce a new kind of number, the "b" number, which is neither a real number nor a complex number except for $b = 0$, and has special properties.

Since Grassmann expounded his algebra, division in the set area of "a" numbers has not been defined up to now. What is the source of the difficulty? It seems that $\theta_i \times 1$ does not have a definite value. For example, defining $\theta_i/\theta_i = \theta_i \times \theta_i^{-1} = 1$ (θ_i is an "a" number), it should be $\theta_i \times 1 = \theta_i$ seemingly. But when we use the multiplicative associative law to calculate, then

$$\theta_i \times 1 = \theta_i \times (\theta_i \times \theta_i^{-1})$$
$$= (\theta_i \times \theta_i) \times \theta_i^{-1} = 0 \times \theta^{-1} = 0.$$

According to the traditional idea of real 1, it seems that the former is true and the latter is wrong. It seems that to define division by an "a" number, the associative law of "a" numbers is destroyed. However, we will see that the former is wrong and the latter is true, because in $\theta_i \times \theta_i^{-1}$ the θ_i and θ_i^{-1} are all "a" numbers, and the product of these two "a" numbers is unity (i.e., 1). This 1 is in the category of "b" numbers rather than real numbers. We will show that $\theta_j (\neq \theta_i)$ multiplied by the "b" number 1 is equal to θ_j (i.e., $\theta_j \times 1 = \theta_j$) and if $\theta_j = \theta_i$, then $\theta_i \times 1 = 0$. It follows that the "b" number 1 relates to the "a" number θ_i. Hereafter we write the "b" number 1 as $1(\theta_i)$ for clarity. Thus we have this important result: Within the range of "b" numbers the product of some "a" number and $1(\theta_i)$ is equal either to this "a" number or to zero. Then the associative law of "a" numbers holds true. After "b" numbers are introduced we can define division by an "a" number.

II. OPERATIVE RULES OF "a" NUMBERS

In a number of references[5,6,7] it is implied that the multiplication of "a" numbers satisfies the associative law and the anticommutative law. On the basis of this point this paper expounds the operative rules of "a" numbers in greater detail. Let θ_α be an "a" number. From the definition of the external product in Grassmann's algebra we know[2] the anticommutive law

$$\theta_{\alpha_1}\theta_{\alpha_2} = -\theta_{\alpha_2}\theta_{\alpha_1}, \tag{2.1}$$

$$\theta_\alpha \cdot \theta_\alpha = \theta_\alpha^2 = 0, \tag{2.2}$$

in which α is not a repetitive index.

Hypothesis 1: An "a" number multiplied by zero is zero still, that is,

$$\theta_\alpha \cdot 0 = 0 \cdot \theta_\alpha = 0. \tag{2.3}$$

Corollary 1: Zero (0) is an "a" number.

Definition 1: The continuous product of an "a" number is as follows:

$$\theta_1\theta_2\theta_3\theta_4 = [(\theta_1\theta_2)\cdot\theta_3]\cdot\theta_4. \tag{2.4}$$

Hypothesis 2: The multiplication of "a" numbers obeys the associative law, i.e.,

$$\theta_1\theta_2\theta_3\theta_4 = [(\theta_1\theta_2)\cdot\theta_3]\theta_4 = [\theta_1(\theta_2\theta_3)]\theta_4$$
$$= (\theta_1\theta_2)(\theta_3\theta_4) = \theta_1[\theta_2(\theta_3\theta_4)]. \tag{2.5}$$

Theorem 1: Any two adjacent "a" numbers $\theta_i\theta_j$ in a continuous product $\theta_1\theta_2\cdots\theta_i\theta_j\cdots\theta_n$ obey the anticommutative law.

Proof: Note $\theta_3\theta_4$ in the continuous product $\theta_1\theta_2\theta_3\theta_4$. From Eqs. (2.5) and (2.1) we have

 0022-2488/88/081780-06$02.50

$$\theta_1\theta_2\theta_3\theta_4 = (\theta_1\theta_2)(\theta_3\theta_4)$$
$$= (\theta_1\theta_2)(-\theta_4\theta_3) = -\theta_1\theta_2\theta_4\theta_3.$$

To define division by an "a" number we introduce the new "b" number and study its regularity.

III. THE "b" NUMBER

Definition 2: Let θ_α be an "a" number and

$$b = \theta_{\alpha_1}\theta_{\alpha_2}\cdots\theta_{\alpha_n} \quad (n = 2,4,6,...). \tag{3.1}$$

Equation (3.1) defines a "b" number.

Theorem 2: Zero is also a "b" number.

From Corollary 1 we know that zero is an "a" number. But from Definition 2, $0\cdot\theta_\alpha$ and $\theta_\alpha\cdot 0$ have to be "b" numbers. According to Eq. (2.3) the "b" numbers $0\cdot\rho_\alpha$ and $\theta_\alpha\cdot 0$ are zero. Obviously the result zero is not only an "a" number but also a "b" number.

Theorem 3: A "b" number obeys the commutative law of multiplication.

This can be proved at once by Theorem 1.

Theorem 4: A "b" number obeys the associative law of multiplication.

IV. SPECIAL PROPERTIES OF THE "b" NUMBER

To sum up, "b" numbers and real numbers have only one common element, zero. At the same time they both obey the commutative law and the associative law. Naturally, we can also determine that "b" numbers obey the distributive law; this determination is not in contradiction to what was said above. However, whether or not the "b" number is a real (or a complex) number is still in question.

The real number has two properties, as everyone knows.

(1) If $C_1\cdot C_2 = 0$, then either $C_1 = 0$ or $C_2 = 0$; or if $C_1^2 = 0$, then $C_1 = 0$ must be satisfied. (4.1)

(2) If $C_1 \neq 0$ and $C_2 \neq 0$, then $C_1C_2 \neq 0$. (4.2)

However, the "b" number does not have these properties, but does have its own special characteristics.

Property 1: The product of a "b" number multiplied by itself is always equal to zero.

Let $\theta_\beta \neq 0$, $\theta_{\beta'} \neq 0$, and $b = \theta_\beta\theta_{\beta'}$. Then we can obtain

$$b_\beta^2 = \theta_\beta\theta_{\beta'}\theta_\beta\theta_{\beta'} = \theta_\beta(-\theta_\beta\theta_{\beta'})\times\theta_{\beta'}$$
$$= -\theta_\beta\theta_\beta\theta_{\beta'}\theta_{\beta'} = 0. \tag{4.3}$$

In general we have

$$b_\beta^n = 0 \quad (n = 2,3,4,...). \tag{4.4}$$

Property 2: The product of two nonzero "b" numbers is probably zero.

For example, $b_{\beta_1} = \theta_{\beta_1}\cdot\theta_\alpha \neq 0$, $b_{\beta_2} = \theta_{\beta_2}\cdot\theta_\alpha \neq 0$, but from Eqs. (2.1) (2.2), and (2.4) we have

$$b_{\beta_1}\cdot b_{\beta_2} = \theta_{\beta_1}\theta_\alpha\cdot\theta_{\beta_1}\cdot\theta_\alpha = \theta_{\beta_1}(-\theta_{\beta_2}\theta_\alpha)\theta_\alpha = 0. \tag{4.5}$$

Property 3: The product of a nonzero "b" number and a nonzero "a" number is probably zero.

For example, consider $b = \theta_1\theta_2$, where $\theta_1 \neq 0$, $\theta_2 \neq 0$ and $\theta_1 \neq \theta_2$. Then $b \neq 0$. However,

$$b\theta_1 = \theta_1\theta_2\theta_1 = -\theta_1\theta_1\theta_2 = 0, \quad b\theta_2 = \theta_1\theta_2\theta_2 = 0. \tag{4.6}$$

It can be readily found that whether or not b equals zero depends upon whether or not the factor θ_1 is contained in b.

From the three special properties of "b" numbers it follows that "b" numbers (except for zero) are neither real nor complex numbers. Now that we have introduced "b" numbers, division by an "a" number can be defined.

V. THE REGULARITY OF DIVISION BY AN "a" NUMBER

The key to division by an "a" number is to define its inverse elements. The inverse element of θ_i is θ_i^{-1} in the case of "a" numbers. We write down the equation

$$\theta_i/\theta_i \equiv \theta_i\times\theta_i^{-1} \equiv 1(\theta_i). \tag{5.1}$$

Because $1(\theta_i)$ is the product of two "a" numbers, it should have all the characteristics of "b" numbers as seen in the previous section. Then $1(\theta_i)$ has the following results:

$$\theta_j\cdot 1(\theta_i) = \theta_j\times\theta_i\times\theta_i^{-1}\begin{cases} = 0, & \text{for } \theta_i = \theta_j, \quad (5.2)\\ \neq 0, & \text{for } \theta_i \neq \theta_j, \quad (5.3)\end{cases}$$

$$\theta_j^{-1}\cdot 1(\theta_i) = \theta_j^{-1}\times\theta_i\times\theta_i^{-1}\begin{cases} = 0, & \text{for } \theta_i = \theta_j,\\ \neq 0, & \text{for } \theta_i \neq \theta_j.\end{cases} \tag{5.4}$$

From the special properties of the "b" number we know only that $\theta_j\times 1(\theta_i)$ and $\theta_j^{-1}\times 1(\theta_i)$ $(\theta_j \neq \theta_i)$ may be nonzero, but according to the traditional concept of unity 1 we can define

$$\left.\begin{aligned}\theta_j\times 1(\theta_i) &= \theta_j,\\ \theta_j^{-1}\times 1(\theta_i) &= \theta_j^{-1},\end{aligned}\right\} \text{for } \theta_j \neq \theta_i. \tag{5.4'}$$

By virtue of Eqs. (5.3), (5.4), and (5.4') we obtain

$$\theta_j\cdot 1(\theta_i) = (1-\delta_{ij})\theta_j, \tag{5.5}$$

$$\theta_j^{-1}\cdot 1(\theta_i) = (1-\delta_{ij})\theta_j^{-1}, \tag{5.6}$$

in which j is not a repetitive index.

From the above discussion we see that we can only find the inverse element of an "a" number, but do not find the unity number 1 within the realm of "a" numbers. The $1(\theta_i)$ must belong in the realm of "b" numbers. On this point there is a difference between the "b" number and the real (or complex) number in principle.

Now let us define division by many "a" numbers. The equality $\theta_1\theta_2/\theta_1\theta_2 = 1$ is used as a convention.

The definition can be written as follows:

$$\theta_1\theta_2/\theta_1\theta_2 = 1(\theta_1\theta_2), \tag{5.7}$$

or

$$1(\theta_1\theta_2) = \frac{\theta_1\theta_2}{\theta_1\theta_2} = \theta_1\theta_2\times(\theta_1\theta_2)^{-1} = \theta_1\theta_2\theta_2^{-1}\theta_1^{-1}$$
$$= \theta_1 1(\theta_2)\theta_1^{-1} = \theta_1\theta_1^{-1}\times 1(\theta_2) = 1(\theta_1)1(\theta_2), \quad (5.8)$$

so

$$1(\theta_1\theta_2) = 1(\theta_1)1(\theta_2). \quad (5.9)$$

Similarly we may define

$$\theta_1\theta_2\theta_3\theta_4/\theta_7\theta_8\theta_9 = \theta_1\theta_2\theta_3\theta_4\times(\theta_7\theta_8\theta_9)^{-1}$$
$$= \theta_1\theta_2\theta_3\theta_4\theta_9^{-1}\theta_8^{-1}\theta_7^{-1}. \quad (5.10)$$

Note that we use the equality $(\theta_i\theta_j\theta_k)^{-1} = \theta_k^{-1}\theta_j^{-1}\theta_i^{-1}$.

A rule for reduction of a fraction can be obtained by Eq. (5.8):

$$\frac{\overset{1(\theta_1)1(\theta_2)}{\not\theta_1\not\theta_2}}{\not\theta_1\not\theta_2} = 1(\theta_1)1(\theta_2). \quad (5.11)$$

From Eq. (5.10) we have

$$\theta^2/\theta = \theta^2\times\theta^{-1} = 0\cdot\theta^{-1} = 0. \quad (5.12)$$

From Eqs. (5.11) and (5.5),

$$\theta^2/\overset{1(\theta)}{\not\theta} = \theta\cdot 1(\theta) = 0. \quad (5.13)$$

Note that in calculating the reduction of a fraction we must write the $1(\theta_i)$ in the numerator of the fraction. If we leave out the $1(\theta_i)$, then $\theta^2/\not\theta = \theta$. This is certainly wrong.

If we write "c" for "real number", then

$$c/c = 1(c) = 1. \quad (5.14)$$

The equality $1(c) = 1$ shows that the unity number 1 in "c" numbers is independent of any number in the realm of "c" numbers; that is, $1(c_1) = 1(c_2) = \cdots = 1$. This is to distinguish from $1(\theta)$. So the rule for calculating $1(c\theta)$ is

$$1(c\theta_1\theta_2\theta_3) = 1(c)1(\theta_1)1(\theta_2)1(\theta_3)$$
$$= 1\cdot 1(\theta_1)1(\theta_2)1(\theta_3)$$
$$= 1(\theta_1)1(\theta_2)1(\theta_3). \quad (5.15)$$

VI. THE LIMITATIONS OF BEREZIN'S FORMULA

In 1965 Berezin[3] gave a formula to find the rule for the functional derivative of "a" numbers, i.e. (see Appendix A),

$$\frac{\delta(a(x_1)a(x_2)\cdots a(x_n))}{\delta a(x)}$$
$$= \delta(x-x_1)a(x_2)\cdots a(x_n)$$
$$-\delta(x-x_2)a(x_1)a(x_3)\cdots a(x_n)$$
$$+\cdots+(-1)^{n-1}\delta(x-x_n)a(x_1)\cdots a(x_{n-1}). \quad (6.1)$$

One observes that this rule cannot relate to the concept of ordinary derivatives. In this paper we recognize that after division by an "a" number is defined as above, then from the concept of ordinary derivatives we not only derive Eq. (6.1) directly but may also solve new problems which Eq. (6.1) cannot solve. For example, let the increment of $\psi(x)$ be

$$\Delta\psi(x) \equiv \xi\theta\psi(x), \quad (6.2)$$

where the "a" number ξ is independent of space-time and an infinitesimal quantity, while θ is also an "a" number independent of space-time. Consider the functional

$$f[\psi(x)] = \theta\psi(x), \quad (6.3)$$

in which $\theta\cdot\psi(x)$ are all "a" numbers. Then by Berezin's formula the functional derivative of Eq. (6.3) is

$$\frac{\delta f[\psi(x)]}{\delta\psi(y)} = \frac{\delta(\theta\psi(x))}{\delta\psi(y)} = -\theta\delta(x-y)\neq 0. \quad (6.4)$$

Under the conditions of Eq. (6.2), the increment Δf obviously has two "a" numbers, and is then equal to zero. So the result of Eq. (6.4) is wrong.

VII. FINDING A FUNCTIONAL DERIVATIVE FOR "a" NUMBERS FROM THE CONCEPT OF THE ORDINARY FUNCTIONAL DERIVATIVE

Now let us consider the functional derivative. Let there be a functional $f[\psi(x)]$. By the definition of variation[8] we have

$$\delta f[\psi(x)] = \int d^4y\,\frac{\delta f[\psi(x)]}{\delta\psi(y)}\,\delta\psi(y). \quad (7.1)$$

Let us divide the four-dimensional space-time into arbitrarily small cells of volume, with index i. Then $f[\psi(i)]$ is an ordinary function of $\psi(i)$,[9] and $\psi(j)$ is the value of ψ in the jth cell. Then we may obtain

$$\delta f[\psi(i)] = \sum_j \frac{\partial f[\psi(i)]}{\partial\psi(j)}\,\delta\psi(j)$$
$$= \sum_j \frac{1}{\Delta V_j}\frac{\partial f[\psi(i)]}{\partial\psi(j)}\,\delta\psi(j)\Delta V_j. \quad (7.2)$$

As $\Delta V_j\to 0$, Eq. (7.2) becomes Eq. (7.1). Thus we can obtain

$$\frac{\delta f[\psi(x)]}{\delta\psi(y)} = \lim_{\Delta V_j\to 0}\frac{1}{\Delta V_j}\frac{\partial f[\psi(i)]}{\partial\psi(j)}$$
$$= \lim_{\Delta V_j\to 0}\frac{\delta_{ij}}{\Delta V_j}\frac{\partial f[\psi(j)]}{\partial\psi(j)}$$
$$= \delta^4(x-y)\,\frac{\partial f[\psi(j)]}{\partial\psi(j)}, \quad (7.3)$$

where $\partial f[\psi(j)]/\partial\psi(j)$ is a derivative of an ordinary function, i.e.,

$$\frac{\partial f[\psi(j)]}{\partial\psi(j)} = \lim_{\Delta\psi(j)\to 0}\frac{\Delta f[\psi(j)]}{\Delta\psi(j)}. \quad (7.4)$$

Since we have the definition of division by an "a" number, it also holds true that the numerator and the denominator in Eq. (7.4) both contain "a" numbers. Thus we must interpret the equation as follows:

$$\frac{\Delta f[\psi(j)]}{\Delta\psi(j)} = \Delta f[\psi(j)]\times(\Delta\psi(j))^{-1}. \quad (7.5)$$

Now let us find the derivative for (6.3) under the conditions of (6.2). Note that $1(\theta)$ is a "b" number, and $\theta\times 1(\theta) = 0$. By means of Eqs. (7.3), (6.2), and (5.10) we can get

$$\frac{\delta f[\psi(x)]}{\delta\psi(y)}$$

$$= \delta^4(x-y)\,\frac{\partial f[\psi(j)]}{\partial\psi(j)}$$

$$= \delta^4(x-y)\,\frac{\partial(\theta\psi(j))}{\partial\psi(j)}$$

$$= \delta^4(x-y)\,\lim_{\Delta\psi(j)\to 0}\frac{\theta\Delta\psi(j)}{\Delta\psi(j)}$$

$$= \delta^4(x-y)\lim_{\Delta\psi_j\to 0}\theta\Delta\psi(j)\times(\Delta\psi(j))^{-1}$$

$$= \delta^4(x-y)\lim_{\xi\to 0}\theta\xi\theta\psi(j)\times(\xi\theta\psi(j))^{-1} \qquad (7.6)$$

$$= \delta^4(x-y)\lim_{\xi\to 0}\theta\xi\times\xi^{-1}\times\theta\times\theta^{-1}$$

$$\times\psi(j)\times\psi(j)^{-1}$$

$$= \delta^4(x-y)\theta\,1(\theta)1(\xi)1(\psi(j)) = 0\,. \qquad (7.7)$$

The result of (7.7) may be directly obtained from Eq. (7.6) since $\theta^2 = 0$.

VIII. A NEW FORMULA FOR THE FUNCTIONAL DERIVATIVE CONTAINING "a" NUMBERS

Let the continuous product of n "a" numbers be $(a_1(x_1)a_2(x_2)\cdots a_n(x_n))$, where each $a_i(x_i)$ may be independent or the increment $\Delta a_i(x_i)$ $(i = 1,2,...,n)$ is not independent. For example,

$$\Delta a_i(x_i) = \xi a_i(x_i)a_{i+p}(x_{i+p})a_{i-s}(x_{i-s}).$$

From (7.3) the functional derivative of this continuous product is

$$\frac{\delta(a_1(x_1)a_2(x_2)\cdots a_n(x_n))}{\delta a_k(x)}$$

$$= \delta_{1k}\delta(x-x_1)\,\frac{\partial a_1(x_1^j)}{\partial a_1(x_1^j)}\,a_2(x_2)a_3(x_3)\cdots a_n(x_n)$$

$$-\,\delta_{2k}\delta(x-x_2)a_1(x_1)\,\frac{\partial a_2(x_2^j)}{\partial a_2(x_2^j)}\,a_3(x_3)\cdots a_n(x_n)$$

$$+\cdots+(-1)^{n-1}\delta(x-x_n)\delta_{nk}a_1(x_1)$$

$$\times\cdots a_{n-1}(x_{n-1})\,\frac{\partial a_n(x_n^j)}{\partial a_n(x_n^j)}, \qquad (8.1)$$

where the index j in x_i^j represents the cell j of space-time x_i. From Eq. (5.1) $\partial a_i(x_i^j)/\partial a_i(x_i^j)$ will become (see Appendix B)

$$\frac{\partial a_i(x_i^j)}{\partial a_i(x_i^j)} = \lim_{\Delta a_i(x_i^j)\to 0}\frac{\Delta a_i(x_i^j)}{\Delta a_i(x_i^j)} = 1\left[\lim_{\Delta a_i(x_i^j)\to 0}\Delta a_i(x_i^j)\right]$$

$$\equiv 1[\Delta a_i(x_i)]. \qquad (8.2)$$

The last steps have been written in abbreviated form for convenience in writing. Then (8.1) becomes, in the notation of (8.2),

$$\frac{\delta(a_1(x_1)a_2(x_2)\cdots a_n(x_n))}{\delta a_k(x)}$$

$$= \delta_{1k}\delta(x-x_1)1[\Delta a_1(x_1)]a_2(x_2)\cdots a_n(x_n)$$

$$-\cdots+(-1)^{n-1}\delta_{nk}\delta(x-x_n)a_1(x_1)$$

$$\times\cdots a_{n-1}(x_{n-1})1[\Delta a_n(x_n)]. \qquad (8.3)$$

This is a new derivative formula that contains Berezin's formula. When $\Delta a_i(x_i)$ $(i = 1,2,...,n)$ and

$$a_1(x_1)a_2(x_2)\cdots a_{i-1}(x_{i-1})a_{i+1}(x_{i+1})\cdots a_n(x_n)$$

have no common "a" number factor, then from (5.5) we can obtain

$$a_1(x_1)\cdots a_{i-1}(x_{i-1})1[\Delta a_i(x_i)]\cdot a_{i+1}(x_{i+1})\cdots a_n(x_n)$$

$$= a_1(x_1)\cdots a_{i-1}(x_{i-1})a_{i+1}(x_{i+1})\cdots a_n(x_i). \qquad (8.4)$$

Substituting (8.4) into (8.3), Berezin's formula (6.1) can be obtained.

If $\Delta a_i(x_i)$ $(i = 1,2,3,..,n)$ and

$$a_1(x_1)\cdots a_{i-1}(x_{i-1})a_{i+1}(x_{i+1})\cdots a_n(x_n)$$

have a common "a" number factor, then from (5.5) and (5.6) the ith term in (8.3) is zero. This is the result of Sec. VI. Using the new formula (8.3) as the derivative for (6.3) under the conditions of (6.2), we can obtain

$$\frac{\delta(\theta\psi(x))}{\delta\psi(y)} = -\,\delta^4(x-y)\theta\,\frac{\partial\psi(j)}{\partial\psi(j)}$$

$$= -\,\delta^4(x-y)\theta\,1(\Delta\psi)$$

$$= -\,\delta^4(x-y)\theta\,1[\xi\theta\psi]$$

$$= -\,\delta^4(x-y)\theta\,1[\xi]1(\theta)1[\psi] = 0. \quad (8.5)$$

According to the traditional idea of real numbers the result for (8.5) can only be

$$\theta\,1[\xi]1(\theta)1[\psi] = \theta\cdot 1\cdot 1\cdot 1 = \theta.$$

Then $\delta(\theta\psi(x))/\delta\psi(y) = -\,\delta^4(x-y)\theta$. Clearly this is a wrong result.

IX. THE INFINITESIMAL CALCULUS OF "b" NUMBERS

Now we will find the functional derivative containing "b" numbers by analogy with the derivation of "a" numbers. Since $b^2 = 0$ and "b" numbers obey the commutative law, the sign of every term is positive. Then the following equality holds true:

$$\frac{\delta(b(x_1)b(x_2)\cdots b(x_n))}{\delta b(x)}$$

$$= \delta(x-x_1)1[\Delta b(x_1)]b(x_2)\cdots b(x_n)$$

$$+\cdots+\delta(x-x_n)b(x_1)\cdots b(x_{n-1})1[\Delta bx_n], \qquad (9.1)$$

in which the $1(b_i)$ is

$$b_i\times b_i^{-1} = 1[b_i]. \qquad (9.1')$$

Other definitions are analogous to those in Sec. V. For example,

$$1(b_i)\times b_j=(1-\delta_{ij})b_j,\quad \text{etc.} \tag{9.1''}$$

Concerning the integral of "b" numbers, the linear term of a "b" number occurs only because $b^2=0$, so we need only consider the integral $\int b\,db$. Let β be a "b" number constant. Because an integral is invariant under translational transformation, we can get

$$\int b\,db=\int(b+\beta)\,d(\beta+b)$$

$$=\int(b+\beta)db=\int b\,db+\int\beta\,db. \tag{9.2}$$

It is evident that

$$\int db=0. \tag{9.3}$$

Since the integral $\int b\,db$ is a "b" number constant, then let this "b" number equal 1. We have

$$\int b\,db=1. \tag{9.4}$$

For the integral of the product of many "b" numbers, we have following result:

$$\int b_ib_jb_k\cdots b_lb_m\,db_n$$

$$=b_jb_k\cdots b_lb_m\delta_{in}+b_ib_k\cdots b_lb_m\delta_{jn}$$

$$+\cdots+b_ib_j\cdots b_l\delta_{mn}. \tag{9.5}$$

X. APPLICATION OF THE NEW DERIVATIVE FORMULA IN GAUGE FIELD THEORY

In gauge field theory, when we deal with functional integrals containing "a" numbers the transformation of integral variables is often used. At the same time we must calculate the value of the Jacobian in the transformation. However, in finding some functional derivative using only Berezin's formula, we would obtain the wrong result. To correctly solve this problem we must use the new formula (8.3).

The BRS transformation is[10,11]

$$\psi(x)\to\psi'(x)=\psi(x)-i\xi\frac{\lambda^a}{2}C^a\psi(x),$$

$$\Delta\psi(x)=-i\xi\frac{\lambda^a}{2}C^a\psi(x),$$

$$\bar\psi(x)\to\bar\psi'(x)=\bar\psi(x)+i\bar\psi(x)\xi\frac{\lambda^a}{2}C^a,$$

$$\Delta\bar\psi(x)=i\bar\psi(x)\xi\frac{\lambda^a}{2}C^a,$$

$$A^a_\mu\to A^a_\mu=A^a_\mu-\frac{1}{g}\xi D^{ab}_\mu C_b,$$

$$\Delta A^a_\mu=-\frac{1}{g}\xi D^{ab}_\mu C_b,\quad \text{etc.}$$

To find the value of the Jacobian we must calculate $\delta\psi'(x)/\delta\psi(y)$, etc. From the new formula (8.3) we can obtain

$$\frac{\delta\psi'(x)}{\delta\psi(y)}=\delta^4(x-y)-i\xi\frac{\lambda^a}{2}C^a1[\Delta\psi]$$

$$=\delta^4(x-y)-i\xi\frac{\lambda^a}{2}C^a1\left[-i\xi\frac{\lambda^a}{2}C^a\psi\right]$$

$$=\delta^4(x-y)+i^2\xi\frac{\lambda^a}{2}C^a1(\xi)1\left(\frac{\lambda^a}{2}C^a\right)1(\psi)$$

$$=\delta^4(x-y)+i^2\xi\,1(\xi)\frac{\lambda^a}{2}C^a1\left(\frac{\lambda^a}{2}C^a\right)1(\psi)$$

$$=\delta^4(x-y),\quad \text{etc.} \tag{10.1}$$

However, according to Berezin's formula we can obtain

$$\frac{\delta\psi'(x)}{\delta\psi(y)}=\delta^4(x-y)-i\xi\frac{\lambda^a}{2}C^a. \tag{10.2}$$

This result is clearly wrong.

XI. CONCLUSION AND DISCUSSION

(1) Based on the existence of both fermions and antifermions in particle physics, we can assume that every "a" number has an inverse element "a^{-1}". They are both anticommutative numbers.

(2) Given the equation $\theta_i/\theta_i\equiv\theta_i\times\theta_i^{-1}\equiv 1(\theta_i)$, then according to the special property 3 of "b" numbers, $\theta_j\times 1(\theta_i)\neq 0$ and $\theta_j^{-1}\times 1(\theta_i)\neq 0$ $(\theta_j\neq\theta_i)$. What are the results of these products? The answer cannot be found by the special properties of the "b" number, but by virtue of the traditional concept of unity 1 we may define $\theta_j\times 1(\theta_i)=\theta_j$ and $\theta_j^{-1}\times 1(\theta_i)=\theta_j^{-1}$.

(3) However, when we calculate $\theta_i\times 1(\theta_i)$ and $\theta_i^{-1}\times 1(\theta_i)$, we must destroy the traditional concept of the unity 1. Because of the special properties of "b" numbers we know that $\theta_i\times 1(\theta_i)=\theta_i^{-1}\times 1(\theta_i)=0$.

(4) When we define division by an "a" number, sometimes we use the traditional concept of unity 1, and sometimes we destroy it. But we use the special properties of the "b" number from beginning to end. In other words, the associative law and the anticommutative law of the "a" number are used through the whole process of division.

(5) According to the definition of division by an "a" number we obtain the new derivation formula, which encompasses Berezin's formula as well.

(6) We introduced the "b" number, which may be useful in solving problems of supersymmetry.

APPENDIX A: INDEFINITE VALUE OF THE DERIVATIVE OF $f[\psi(j)]$

Berezin's formula of derivation is essentially equivalent to the definition

$$f[\psi(j)+\Delta\psi(j)]-f[\psi(j)]=\Delta\psi(j)\frac{\partial f}{\partial\psi(j)}, \tag{A1}$$

in which $f[\psi(j)]$ is the ordinary function of $\psi(j)$, and the derivation of $f[\psi(j)]$ is the coefficient of the linear term of $\Delta\psi(j)$. However, in the realm of "a" numbers the derivative defined above has an indefinite value. For example, if the increment $\Delta\psi(j)$ is

$$\Delta\psi(j) = \xi\theta\psi(j), \quad \text{(A2)}$$

in which $\xi\cdot\theta$ and $\psi(j)$ are all "a" numbers, ξ is infinitesimal, and the function containing "a" numbers is

$$f(\psi(j)) = \theta\psi(j), \quad \text{(A3)}$$

then according to the definition of the derivative in Eq. (1) under the condition of Eq. (2), the derivative of $f[\psi(j)]$ with respect to $\psi(j)$ is

$$\begin{aligned} &f[\psi(j) + \Delta\psi(j)] - f[\psi(j)] \\ &\quad = \theta(\psi(j) + \Delta\psi(j)) - \theta\psi(j) \\ &\quad = \theta\Delta\psi(j) = -\Delta\psi(j)\theta = \xi\theta\psi(j)(-\theta) = 0. \end{aligned} \quad \text{(A4)}$$

Since we have

$$0 = \xi\theta\psi(j)(a_1\theta) = \Delta\psi(j)(a_1\theta), \quad \text{(A5)}$$

$$0 = \xi\theta\psi(j)(a_2\psi(j)) = \Delta\psi(j)a_2\psi(j), \quad \text{(A6)}$$

in which a_1 and a_2 are arbitrary functions, then

$$\frac{\partial f}{\partial\psi(j)} = -\theta, \quad \text{(A7)}$$

$$\frac{\partial f}{\partial\psi(j)} = a_1\theta, \quad \text{(A8)}$$

$$\frac{\partial f}{\partial\psi(j)} = a_2\psi(j). \quad \text{(A9)}$$

Thus, according to the above definition, the derivative of $f[\psi(j)]$ has sometimes an indefinite value. This conclusion is right only for the realm of "a" numbers, not for real or complex numbers.

APPENDIX B: THE DEFINITION OF LEFT DIVISION

Since we use the left derivative in Berezin's formula, we should use the following definition of left division: equality:

$$\frac{\theta_i}{\theta_i} \equiv \theta_i^{-1}\times\theta_i \equiv 1(\theta_i); \quad \text{(B1)}$$

definition 1:

$$\left.\begin{aligned} \theta_j\times 1(\theta_i) &= \theta_j \\ \theta_j^{-1}\times 1(\theta_i) &= \theta_j^{-1} \end{aligned}\right\}, \quad \text{for } \theta_i\neq\theta_j; \quad \text{(B2)}$$

definition 2:

$$\begin{aligned} \theta_1\theta_2\theta_3/\theta_4\theta_5\theta_6 &= (\theta_4\theta_5\theta_6)^{-1}\times(\theta_1\theta_2\theta_3) \\ &= \theta_6^{-1}\theta_5^{-1}\theta_4^{-1}\times\theta_1\theta_2\theta_3. \end{aligned} \quad \text{(B3)}$$

From the special properties of "b" numbers we obtain

$$\theta_i\times 1(\theta_i) = 0, \quad \theta_i^{-1}\times 1(\theta_i) = 0. \quad \text{(B4)}$$

[1]H. G. Grassmann, *Linear Ausdehnungslehre* (Chelsea, New York, 1844) (reprint: Chelsea, New York, 1969).

[2]M. Kline, *Mathematical Thought from Ancient to Modern Times* (Oxford U.P., New York, 1972), Vol. 3, p. 181.

[3]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).

[4]R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. **56**, 81 (1975).

[5]J. Wess and B. Zumino, Nucl. Phys. B **70**, 310 (1974).

[6]A. Salam and J. Strasdee, Phys. Rev. D **11**, 1521 (1975).

[7]P. Fayet and Ferraras, Phys. Rep. C **32**, 249 (1977).

[8]N. N. Bogolubov and D. V. Shikov, *An Introduction to the Theory of Quantized Fields* (Moscow, Nauka, 1976).

[9]D. Lurie, *Particles and Fields* (Interscience, New York, 1968), pp. 63 and 64.

[10]C. Bechi, A. Rouet, and R. Stora, Commun. Math. Phys. **42**, 127 (1975).

[11]Hu Yao-Guang, *Gauge Field Theory* (Hua Dong Normal U.P., Shanghai, 1984).

The existence of regular partially future asymptotically predictable space-times

Andrzej Królak[a)]
Institute of Mathematics, Polish Academy of Sciences, Sniadeckich 8, 00-950 Warsaw, Poland

(Received 12 January 1988; accepted for publication 13 April 1988)

A new definition of asymptotically flat space-times is proposed. It is shown that singularities in these new space-times arise from regular initial data. This leads to a simplification of the assumptions of the cosmic censorship theorem recently proved by the author [Class. Quantum Gravit. **3**, 267 (1986)]. It is also shown that in the new asymptotically flat space-times a stronger censorship theorem than that proved so far by the author holds.

I. INTRODUCTION

Recently the present author has demonstrated the existence of Penrose's "cosmic censor" in a wide class of weakly asymptotically simple and empty space-times. The censorship theorem proved by the author (Ref. 1, Theorem 3.1) contains three classes of assumptions. The first class of assumptions consists of two standard conditions: the energy condition and the causality condition. The second class are the main assumptions of the theorem: conditions on space-time singularities and a further restriction on the global causal structure of space-time. There is also a third class of assumptions consisting of conditions ensuring that singularities occurring in space-time arise from regular initial data. These conditions do not appear explicitly in the statement of the theorem, but they are contained in the definition of the regular partially future asymptotically predictable space-times from a partial Cauchy surface $\mathscr{S}$ (Ref. 1, Definition 2.9). In this paper we shall define a new class of asymptotically flat space-times called regular weakly asymptotically simple and empty space-times. We shall prove that this definition ensures the existence of a partial Cauchy surface from which space-time is partially future asymptotically predictable. Thus with this new definition of asymptotically flat space-times our censorship theorem is considerably simplified.

In Sec. II, after recalling some basic notions, we shall introduce the definition of regular weakly asymptotically simple and empty space-times. Then by proving a number of lemmas and propositions we shall prove the existence in such a space-time of a partial Cauchy surface from which space-time is partially future asymptotically predictable. In Sec. III we shall show that with our new definition one can prove stronger versions of cosmic censorship: strong future asymptotic predictability and regular predictability.

II. REGULAR WEAKLY ASYMPTOTICALLY PREDICTABLE SPACE-TIMES

By space-time we shall mean a pair $(\mathscr{M},g)$, where $\mathscr{M}$ is a connected orientable four-dimensional Hausdorff C^∞ manifold and g is a C^∞ Lorentz metric on $\mathscr{M}$. Two space-times $(\mathscr{M},g)$ and $(\mathscr{M}',g')$ are said to be isometric if there is a diffeomorphism $\varphi: \mathscr{M} \to \mathscr{M}'$ that carries the metric g into the metric g', i.e., $\varphi_* g = g'$.

Before we can introduce the notion of the regular weakly asymptotically simple and empty space-time we shall need the concept of an asymptotically simple and empty space-time.

Definition 1: A space-time $(\mathscr{M},g)$ is said to be asymptotically simple and empty if there exists a strongly causal space-time $(\tilde{\mathscr{M}},\tilde{g})$ and an imbedding Θ: $\mathscr{M} \to \tilde{\mathscr{M}}$ that imbeds $\mathscr{M}$ as a manifold with smooth boundary $\partial\mathscr{M}$ in $\tilde{\mathscr{M}}$, such that

(1) there is a smooth function Ω on $\tilde{\mathscr{M}}$ such that on $\Theta(\mathscr{M})$ Ω is positive and $\Omega^2 g = \Theta(g)$;
(2) on $\partial\mathscr{M}$, $\Omega = 0$ and $d\Omega \neq 0$;
(3) every null geodesic in $\mathscr{M}$ has two end points on $\partial\mathscr{M}$;
(4) $\mathrm{Ricc}(g) = 0$ on an open neighborhood of $\partial\mathscr{M}$ in $\mathscr{M}\cup\partial\mathscr{M}$. We shall write $\bar{\mathscr{M}}$ for $\mathscr{M}\cup\partial\mathscr{M}$ and $\bar{g}$ for $\tilde{g}|\bar{\mathscr{M}}$.

The following theorem and lemma give the main properties of asymptotically simple and empty space-times.

Theorem 1: Let $(\mathscr{M},g)$ be an asymptotically simple and empty space-time. Let $\mathscr{I}^+ = \partial\mathscr{M}\cap I^+(\bar{\mathscr{M}},\bar{\mathscr{M}})$ and $\mathscr{I}^- = \partial\mathscr{M}\cap I^-(\bar{\mathscr{M}},\bar{\mathscr{M}})$. Then

(1) $d\Omega$ is null with respect to $\tilde{g}$ everywhere on $\partial\mathscr{M}$;
(2) $\partial\mathscr{M}$ is disconnected, its components are $\mathscr{I}^+$ and $\mathscr{I}^-$;
(3) $\mathscr{I}^+$ and $\mathscr{I}^-$ are achronal and without edge in $(\bar{\mathscr{M}},\bar{g})$, and are generated by inextendible null geodesics thereof;
(4) $(\mathscr{M},g)$ is globally hyperbolic, admitting a Cauchy surface diffeomorphic to R^3;
(5) $\mathscr{I}^+$ and $\mathscr{I}^-$ are diffeomorphic to $S^2\times R$;
(6) $(\bar{\mathscr{M}},\bar{g})$ is causally simple.

Lemma 1: Let $\mathscr{W}$ be a compact subset of an asymptotically simple and empty space-time $(\mathscr{M},g)$. Then the set $\dot{J}^+(\mathscr{W},\bar{\mathscr{M}})\cap\mathscr{I}^+$ is compact and acausal, and intersects every generator of $\mathscr{I}^+$.

Asymptotically simple and empty space-times defined above are space-times that may contain bounded objects such as stars that do not undergo gravitational collapse resulting in singularities. To be able to consider the general collapse situations with singularities, Hawking and Ellis introduced the notion of the weakly asymptotically simple and empty space-time.

Definition 2: A space-time $(\mathscr{M},g)$ is said to be weakly

[a)] Present address: Department of Applied Mathematics and Astronomy, University College Cardiff, CFI 1XL, United Kingdom.

asymptotically simple and empty if there exists an open set $\mathscr{U}$ in $\mathscr{M}$ and an extension $(\mathscr{M}',g')$ of $(\mathscr{U},g/\mathscr{U})$ such that

(1) $(\mathscr{M}',g')$ is asymptotically simple and empty;

(2) $\mathscr{U}$ is a neighborhood of $\partial\mathscr{M}'$ in $\mathscr{M}'$.

Since the cosmic censorship hypothesis concerns only singularities that arise to the future of a regular initial surface we have introduced the following class of space-times.

Definition 3 (Ref. 1, Definition 2.9): A weakly asymptotically simple and empty space-time is said to be regular partially future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ if

(1) $\mathscr{J}^+\subset J^+(\mathscr{S},\overline{\mathscr{M}})$,

(2) $\overline{D^+(\mathscr{S},\overline{\mathscr{M}})}\cap\lambda\neq\varnothing$ for all generators λ of $\mathscr{J}^+$,

(3) for any past-endless causal curve γ in $J^+(\mathscr{S})$, there exists a point s in $\mathscr{S}$ such that $\gamma\subset I^+(s)$.

The necessity of all the above conditions to ensure the existence of regular initial data has been thoroughly discussed in Ref. 1. Here we shall propose a new class of asymptotically flat space-times called regular, weakly, asymptotically simple and empty. Then we shall establish that this new class of space-times admits a partial Cauchy surface from which space-time is partially future asymptotically predictable.

Definition 4: A weakly asymptotically simple and empty space-time is said to be regular, weakly, asymptotically simple and empty if

(1) for every point $p\in\mathscr{M}'-\mathscr{U}$, the set $\mathscr{M}'-(\mathscr{U}\cup I^+(p,\mathscr{M}'))$ is compact,

(2) every achronal set of $(\mathscr{M}',g)$ intersects $\mathscr{U}$ in an achronal set of $(\mathscr{M},g)$.

The importance of conditions (1) and (2) in any definition of an asymptotically flat space-time has been underlined by Newman.[2] The particular form of condition (1) is due to Clarke and de Felice.[3] Condition (1) ensures that the set $\mathscr{U}$ is a good neighborhood of both null infinity $\mathscr{J}^+\cup\mathscr{J}^-$ and the spacelike infinity. Condition (2) means that global causal structures of space-times $(\mathscr{M},g)$ and $(\mathscr{M}',g')$ are compatible.

We shall prove the main result of our paper by establishing a series of propositions. First, we shall prove the existence of a suitable partial Cauchy surface $\mathscr{S}$ in a regular, weakly, asymptotically predictable space-time and then we shall prove that space-time is regular partially future asymptotically predictable from $\mathscr{S}$ by demonstrating three propositions, each establishing one of the conditions of Definition 3.

Proposition 1: Let $(\mathscr{M},g)$ be a regular, weakly, asymptotically simple and empty space-time; then there exist a partial Cauchy surface $\mathscr{S}$ and a compact set $\mathscr{K}$ in $\mathscr{M}$ such that $\mathscr{K}\subset\mathscr{S}\subset\mathscr{U}$, $\mathscr{M}'-\mathscr{U}\subset I^+(\mathscr{K},\mathscr{M}')$, and $\mathscr{S}$ is a Cauchy surface in $\mathscr{M}'$.

Proof: Let $(\mathscr{M}',g')$ be the asymptotically simple and empty space-time given in Definition 1. Since $(\mathscr{M}',g')$ is globally hyperbolic it contains a sequence of Cauchy surfaces $\{\mathscr{S}'_n\}$ with $\mathscr{S}'_n\subset I^+(\mathscr{S}_{n+1},\mathscr{M}')$ and $\cup_n I^+(\mathscr{S}'_n,\mathscr{M}')=\mathscr{M}'$. Let p be any point in $\mathscr{M}'-\mathscr{U}$. The sets $I^+(\mathscr{S}'_n,\mathscr{M}')$ cover the compact set $\mathscr{M}'-(\mathscr{U}\cup I^+(p,\mathscr{M}'))$, and so there exists an m with $\mathscr{M}'-(\mathscr{U}\cup I^+(p,\mathscr{M}'))\subset I^+(\mathscr{S}'_m,\mathscr{M}')$. From global hyperbolicity of $(\mathscr{M}',g')$ it follows that the set $\mathscr{K}=\mathscr{S}'_m\cap I^-(\mathscr{M}'-(\mathscr{U}\cup I^+(p,\mathscr{M}')))$ is compact. Writing $\mathscr{S}$ for $\mathscr{S}'_m$ gives the desired result.

Proposition 2: Let $(\mathscr{M},g)$ be a regular, weakly, asymptotically simple and empty space-time and let $\mathscr{S}$ be the partial Cauchy surface given in Proposition 1. Then $\overline{D^+(\mathscr{S},\overline{\mathscr{M}})}\cap\lambda\neq\varnothing$ for all null geodesic generators of $\mathscr{J}^+$.

Proof: Let $\mathscr{K}$ be the compact set on $\mathscr{S}$ given in Proposition 1. Let λ be a null geodesic generator of $\mathscr{J}^+$. Since $\mathscr{M}-\mathscr{U}\subset I^+(\mathscr{K},\overline{\mathscr{M}})$ the boundary $\dot{I}^+(\mathscr{K},\overline{\mathscr{M}})$ is contained in $\mathscr{U}$. Hence it is also contained in the asymptotically simple empty space-time $(\mathscr{M}',g')$. Thus by Lemma 1, λ must intersect the boundary $\dot{I}^+(\mathscr{K},\overline{\mathscr{M}})$ and leave $I^+(\mathscr{K},\mathscr{M})$. Let q be any point on λ not in $J^+(\mathscr{K},\overline{\mathscr{M}})$. Then $I^-(q,\overline{\mathscr{M}})\cap\mathscr{M}$ belongs to $\mathscr{U}$ and hence to $\mathscr{M}'$. Thus all the past-directed timelike curves from q must intersect $\mathscr{S}$ since $\mathscr{S}$ is a Cauchy surface in $\mathscr{M}'$. Hence by definition $q\in\overline{D^+(\mathscr{S},\overline{\mathscr{M}})}$. Consequently $\lambda\cap\overline{D^+(\mathscr{S},\overline{\mathscr{M}})}\neq\varnothing$.

Proposition 3: Let the conditions of Proposition 2 hold. Then $\mathscr{J}^+\subset J^+(\mathscr{S},\overline{\mathscr{M}})$.

Proof: The proof follows immediately from the fact that $\mathscr{S}$ is a Cauchy surface for $\mathscr{M}'$. Hence $\mathscr{J}^+\subset J^+(\mathscr{S},\overline{\mathscr{M}})$ by the definition of the domain of dependence $D(\mathscr{S},\overline{\mathscr{M}})$.

Proposition 4: Let the conditions of Proposition 2 hold. Let γ be a past-endless causal curve in $J^+(\mathscr{S})$. Then there exists a point s in $\mathscr{S}$ such that $\gamma\subset I^+(s)$.

Proof: First we show that γ cannot enter and remain in $\mathscr{U}$. Otherwise there would be a past endless curve in $\mathscr{M}'$ (as $\mathscr{U}\subset\mathscr{M}'$) that does not intersect $\mathscr{S}$. This would be impossible since $\mathscr{S}$ is a Cauchy surface in $\mathscr{M}'$. Thus γ must enter and remain in $\mathscr{M}-\mathscr{U}$.

Consider a sequence of points q_i on γ in $\mathscr{M}-\mathscr{U}$ satisfying $q_{i+1}\in I^-(q_i)$ for all i, and having no limit point in $\mathscr{M}$. Let $\mathscr{K}$ be the compact set on $\mathscr{S}$ given in Proposition 1. All the q_i lie in $\mathscr{M}\cap I^+(\mathscr{K},\overline{\mathscr{M}})$ since $\mathscr{M}-\mathscr{U}\subset I^+(\mathscr{K},\overline{\mathscr{M}})$. Consider the sequence μ_i of endless timelike curves such that each μ_i passes through q_i and intersects $\mathscr{K}$ in a point s_i. Since $\mathscr{K}$ is compact the sequence of points s_i has the limit point s in $\mathscr{K}$. Consider the chronological future $I^+(s)$ of s. Since s is the limit point of the timelike curves μ_i, all the μ_i for i greater than some I intersect $I^+(s)$. Thus all q_i for $i>I$ are in $I^+(s)$. It follows that $\gamma\subset I^+(s)$.

By the above three propositions we have established the following result.

Theorem 2: Let $(\mathscr{M},g)$ be a regular, weakly, asymptotically simple and empty space-time; then there exists a partial Cauchy surface $\mathscr{S}$ such that $(\mathscr{M},g)$ is regular partially future asymptotically predictable from $\mathscr{S}$.

III. CENSORSHIP THEOREMS

With the new definition of regular, weakly, asymptotically simple and empty space-times the statement of our censorship theorem given in Ref. 1 is as follows.

Theorem 3: Let the space-time $(\mathscr{M},g)$ be regular, weakly, asymptotically simple and empty. Then there exists a partial Cauchy surface $\mathscr{S}$ such that $(\mathscr{M},g)$ is future asymptotically predictable from $\mathscr{S}$ if the following conditions hold:

(1) $R_{ab}K^aK^b \geq 0$ for every null vector K^a;

(2) the strong causality condition holds on $(\mathscr{M},g)$;

and either

(3) the simplicity and the strong curvature conditions hold,

or

(3b) the trapped surface condition holds.

For definitions and discussion of various terms in the above theorem, see Ref. 1.

In regular, weakly, asymptotically simple and empty space-times we can prove stronger versions of cosmic censorship: strong future asymptotic predictability and regular predictability. We shall first define these conditions and then we shall establish a censorship theorem.

Definition 5 (Ref. 4)*:* A weakly asymptotically simple and empty space-time $(\mathscr{M},g)$ is said to be future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ in $\mathscr{M}$ if $\mathscr{I}^+ \subset \overline{D^+(\mathscr{S},\overline{\mathscr{M}})}$.

Definition 6 (Ref. 5)*:* A space-time $(\mathscr{M},g)$ future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ is said to be strongly future asymptotically predictable from $\mathscr{S}$ if $J^+(\mathscr{S}) \cap \overline{J^-(\mathscr{I}^+,\overline{\mathscr{M}})}$ is contained in $D^+(\mathscr{S})$.

Definition 7 (Ref. 6)*:* A space-time $(\mathscr{M},g)$ strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ is said to be a regular predictable space if

(α) $\mathscr{S} \cap J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ is homeomorphic to R^3_- (an open set with compact closure);

(β) $\mathscr{S}$ is simply connected;

(γ) for sufficiently large τ, $\mathscr{S}(\tau) \cap J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ is contained in $J^+(\mathscr{I}^-,\overline{\mathscr{M}})$.

This $\mathscr{S}(\tau)$ is a slicing of $D^+(\mathscr{S})$ constructed in Proposition 9.2.3 of Ref. 7.

For the discussion of the above conditions we refer the reader to Chap. 9 of Ref. 7.

Theorem 4: Let the conditions of Theorem 3 hold; then the weakly asymptotically simple and empty space-time $(\mathscr{M},g)$ is strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$.

Proof: The difference between the future predictability and strong future predictability is that the latter requires that the closure of $J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ as well as $J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ to the future of $\mathscr{S}$ belong to the domain of dependence of $\mathscr{S}$. Scrutiny of the definitions and proof of Theorem 3.1 in Ref. 1 show that the arguments of the proof of that theorem establish strong predictability as well.

Proposition 5: Let $(\mathscr{M},g)$ be a regular weakly asymptotically simple and empty space-time strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ constructed in Proposition 1. Then $(\mathscr{M},g)$ is a regular predictable space.

Proof: Let us consider the intersection $\mathscr{S} \cap \overline{J^-(\mathscr{I}^+,\overline{\mathscr{M}})}$. It cannot be empty since by Proposition 3 $\mathscr{I}^{+-} \subset J^+(\mathscr{S},\overline{\mathscr{M}})$. Let us consider the open set $\mathscr{B} := \mathscr{S} - J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ and suppose that $\mathscr{B} \neq \varnothing$. Let $\mathscr{K}$ be the compact set constructed in Proposition 1. Since $\mathscr{M}' - \mathscr{U} \subset I^+(\mathscr{K},\mathscr{M}')$ and since all the future-directed nonspacelike curves with past end point at $\mathscr{B}$ must enter $\mathscr{M}' - \mathscr{U}$ or otherwise they would have future end point on $\mathscr{I}^+$, which is impossible, the set $\mathscr{B}$ must be contained in $\mathscr{K}$. Hence the closure $\overline{\mathscr{B}}$ is compact as $\mathscr{K}$ is a compact set. Since $\mathscr{S}$ is a Cauchy surface in the asymptotically simple and empty space-time $(\mathscr{M},g')$, it is topologically R^3 by Theorem 1(4). Therefore $\mathscr{S} \cap J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ is homeomorphic to R^3_- (an open set with a compact closure) or to R^3, if the set $\mathscr{B}$ is empty. Thus condition (α) holds. Since $\mathscr{S}$ is homeomorphic to R^3, it is simply connected and consequently condition (β) is fulfilled. Since $\mathscr{S} \subset J^+(\mathscr{I}^-,\overline{\mathscr{M}})$ as $(\mathscr{M},g)$ has an asymptotically simple past it follows that $\mathscr{S}(\tau) \cap J^-(\mathscr{I}^+,\overline{\mathscr{M}})$ is contained in $J^+(\mathscr{I}^-,\overline{\mathscr{M}})$ for all τ and thus condition (γ) holds as well.

ACKNOWLEDGMENTS

I am grateful to the Department of Applied Mathematics and Astronomy, University College Cardiff, and in particular to Professor B. F. Schutz, for their hospitality. I would like to thank Dr. St. Bazański, Professor C. J. S. Clarke, and Dr. R. P. A. C. Newman for helpful discussions.

[1]A. Królak, Class. Quantum Grav. **3**, 267 (1986).

[2]R. P. A. C. Newman, Gen. Relativ. Gravit. **13**, 641 (1983).

[3]C. J. S. Clarke and F. de Felice, Gen. Relativ. Gravit. **16**, 139 (1984).

[4]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973), p. 310.

[5]See Ref. 2, p. 313.

[6]See Ref. 2, p. 318.

[7]See Ref. 2, p. 222.

Cohomology of the structure sheaf of real and complex supermanifolds

Claudio Bartocci and Ugo Bruzzo
Dipartimento di Matematica, Università di Genova, Via L. B. Alberti 4, 16132 Geneva, Italy

(Received 5 June 1987; accepted for publication 9 March 1988)

The cohomology of the structure sheaf of real and complex supermanifolds is studied. It is found to be nontrivial (also in the real case), unless the supermanifold is De Witt, i.e., it is a fiber bundle on an ordinary manifold with a vector fiber. As a consequence, the Dolbeault theorem can be extended only to complex De Witt supermanifolds. The relevance of the nontrivial cohomology of the structure sheaf to the classification of complex line superbundles is discussed. The relationship between the Picard group of a complex De Witt supermanifold and the Picard group of its body is shown.

I. INTRODUCTION

In this paper we study the cohomology of a supermanifold M with values in the structure sheaf $\mathscr{G}$ of M. Supermanifolds are considered in the sense of De Witt and Rogers,[1,2] i.e., they are a manifold whose coordinate maps take values in an exterior algebra. Our interest in studying the cohomological properties of supermanifolds stems from possible applications to field and string theory, e.g., in connection with anomalies in supersymmetric field theories and in superstring theories, or related to the extension to supermanifolds of the complex manifold techniques currently used in superstring theory.

In a previous article[3] (see also Ref. 4) we investigated the de Rham cohomology of the complex of superdifferentiable forms on a supermanifold M (SDR cohomology). In the case of a De Witt supermanifold (a fiber bundle on an ordinary manifold), SDR cohomology is trivially equivalent to de Rham cohomology of the base manifold. In the general case, SDR cohomology is different from the ordinary de Rham cohomology, and, indeed, it comes out that it is neither a topological nor a differentiable invariant, while it is a superdifferentiable invariant. This state of affairs is basically due to the fact that the cohomology of the structure sheaf $\mathscr{G}$ of a supermanifold is in general not trivial, contrary to what happens in the case of real manifolds.

In this paper we develop some basic techniques to study the cohomology $H^*(M,\mathscr{G})$ of the structure sheaf $\mathscr{G}$ of a supermanifold M. Apart from the trivial case of De Witt supermanifolds, $H^*(M,\mathscr{G})$ does not vanish. We show that the cohomology of $\mathscr{G}$ can be computed in terms of the Čech cohomology of a "good" cover of M. The cohomology of the structure sheaf of complex supermanifolds is introduced and a generalization of Dolbeault theorem is proved to hold in the case of complex De Witt supermanifolds. As an application, we introduce the Chern class of complex line superbundles and study the Picard group of real and complex supermanifolds.

II. COHOMOLOGY OF SHEAVES

For the reader's convenience, in this section we recall some basic definitions and results in sheaf theory.

Sheaves[5]*:* Let X be a topological space. A *sheaf* $\mathscr{F}$ of Abelian groups on X is a correspondence that to each open set U in X assigns an Abelian group $\mathscr{F}(U)$, called the group of *sections* of $\mathscr{F}$ over U, so as to verify the following properties.

(i) For any inclusion of open sets $V \subset U$ there exists a group morphism ρ^U_V: $\mathscr{F}(U) \to \mathscr{F}(V)$, called restriction morphism, such that $\rho^U_U = \mathrm{id}$.

(ii) If $W \subset V \subset U$ are inclusions of open sets, then $\rho^U_V \circ \rho^V_W = \rho^U_W$.

(iii) If $\{U_i, i \in I\}$ are open sets in X, $U = \cup_{i\in I} U_i$, and $s,t \in \mathscr{F}(U)$ are such that $\rho^U_{U_i}(s) = \rho^U_{U_i}(t)$, $\forall i \in I$, then $s = t$.

(iv) If $\{U_i, i \in I\}$ are as above, and a collection $\{s_i \in \mathscr{F}(U_i),\ i \in I\}$ is given such that $\rho^{U_i}_{U_i \cap U_j}(s_i) = \rho^{U_j}_{U_i \cap U_j}(s_j)$, then there exists a section $s \in \mathscr{F}(U)$ such that $\rho^U_{U_i}(s) = s_i$.

The *stalk* $\mathscr{F}_x$ of $\mathscr{F}$ at $x \in M$ is defined as the direct limit of the $\mathscr{F}(U)$'s over all open neighborhoods U of x. $\mathscr{F}_x$ is an Abelian group, and its elements are called the *germs of sections* of $\mathscr{F}$ at x.

In the following, Γ will denote the functor that to any sheaf associates the group of its global sections, i.e., $\Gamma\mathscr{F} = \mathscr{F}(X)$; Γ is left exact, but not exact.

One can define sheaves not only of Abelian groups, but also of other algebraic objects, such as sets, rings, modules, etc. However, if not otherwise stated, by "sheaf" we shall mean "sheaf of Abelian groups."

Sheaf cohomology: A sheaf $\mathscr{F}$ is said to be *injective* if, given any exact sequence of sheaves[5] $0 \to \mathscr{F}' \to \mathscr{F}''$, together with a morphism $\mathscr{F} \to \mathscr{F}'$, there is a morphism $\mathscr{F} \to \mathscr{F}''$ such that the diagram

$$\begin{array}{ccccc} 0 & \to & \mathscr{F}' & \to & \mathscr{F}'' \\ & & \uparrow & \nearrow & \\ & & \mathscr{F} & & \end{array}$$

commutes. Given a sheaf $\mathscr{F}$ on a topological space X, one can always find an *injective resolution* $\mathscr{L}^*$ of $\mathscr{F}$, namely, an exact sheaf sequence

$$0 \to \mathscr{F} \to \mathscr{L}^0 \to \mathscr{L}^1 \to \cdots, \tag{2.1}$$

where all the $\mathscr{L}$'s are injective. Applying the functor Γ, the resulting sequence of Abelian groups

$$0 \to \Gamma\mathscr{F} \to \Gamma\mathscr{L}^0 \to \Gamma\mathscr{L}^1 \to \cdots \tag{2.2}$$

 0022-2488/88/081789-06$02.50

is no longer exact. The cohomology of the complex (2.2), denoted by

$$H^p(X,\mathscr{F}) \equiv \frac{\ker(\Gamma\,\mathscr{L}^p \to \Gamma\mathscr{L}^{p+1})}{\operatorname{Im}(\Gamma\mathscr{L}^{p-1} \to \Gamma\,\mathscr{L}^p)}, \quad \text{if } p>0;$$

$$H^0(X,\mathscr{F}) = \ker(\Gamma\,\mathscr{L}^0 \to \Gamma\,\mathscr{L}^1),$$

is called the *cohomology of X with values in the sheaf $\mathscr{F}$*. Left exactness of Γ implies $H^0(X,\mathscr{F}) = \Gamma\,\mathscr{F}$.

It can be proved that this cohomology does not depend on the particular injective resolution. Whenever a sheaf $\mathscr{F}$ has trivial cohomology, in the sense that $H^p(X,\mathscr{F}) = 0$ for $p>0$, it is said to be *acyclic.*

Finally, we have the following important result.

Theorem 2.1: If

$$0 \to \mathscr{F} \to \mathscr{G} \to \mathscr{A} \to 0$$

is an exact sequence of sheaves on X, there is a long exact sequence in sheaf cohomology

$$\begin{aligned} 0 &\to H^0(X,\mathscr{F}) \to H^0(X,\mathscr{G}) \to H^0(X,\mathscr{A}) \\ &\overset{\partial}{\to} H^1(X,\mathscr{F}) \to \cdots \\ &\to H^p(X,\mathscr{F}) \to H^p(X,\mathscr{G}) \to H^p(X,\mathscr{A}) \\ &\overset{\partial}{\to} H^{p+1}(X,\mathscr{F}) \to \cdots, \end{aligned}$$

where the ∂'s are the so-called connecting morphisms.[5]

Čech cohomology[5]*:* Let $\mathscr{F}$ be a sheaf on a topological space X, and $\mathscr{U} = \{U_\alpha, \alpha \in J\}$ an open cover of X, with J an ordered set; for all $\alpha_0 \cdots \alpha_p \in J$, define $U_{\alpha_0\cdots\alpha_p} = U_{\alpha_0} \cap \cdots \cap U_{\alpha_p}$. Then define the complex of Abelian groups $C^*(\mathscr{U},\mathscr{F})$ whose pth term is

$$C^p(\mathscr{U},\mathscr{F}) = \prod_{\alpha_0 < \cdots < \alpha_p} \mathscr{F}(U_{\alpha_0\cdots\alpha_p})$$

and a differential operator δ: $C^p(\mathscr{U},\mathscr{F}) \to C^{p+1}(\mathscr{U},\mathscr{F})$ as follows:

$$(\delta f)_{\alpha_0\cdots\alpha_{p+1}} = \sum_{k=0}^{p+1} (-1)^k f_{\alpha_0\cdots\hat{\alpha}_k\cdots\alpha_{p+1}},$$

where the caret denotes omission. The *Cech cohomology of X with values in $\mathscr{F}$ with respect to the cover $\mathscr{U}$* is defined as the cohomology of the differential complex $(C^*(\mathscr{U},\mathscr{F}),\delta)$. There are natural group morphisms

$$\check{H}^p(\mathscr{U},\mathscr{F}) \to H^p(X,\mathscr{F}) \tag{2.3}$$

of the Čech cohomology into the sheaf cohomology of $\mathscr{F}$.[5]

Theorem 2.2[5]**:** A sufficient condition for the morphism (2.3) to be one to one is that for all nonvoid intersections $U_{\alpha_0\cdots\alpha_q}$, the restricted sheaf $\mathscr{F}|U_{\alpha_0\cdots\alpha_q}$ is acyclic. □

The direct limit of the groups $\check{H}^*(\mathscr{U},\mathscr{F})$ over all the open covers of X gives the Čech cohomology of X with values in $\mathscr{F}$, denoted by $\check{H}^*(X,\mathscr{F})$ (this involves set-theoretical subtleties[5]).

Theorem 2.3[5]**:** If X is paracompact, and $\mathscr{F}$ is a sheaf on X, the Čech and sheaf cohomologies of X with values in $\mathscr{F}$ are isomorphic.

III. SUPERMANIFOLDS

Let us denote by B_L the exterior algebra over $\mathbb{R}^L$, $L<\infty$. Here B_L has a natural $\mathbb{Z}_2$ gradation $B_L = (B_L)_0 \oplus (B_L)_1$ (by "graded" will shall always mean $\mathbb{Z}_2$ graded). A graded vector space basis for B_L can be indexed by M_L, the set of strictly increasing sequences of integers $\mu = \{0 \leqslant \mu_1 < \cdots < \mu_s \leqslant L\}$: if $\{e_1 \cdots e_L\}$ generate B_L, then $\beta_\mu = e_{\mu_1} \wedge \cdots \wedge e_{\mu_s}$; we shall also set $\beta_0 = 1$. Here B_L, equipped with the wedge product, is a graded commutative algebra (in the following, the wedge product symbol will be omitted). Let N_L denote the ideal of nilpotents of B_L, so that $B_L = \mathbb{R} \oplus N_L$. The projection $B_L \to \mathbb{R}$ is usually called *body map* and will be here denoted by σ. The Cartesian product $(B_L)^{m+n}$ can be turned into a graded B_L module by setting

$$(B_L)^{m+n} = B_L^{m,n} \oplus B_L^{\bar{m},\bar{n}}, \quad B_L^{m,n} = (B_L)_0^m \times (B_L)_1^n,$$

$$B_L^{\bar{m},\bar{n}} = (B_L)_1^m \times (B_L)_0^n.$$

A body map $\sigma^{m,n}$: $B_L^{m,n} \to \mathbb{R}^m$ is defined by letting $\sigma^{m,n}(x^1 \cdots x^m, y^1 \cdots y^n) = (\sigma(x^1) \cdots \sigma(x^m))$. $B_L^{m,n}$ will be equipped with its vector space topology.

Following Rogers,[6] we introduce on $B_L^{m,n}$ a distinguished sheaf of B_L-valued functions. Let $\mathscr{C}[V;Q]$ denote the sections over $V \subset X$ of the sheaf of Q-valued C^∞ functions on a manifold X. Let U be an open set in $\mathbb{R}^m$, L and L' two positive integers with $L' \leqslant L$, and define a map

$$Z_{L',L}\colon \mathscr{C}[U;B_{L'}] \to \mathscr{C}[(\sigma^{m,0})^{-1}(U); B_L]$$

according to

$$\begin{aligned} Z_{L',L}(f)(x^1 \cdots x^m) &= \sum_{i_1 \cdots i_m = 0}^{L} \frac{1}{i_1! \cdots i_m!} (\partial_1^{i_1} \cdots \partial_m^{i_m} f)|(\sigma(x^1) \cdots \sigma(x^m)) \\ &\quad \times s(x^1)^{i_1} \cdots s(x^m)^{i_m}. \end{aligned} \tag{3.1}$$

The image of the injective map $Z_{L',L}$ will be denoted by $\mathscr{G}[(\sigma^{m,0})^{-1}(U)]$; it consists of the GH^∞ functions of even variables on $(\sigma^{m,0})^{-1}(U)$. The ring $\mathscr{G}[(\sigma^{m,n})^{-1}(U)]$ of GH^∞ functions of even and odd variables on the open set $(\sigma^{m,n})^{-1}(U)$, where U is open in $\mathbb{R}^m$, is formed by elements of the type

$$F(x^1 \cdots x^m, y^1 \cdots y^n) = \sum_{\mu \in M_n} F_\mu(x^1 \cdots x^m)\, y^\mu, \tag{3.2}$$

where $y^\mu = y^{\mu_1} \cdots y^{\mu_s}$ and $F_\mu \in \mathscr{G}[(\sigma^{m,0})^{-1}(U)]$. The derivatives of F are uniquely determined by a Taylor expansion, provided that $L - L' \geqslant n$.[6] In the following, we shall always assume that this condition is fulfilled.

If, for any open set V in $B_L^{m,n}$, we let

$$\mathscr{G}(V) = \mathscr{G}[(\sigma^{m,n})^{-1}\sigma^{m,n}(V)],$$

we define on $B_L^{m,n}$ a sheaf $\mathscr{G}$ of graded $B_{L'}$ modules, whose sections are the GH^∞ functions on $B_L^{m,n}$. Here $\mathscr{G}$ is apparently not soft (i.e., a section over a closed subset of $B_L^{m,n}$ is not necessarily extendable to all of $B_L^{m,n}$). As a consequence, $B_L^{m,n}$ (and, *a fortiori*, any supermanifold) has no GH^∞ partitions of unity.

Definition 3.1: An (m,n)-dimensional GH^∞ supermanifold is a Hausdorff, second countable topological space M

together with an atlas $\mathscr{F} = \{(U_\alpha, \psi_\alpha) | \psi_\alpha: U_\alpha \to B_L^{m,n}\}$ such that the transition functions are GH^∞ mappings. □

The sheaf of GH^∞ functions on M (the *structure sheaf* of M) will be again denoted by $\mathscr{G}$. The rest of this paper is devoted to the study of the cohomology of $\mathscr{G}$.

In order to deal with complex supermanifolds we consider the tensor product $C_L = B_L \otimes \mathbb{C}$. C_L is a complex $\mathbb{Z}_2$-graded commutative algebra. A body map σ: $C_L \to \mathbb{C}$ is defined by $\sigma(a \otimes z) = \sigma(a) \otimes z$. The complex vector superspaces $C_L^{m,n}$ are defined in the obvious way. A function f: $\mathbb{C}^m \to C_L$ is said to be complex analytic if $f(z^1 \cdots z^m) = f^\mu(z^1 \cdots z^m)\, \beta_\mu$, with $\{\beta_\mu\}$ a basis in B_L and all the f^μ complex analytic. Complex superanalytic functions F: $C_L^{m,n} \to C_L$ are introduced starting from complex analytic mappings f: $\mathbb{C}^m \to C_{L'}$ in the same way as GH^∞ functions are introduced in terms of C^∞ mappings f: $\mathbb{R}^m \to B_{L'}$. A complex supermanifold is characterized as in Definition 3.1, with $B_L^{m,n}$ and GH^∞, respectively, replaced by $C_L^{m,n}$ and "complex superanalytic."

IV. COHOMOLOGY OF THE STRUCTURE SHEAF OF REAL SUPERMANIFOLDS

In the study of the cohomology of the structure sheaf of a supermanifold, a central role is played by the so-called De Witt supermanifolds. A supermanifold M is De Witt if it has an atlas $\{(\psi_\alpha, U_\alpha)\}$ such that the sets $\psi_\alpha(U_\alpha) \subset B_L^{m,n}$ have the form $\psi_\alpha(U_\alpha) = \sigma^{m,n}(\psi_\alpha(U_\alpha)) \times P_L^{m,n}$, where $P_L^{m,n} = P_L^m \times (B_L)_1^n$, P_L being the ideal of nilpotents in $(B_L)_0$. It is easily shown that an (m,n)-dimensional De Witt supermanifold M is a locally trivial C^∞ bundle Φ: $M \to M_0$ over an m-dimensional real differentiable manifold M_0, with typical fiber $P_L^{m,n}$. The real manifold M_0 is usually called the *body* of M. It should be noted that M need not be a vector bundle, i.e., its transition functions are not necessarily vector space morphisms.

Theorem 4.1: The structure sheaf $\mathscr{G}$ of a De Witt supermanifold M is acyclic.

The proof of this theorem will require a few lemmas.

Lemma 4.1: Let X and Y be connected topological spaces, Σ a ring with identity, $\mathscr{A}$ a sheaf of Σ modules on X, $\mathscr{K}$ the constant sheaf on Y with stalk a free Σ-module K. Then

$$H^*(X \times Y, \mathscr{A} \hat{\otimes} \mathscr{K}) \simeq H^*(X, \mathscr{A}) \otimes_\Sigma K,$$

where $\hat{\otimes}$ denotes the total tensor product of sheaves over Σ.[5]

Proof: Even though the proof of this result is a standard matter, for the reader's convenience we include it here. Let $0 \to \mathscr{A} \to \mathscr{A}^*$ be an injective resolution of $\mathscr{A}$. The sheaf sequence

$$0 \to \mathscr{A} \hat{\otimes} \mathscr{K} \to \mathscr{A}^* \hat{\otimes} \mathscr{K} \tag{4.1}$$

is exact because on the stalks it reads

$$0 \to \mathscr{A}_x \otimes K \to \mathscr{A}_x^* \otimes K \tag{4.2}$$

and this is exact since K is flat. In order to prove that (4.1) is injective it suffices to show that (4.2) is injective for all x. Now, every $\mathscr{A}_x^*$ is an injective Σ-module; moreover, since K is free,

$$A_x^* \oplus K \simeq \oplus A_x^*,$$

so that $A_x^* \otimes K$ is injective. Then the sequence (4.1) can be used to compute the cohomology of $\mathscr{A} \hat{\otimes} \mathscr{K}$:

$$H^*(X \times Y, \mathscr{A} \hat{\otimes} \mathscr{K}) \simeq H^*[\Gamma(\mathscr{A}^* \hat{\otimes} \mathscr{K})].$$

Since the right-hand side equals $H^*[\Gamma\, \mathscr{A}^* \otimes K]$, and K is flat, we obtain

$$H^*(X \times Y, \mathscr{A} \hat{\otimes} \mathscr{K}) \simeq H^*[\Gamma\, \mathscr{A}^*] \otimes K$$
$$\simeq H^*(X, \mathscr{A}) \otimes K. \quad \square$$

Lemma 4.2: Let $\hat{\mathscr{G}}$ denote the sheaf of GH^∞ functions on $B_L^{m,0}$. For all $p > 0$, $H^p(B_L^{m,0}, \hat{\mathscr{G}}) = 0$.

Proof: $\hat{\mathscr{G}}$ is isomorphic to the sheaf $\mathscr{C} \hat{\otimes} \mathscr{B}_{L'}$, where $\mathscr{C}$ is the sheaf of $\mathscr{C}^\infty$ functions on $\mathbb{R}^m$ and $B_{L'}$ is the constant sheaf on $P_L^{m,0}$ with stalk $B_{L'}$. Then the previous lemma yields $H^*(B_L^{m,0}, \hat{\mathscr{G}}) \simeq H^*(\mathbb{R}^m, \mathscr{C}) \otimes B_{L'}$. Since $\mathscr{C}$ is acyclic (as it is fine), this proves the claim. □

Lemma 4.3: For all $p > 0$, $H^p(B_L^{m,n}, \mathscr{G}) = 0$.

Proof: Let us denote by $\mathscr{P}$ the sheaf of polynomial functions on $B_L^{0,n}$ with coefficients in $B_{L'}$. One has $\mathscr{G} = \hat{\mathscr{G}} \hat{\otimes} \mathscr{P}$; moreover, $\mathscr{P}$ is constant, and $\mathscr{P}_y \simeq \Lambda$ for all $y \in B_L^{0,n}$, Λ being the exterior algebra generated over $B_{L'}$ by n generators. Here Λ is isomorphic to $(B_L)^{2^n}$ as a $B_{L'}$-module. Lemma 4.1 yields $H^p(B_L^{m,n}, \mathscr{G}) \simeq H^p(B_L^{m,0}, \hat{\mathscr{G}}) \otimes_{B_{L'}} (B_L)^{2^n}$, whence the thesis follows.

Now we can prove Theorem 4.1. Let $\mathscr{U}$ be a good cover of the body M_0 of M (a good cover is an open cover such that any finite, nonvoid intersection of its members is diffeomorphic to an open ball in $\mathbb{R}^m$). Consider on M the cover $\mathscr{W} = \Phi^{-1}\, \mathscr{U}$, where Φ: $M \to M_0$ is the bundle projection. Denoting by $\mathscr{G}$ the structure sheaf of M, we have that the Čech cohomology relative to $\mathscr{W}$ with values in $\mathscr{G}$ is trivial,[3] i.e.,

$$H^p(\mathscr{W}, \mathscr{G}) = 0, \quad \text{for } p > 0.$$

According to Theorem 2.1, Theorem 4.1 is proved if we show that, for all intersections $W_{\alpha_1 \cdots \alpha_q}$ of members of $\mathscr{W}$, the restricted sheaf $\mathscr{G} | W_{\alpha_1 \cdots \alpha_q}$ is acyclic. Since $W_{\alpha_1 \cdots \alpha_q}$ is GH^∞ diffeomorphic to $B_L^{m,n}$, this condition follows from Lemma 4.3, so that Theorem 4.1 holds.

The techniques used to prove Theorem 4.1 provide also a tool for computing the cohomology of the structure sheaf of a generic (non-De Witt) supermanifold in terms of the Čech cohomology of suitable covers. Let $\mathscr{W}$ be a cover of a supermanifold such that all nonvoid, finite intersections of its members are GH^∞ diffeomorphic to open rectangles in $B_L^{m,n}$. Obviously, any supermanifold admits covers of such kind. Then we have the following theorem.

Theorem 4.2: Let M be a supermanifold with structure sheaf $\mathscr{G}$, and let $\mathscr{W}$ be a cover of M as above. Then $\check{H}^*(\mathscr{W}, \mathscr{G}) \simeq H^*(M, \mathscr{G})$.

Proof: The same arguments as in Theorem 4.1 allow us to prove that the restriction to a rectangle in $B_L^{m,n}$ of the structure sheaf of $B_L^{m,n}$ is acyclic. Then Theorem 2.1 implies the thesis.

Example: The open cylinder $M = \mathbb{R} \times S^1$ can be given a structure of (1,0)-dimensional supermanifold with $L = L' = 2$. The sheaf $\mathscr{G}$ is not acyclic, as the analysis of the GH^∞

de Rham cohomology of M shows.[3] Indeed, let $\pi: M \to S^1$ be the projection, and let $\mathscr{U}$ be a good cover of S^1. Then $\mathscr{W} = \pi^{-1}\mathscr{U}$ is a cover of M satisfying the above mentioned requirement, so that $H^*(M,\mathscr{G}) \simeq \check{H}^*(\mathscr{W},\mathscr{G})$. The explicit computation of the right-hand side gives

$$H^p(M,\mathscr{G}) = \begin{cases} C^\infty(\mathbb{R}) \otimes B_L, & \text{for } p = 1, \\ 0, & \text{for } p > 1, \end{cases}$$

where $C^\infty(\mathbb{R})$ is the vector space of real-valued functions on $\mathbb{R}$. Direct computation gives also $H^0(M,\mathscr{G}) = \Gamma\,\mathscr{G} = \mathbb{R} \oplus [C^\infty(\mathbb{R}) \otimes N_L]$ (Ref. 3), where N_L is the nilpotent ideal of B_L.

V. COHOMOLOGY OF COMPLEX SUPERMANIFOLDS

Let M be an (m,n)-dimensional complex supermanifold, and let $\mathscr{O}$ and $\mathscr{G}$ denote, respectively, the sheaf of complex superanalytic and GH^∞ C_L-valued functions on M. The sheaf $\mathscr{D}$ of graded C_L-linear derivations of $\mathscr{G}$ has a natural splitting $\mathscr{D} = \mathscr{D}' \oplus \mathscr{D}''$, where $\mathscr{D}'$ is locally generated over $\mathscr{G}$ by $\{\partial/\partial z^i, \partial/\partial\xi^\alpha, i = 1,...,m, \alpha = 1,...,n\}$, and $\mathscr{D}''$ by $\{\partial/\partial\bar{z}^i, \partial/\partial\bar{\xi}^\alpha\}$, (z^i, ξ^α) being local complex superanalytic coordinates on M. Analogously, the dual sheaf $\mathscr{D}^*$ (whose sections are C_L-valued GH^∞ differential one-forms) has a splitting $\mathscr{D}^* = (\mathscr{D}^*)' \oplus (\mathscr{D}^*)''$. The sheaf $\mathscr{A}^{p,q}$ of C_L-valued GH^∞ differential forms of type (p,q) is defined as

$$\mathscr{A}^{p,q} = \Lambda^p(\mathscr{D}^*)' \otimes_{\mathscr{G}} \Lambda^q(\mathscr{D}^*)''. \tag{5.1}$$

For each p there is an exact sheaf sequence

$$0 \to \mathscr{O}^p \xrightarrow{\text{incl}} \mathscr{A}^{p,0} \xrightarrow{\bar\partial} \mathscr{A}^{p,1} \xrightarrow{\bar\partial} \mathscr{A}^{p,2} \to \cdots,$$

where $\mathscr{O}^p$ is the sheaf of complex superanalytic p-forms on M. The exterior differentials ∂ and $\bar\partial$ are defined in local coordinates by the conditions

$$\partial f(z^1\cdots z^m, \xi^1\cdots\xi^n, \bar{z}^1\cdots\bar{z}^m, \bar\xi^1\cdots\bar\xi^n) = dz^i\frac{\partial f}{\partial z^i} + d\xi^\alpha\frac{\partial f}{\partial\xi^\alpha},$$

$$\bar\partial f(z^1\cdots z^m, \xi^1\cdots\xi^n, \bar{z}^1\cdots\bar{z}^m, \bar\xi^1\cdots\bar\xi^n) = d\bar{z}^i\frac{\partial f}{\partial\bar{z}^i} + d\bar\xi^\alpha\frac{\partial f}{\partial\bar\xi^\alpha},$$

where f is a GH^∞ function, and

$$\partial\partial = \bar\partial\bar\partial = \partial\bar\partial + \bar\partial\partial = 0.$$

For each p, the cohomology groups of the differential complex

$$\Gamma\,\mathscr{A}^{p,0} \xrightarrow{\bar\partial} \Gamma\,\mathscr{A}^{p,1} \xrightarrow{\bar\partial} \Gamma\,\mathscr{A}^{p,2} \to \cdots$$

will be denoted by $H_{\bar\partial}^{p,q}(M)$. In the theory of ordinary complex manifolds, one has the *Dolbeault theorem*,[7] whose generalization to the present situation would read

$$H^q(M,\mathscr{O}^p) \simeq H_{\bar\partial}^{p,q}(M). \tag{5.2}$$

We may wonder to what extent the isomorphism (5.2) applies to the case of complex supermanifolds.

Theorem 5.1: The isomorphism (5.2) holds if M is a complex De Witt supermanifold.

Proof: Obviously, complex De Witt supermanifolds are characterized as in Definition 3.1, but requiring the transition functions to be complex superanalytic. The isomorphism (5.2) can be proved by means of double complex techniques,[5,8] much in the same way as the "super" extension of the De Rham theorem is proved to hold in the case of real De Witt supermanifolds.[3] Alternatively, one can consider the exact sheaf sequences

$$0 \to \mathscr{O}^p \xrightarrow{\text{incl}} \mathscr{A}^{p,0} \xrightarrow{\bar\partial} \mathscr{Z}^{p,1} \to 0, \tag{5.3a}$$

$$0 \to \mathscr{Z}^{p,q} \xrightarrow{\bar\partial} \mathscr{A}^{p,q+1} \xrightarrow{\bar\partial} \mathscr{Z}^{p,q+1} \to 0, \tag{5.3b}$$

where $\mathscr{Z}^{p,q} = \ker(\bar\partial: \mathscr{A}^{p,q} \to \mathscr{A}^{p,q+1})$. Since the sheaves in the middle are acyclic by a similar argument as in Theorem 4.1, the long exact sequences induced by (5.3) in cohomology break into a series of short exact sequences which provide a proof of (5.2). □

Remark: As a corollary to Theorem 5.1, and using an obvious extension of the $\bar\partial$-Poincaré lemma (referred to as "Dolbeault Lemma" in Ref. 7), we obtain $H^q(C_L^{m,n}, \mathscr{O}^p) = 0$, $\forall\, q > 0$, $\forall\, p \geqslant 0$.

VI. CHERN CLASSES OF COMPLEX LINE SUPERBUNDLES

As a first application of the cohomological techniques so far developed, we sketch the main lines of a theory of complex super line bundles. A more comprehensive treatment will be given elsewhere.[9] Obviously, this topic is fundamental to the classification of vector bundles on supermanifolds. For instance, it can be applied to the study of the so-called super Riemann surfaces,[10,11] which presently are under consideration in superstring theory. One could also envisage similar constructions in the case of principal super fiber bundles,[12] with possible applications to anomalous supersymmetric and superstring theories.

Since the distinguishing feature of line bundles is that their structure group is Abelian, we are naturally led to the following definition, where by the GH^∞ bundle on M we mean a pair (E,π), E being a GH^∞ supermanifold, and π a GH^∞ surjective map $\pi: E \to M$.

Definition 6.1: A complex line superbundle on a GH^∞ supermanifold M is a locally trivial GH^∞ fiber bundle $\pi: E \to M$ with standard fiber $(C_L)_0$ and structure group $(C_L)_0^*$. □

Here $(C_L)_0^* = \mathrm{Gl}(1,0;\mathbb{C})$ denotes the multiplicative group of elements in $(C_L)_0$ with a nonvanishing body.

Given a trivializing cover $\mathscr{U} = \{U_\alpha\}$, the specification of the bundle E is equivalent to assigning GH^∞ transition functions $g_{\alpha\beta}: U_\alpha \cap U_\beta \to (C_L)_0^*$ subject to the usual conditions $g_{\alpha\alpha} = 1$, $g_{\alpha\beta}\,g_{\beta\gamma}\,g_{\gamma\alpha} = 1$. So a complex line superbundle E on M is in a one-to-one correspondence with an element (again denoted by E) of the Čech cohomology group $H^1(M,\mathscr{I}^*)$, $\mathscr{I}^*$ being the sheaf of $GH^\infty(C_L)_0^*$-valued functions on M. Let $\mathscr{I}$ denote the sheaf of $GH^\infty(C_L)_0$-valued functions on M, and let $\mathbb{Z}$ be the locally constant sheaf on M with stalk the integers. We have an exact sheaf sequence

$$0 \to \mathbb{Z} \xrightarrow{\text{incl}} \mathscr{I} \xrightarrow{\exp} \mathscr{I}^* \to 0, \tag{6.1}$$

where $\exp f = e^{2\pi i f}$ is defined as a power series. The sequence (6.1) induces in cohomology a long exact sequence

$$\cdots \to \check{H}^1(M,\mathbb{Z}) \to \check{H}^1(M,\mathscr{S}) \to \check{H}^1(M,\mathscr{S}^*) \xrightarrow{-c_1} \check{H}^2(M,\mathbb{Z}) \to \check{H}^2(M,\mathscr{S}) \to \cdots,$$

where $-c_1$ is the canonical connecting morphism. The element $c_1(E) \in \check{H}^2(M,\mathbb{Z})$ is the *(first) Chern class* of E. In general c_1 is not injective, so that nonisomorphic superbundles may have the same Chern class. So already in the GH^∞ case it happens what in the ordinary theory is typical of the holomorphic case. Indeed, one can introduce a Picard group

$$\mathrm{Pic}^0(M) \equiv \check{H}^1(M,\mathscr{S})/\mathrm{Im}\,\check{H}^1(M,\mathbb{Z}), \tag{6.2}$$

which classifies the nontrivial GH^∞ complex line superbundles having vanishing Chen class. However, if the base manifold M is De Witt, the acyclicity of $\mathscr{S}$ (following from Theorem 4.1) implies that c_1 is one to one, so that $\mathrm{Pic}^0(M) = 0$. In this case the complex line superbundles over M are faithfully classified by $\check{H}^2(M,\mathbb{Z})$. Using Theorem 2.2, one shows that $\check{H}^2(M,\mathbb{Z}) \simeq \check{H}^2(M_0,\mathbb{Z})$, where M_0 is the body of M. So we have proved the following theorem.

Theorem 6.1: The complex GH^∞ line superbundles over a De Witt supermanifold M are in a one-to-one correspondence with the ordinary complex line bundles on M_0. □

Particularly interesting is the case of a holomorphic complex line superbundle E over a complex De Witt supermanifold M with body M_0, the transition functions of E being complex superanalytic.[13] In this case both M and M_0 have a Picard group, and, in particular,

$$\mathrm{Pic}^0(M_0) = \check{H}^1(M_0,\mathscr{O})/\mathrm{Im}\,\check{H}^1(M_0,\mathbb{Z}),$$

where $\mathscr{O}$ is the sheaf of holomorphic functions on M_0. Using the same techniques as in Lemma 4.3, one gets $\check{H}^1(M,\mathscr{S}) \simeq \check{H}^1(M_0) \otimes_{\mathbb{C}} (C_L)_0$. Since $\check{H}^1(M,\mathbb{Z}) \simeq \check{H}^1(M_0,\mathbb{Z})$, we have

$$\mathrm{Pic}^0(M) \simeq [\check{H}^1(M_0) \otimes_{\mathbb{C}} (C_L)_0]/\mathrm{Im}\,\check{H}^1(M_0,\mathbb{Z}). \tag{6.3}$$

The body map $\sigma: C_L \to \mathbb{C}$ extends to a morphism ω: $\mathrm{Pic}^0(M) \to \mathrm{Pic}^0(M_0)$. Denoting by $\overline{P}_L$ the nilpotent ideal in $(C_L)_0$, we have the following theorem.

Theorem 6.2: If E is a holomorphic complex super line bundle over a complex De Witt supermanifold M, and ω: $\mathrm{Pic}^0(M) \to \mathrm{Pic}^0(M_0)$ is the morphism defined as above, then

$$\ker \omega \simeq \overline{P}_L. \qquad \square$$

Example: Letting $L=1, L'=0$, the space $M = T^2 \times \mathbb{R}^2$ can be given a structure of (1,1) complex De Witt supermanifold, whose body M_0 is a complex one-dimensional torus. Direct computation shows that $\mathrm{Pic}^0(M) = (C_L)_0^*$; on the other hand, $(C_L)_0^*/\overline{P}_L \simeq \mathbb{C}^*$, which is just the Picard group of M_0.

We can also produce a generalization of the classical result that the Chern class of a complex line bundle on a differentiable manifold is mapped by the de Rham isomorphism into the de Rham cohomology class determined by a curvature form on the bundle.[7] The generalization to supermanifolds needs some care.

Definition 6.2: A connection Δ on a complex line superbundle E is a morphism of sheaves of graded C_L-modules

$$\Delta: \quad E \to (\mathrm{Hom}(TM,E))$$

[where $\mathrm{Hom}(TM,E)$ is the complex superbundle on M whose standard fiber at $x \in M$ is the graded C_L-module of graded B_L-linear morphisms $T_xM \to E_x$] satisfying the property

$$\Delta(f\xi) = df \otimes \xi + f\Delta\xi, \quad \text{for all } f \in \mathscr{S}(U), \quad \xi \in E(U).$$

□

The square Δ^2 of the connection is called the *curvature* of Δ. The curvature determines a global $(C_L)_0$-valued two-form on M, denoted by Ω. The Bianchi identity implies that Ω is closed under d.

We recall from Ref. 3 that a GH^∞ (B_L-valued) de Rham cohomology $H^*_{\mathrm{SDR}}(M)$ can be introduced as the cohomology of the complex

$$\Gamma\,\mathscr{G}^0 \xrightarrow{d} \Gamma\,\mathscr{G}^1 \xrightarrow{d} \Gamma\,\mathscr{G}^2 \to \cdots,$$

where $\mathscr{G}^p$ is the sheaf of GH^∞ B_L-valued p-forms on M. Any curvature form Ω determines cohomology classes $[\mathrm{Re}\,\Omega]$ and $[\mathrm{Im}\,\Omega]$ in $H^2_{\mathrm{SDR}}(M)$, which are shown to be independent of the connection. The inclusion $\mathscr{G}^* \to \mathscr{C}^* \otimes B_L$, where $\mathscr{C}^p$ is the sheaf of C^∞ p-forms, induces in cohomology a map α: $H^*_{\mathrm{SDR}}(M) \to H^*_{\mathrm{DR}}(M) \otimes B_L$, where $H^*_{\mathrm{DR}}(M)$ is the ordinary de Rham cohomology of M.

Theorem 6.3: Let E be a complex vector superbundle that admits a connection whose curvature form is purely imaginary. Then, for all connections Δ on E with curvature form Ω, $[(i/2\pi)\Omega]$ is $(B_L)_0$ valued and

$$-\alpha[(i/2\pi)\Omega] = \beta \circ c_1(E),$$

where β: $\check{H}^2(M, \mathbb{Z}_L) \to H^2_{\mathrm{DR}}(M) \otimes B_L$ is the composition of the map $\check{H}^2(M, \mathbb{Z}_L) \to \check{H}^2(M, B_L)$ with the isomorphism $\check{H}^2(M, B_L) \to H^2_{\mathrm{DR}}(M) \otimes B_L$ induced by the ordinary Čech–de Rham isomorphism.

Proof: The classical proof can be straightforwardly adapted to the present situation.[7] □

The assumption of the existence of a connection with purely imaginary curvature, which plays the role of the existence of a Hermitian structure in the ordinary theory, is necessary, due to the lack of GH^∞ partitions of unity over supermanifolds. However, if M is De Witt, it has "tubular" GH^∞ partitions of unity,[14] so that this assumption can be removed.

We can also consider bundles with fiber $(C_L)_1$. Since the group $\mathrm{Gl}(0,1;\mathbb{C})$ of $(C_L)_0$-linear automorphisms of $(C_L)_1$ is again $(C_L)_0^*$, we obtain the same results as above. Thus the way is open to the study of complex vector superbundles of higher rank, i.e., bundles with fiber C_L^{m+n} and structure group $\mathrm{Gl}(m,n;\mathbb{C})$.[15,16]

ACKNOWLEDGMENTS

This research was partly supported by the National Group for Mathematical Physics of the Italian Research Council (CNR), by the Italian Ministry for Public Education through the research project "Geometria e Fisica," and by the joint Italian–Spanish CNR–CSIC project, "Methods and Applications of Differential Geometry in Mathematical Physics." We would like to thank D. Hernández Ruipérez and J. Muñoz Masqué for useful discussions.

[1]B. De Witt, *Supermanifolds* (Cambridge U.P., London, 1984).
[2]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[3]C. Bartocci and U. Bruzzo, J. Math. Phys. **28**, 2363 (1988).
[4]J. M. Rabin, Commun. Math. Phys. **108**, 375 (1987).
[5]R. Godement, *Théorie des faisceaux* (Hermann, Paris, 1964).
[6]A. Rogers, Commun. Math. Phys. **105**, 375 (1986).
[7]S. S. Chern, *Complex Manifolds Without Potential Theory* (Springer, New York, 1979).
[8]R. Bott and L. W. Tu, *Differential Forms in Algebraic Topology* (Springer, New York, 1982).
[9]C. Bartocci and U. Bruzzo, "Super line bundles," Dept. of Mathematics, Univ. of Genova preprint 49/1988.
[10]D. Friedan, "Notes on string theory and two-dimensional conformal field theory," in *Proceedings of the Workshop on Unified String Theories*, edited by D. Gross and M. Green (World Scientific, Singapore, 1986).
[11]L. Crane and J. M. Rabin, Commun. Math. Phys. **113**, 601 (1988).
[12]U. Bruzzo and R. Cianci, Class. Quantum Gravit. **1**, 213 (1984); C. Bartocci, U. Bruzzo, and G. Landi, "Chern-Simons forms on principal super fibre bundles," Preprint SISSA 109/87/FM (Trieste).
[13]We would like to acknowledge an interesting discussion on this argument with Professor J. Rabin.
[14]C. Bartocci and U. Bruzzo, "Existence of connections on super bundles," Dept. of Mathematics, Univ. of Genova preprint, 1988.
[15]C. Bartocci and U. Bruzzo, J. Geom. Phys. **4** (3), (1987).
[16]U. Bruzzo and D. Hernández Ruipérez, "Characteristic classes of super vector bundles," Preprints de Matemáticas, Universidad de Salamanca, and Dept. of Mathematics, Univ. of Genova preprint.

Global anomalies and algebraic topology

Albert T. Lundell
Department of Mathematics, University of Colorado, Campus Box 426, Boulder, Colorado 80309
Yasunari Tosa
Department of Physics, University of Colorado, Campus Box 390, Boulder, Colorado 80309

(Received 7 January 1988; accepted for publication 23 March 1988)

The algebraic topology aspect of the global pure gauge anomaly calculation is investigated. In particular, the use of a cohomology sequence clarifies the method initiated by Witten [Nucl. Phys. **223**, 422, 433 (1983)] and Elitzur and Nair [Nucl. Phys. B **243**, 205 (1984)]. Examples in SU(N), Sp(N), and SO(N) are discussed.

I. INTRODUCTION

Currently, the construction of a physical theory requires one to show that the theory is not only free of perturbative (local) anomalies, but also free of nonperturbative (global) anomalies, otherwise the theory is inconsistent, i.e., the fermionic functional measure is not invariant under the gauge transformation. As a typical example, Witten[1] has shown that an SU(2) gauge theory with an odd number of Weyl fermion doublets has a global anomaly in $D=4$ dimensions. This fact is relevant in our world, since the standard model is the SU(3)×SU(2)×U(1) gauge theory. Fortunately, so far we have an even number of doublets, paired as (three-colored) quark and lepton doublets.

Because of recent interests in higher-dimensional theories, it is essential to have a general method of calculating anomalies. Witten[2] and others[3] have developed such a method that encompasses not only pure gauge anomalies, but also mixed and gravitational anomalies. It is usually very difficult to estimate the exact values for the anomalies, although the method is very powerful in the sense that Witten[2] was able to show the absence of global anomalies for superstring-induced field theories. The main difficulty is that one must investigate the index theorem for a manifold with boundary.

Witten, Elitzur, and Nair[4] have devised another method in which it is easier to estimate anomalies, but which is limited to the analysis of pure gauge anomalies.[5] Their method has been applied to SU(N) gauge theores[5,6] and has recently been generalized and applied to the analysis of other simple Lie groups.[7,8] The purpose of this paper is to explore the mathematical aspect of the method and, in particular, concentrate on the exact sequence of the anomaly. Because of the unfamiliarity of algebraic topology among physicists, we give a rather detailed account.

This paper is organized as follows. In this section, we establish the general results obtained in Refs. 7 and 8. In Sec. II, we reexpress the anomaly for a typical case for $H=\mathrm{SU}(N)$, Sp(N), and SO(N). The use of a cohomology sequence clarifies the reasoning, although the result can be obtained in a brute force way. In Sec. III, we investigate concrete examples in special unitary groups, symplectic groups, and special orthogonal groups.

A. General results on the anomaly

We explain the general results obtained in Refs. 7 and 8 from a mathematical point of view.

We will say that infinitesimal gauge transformations reach a gauge transformation $g: S^k \to G$, provided that these infinitesimal transformations generate a homotopy of g with the constant map at $1 \in G$.

Nonperturbative anomalies are associated with gauge transformations that cannot be continuously deformed to a constant (the identity), i.e., they occur only when $\Pi_D(H) \neq 0$ for the gauge group H in dimension $D=2n$. If we embed H in group G such that $\Pi_{2n}(G)=0$, then any global anomaly caused by $h: S^{2n} \to H$, which corresponds to a nontrivial element of $\Pi_{2n}(H)$, can be calculated by integrating the perturbative anomaly caused by infinitesimal gauge transformations of G that reaches h. We know how the infinitesimal gauge transformation of G affects the fermionic functional measure.[9] In order to use this method, we must satisfy the embedding condition: the fermion representation $\tilde{\omega}$ of G must yield the representation ω and singlets of H under reduction. Note that this condition may not be satisfied for some choice of G. We will not go into this subject, which has been discussed in Ref. 8.

Let $\Pi_{2n}(G)=0$ and suppose $g: S^{2n} \to G$ can be reached by infinitesimal gauge transformations of G. For a Weyl fermion with a representation $\tilde{\omega}$ of G, the anomaly coefficient $A(\tilde{\omega})$ in dimension $D=2n$ is given by

$$A(\tilde{\omega}) = \exp\left[i \int_{D^{2n+1}} \gamma(\tilde{g},A,F)\right], \tag{1.1}$$

where D^{2n+1} denotes the disk with boundary S^{2n} and $\tilde{g}$ is the extension of g determined by the infinitesimally generated homotopy where $\tilde{g}|_{\text{origin}}$ = identity element and $\tilde{g}|_{\partial D^{2n+1}} = g$. The $(2n+1)$-form, γ, is defined as follows:

$$\operatorname{Tr} F^{n+1} = d\omega_{2n+1}(A,F) = d\omega_{2n+1}(A^{\tilde{g}},F^{\tilde{g}}), \tag{1.2}$$

$$\begin{aligned}\gamma(\tilde{g},A,F) &= \frac{i^{n+1}}{(2\pi)^n (n+1)!}\left[\omega_{2n+1}(A^{\tilde{g}},F^{\tilde{g}}) - \omega_{2n+1}(A,F)\right] \\ &\propto \operatorname{Tr}[(\tilde{g}^{-1}d\tilde{g})^{2n+1}] + d\alpha_{2n}(\tilde{g},A,F).\end{aligned} \tag{1.3}$$

All the group dependent quantities $\tilde{g}$, A, and F are in the representation $\tilde{\omega}$. Note that the action of a gauge transformation g of G is defined by $A \to A^g = g^{-1}Ag + g^{-1}\,dg$ for the one-form A. The two-form F is defined as $F = dA + A^2$. Both A and F are Lie algebra valued forms.

If $\Pi_{2n}(G)=0$, $H \subset G$, and $h: S^{2n} \to H$, there exists a g: $(D^{2n+1}, S^{2n}) \to (G,H)$ with $\partial D^{2n+1} = S^{2n}$ such that

 0022-2488/88/081795-12$02.50

$g|_{S^{2n}} = h$. (In general, for $h: S^m \to H$ with $i_*([h]) = 0$, where $i_*: \Pi_m(H) \to \Pi_m(G)$ is induced by inclusion, there is a map $g: (D^{m+1}, S^m) \to (G,H)$ such that $g|_{S^m} = h$.[10]) Using this g, the global anomaly of H is given by Eq. (1.1). Note that g is classified by the relative homotopy group[10] $\Pi_{2n+1}(G/H)$, which is an Abelian group for $n \geqslant 1$ and is isomorphic to $\Pi_{2n+1}(G/H)$.[11] Hereafter, we omit the forms A and F from γ. When $\gamma(g)$ is integrated over the compact space S^{2n+1}, the dependence on A and F disappears, because of Eqs. (1.2) and (1.3).

In order for the anomaly expression to be independent of the extension, two different extensions $g': D^{2n+1} \to G$ and $g'': D^{2n+1} \to G$ must satisfy

$$\exp i\left[\int_{D^{2n+1}} \gamma(g') - \int_{D^{2n+1}} \gamma(g'')\right] = \exp\left[i \int_{S^{2n+1}} \gamma(g)\right] = 1, \tag{1.4}$$

where $g: S^{2n+1} \to G$ restricts to g' and g'' on opposite hemispheres. This fixes the normalization of the integral over S^{2n+1} to be $2\pi \times$ (integer).

For any Lie group G and $g: S^{2n+1} \to G$, the integral $\int_{S^{2n+1}} \gamma(g)$ satisfies two important properties: (1) it is homotopy invariant; (2) it is a homomorphism $\Pi_{2n+1}(G) \to \mathbb{R}$.

For two smoothly homotopic maps $f, g: S^{2n+1} \to G$ and a closed k-form γ of G, $f^*\gamma - g^*\gamma = d\phi$ for some $(k-1)$-form ϕ, where f^*, g^* are the induced maps, pulling forms on G back to forms on S^{2n+1}.[12] Since $H^k(S^{2n+1})$ is nontrivial only for $k = 0$ or $k = 2n+1$, the integral is trivial except for these dimensions. Thus we use $(2n+1)$-forms. Property (2) means

$$\int_{S^{2n+1}} \gamma(fg) = \int_{S^{2n+1}} \gamma(f) + \int_{S^{2n+1}} \gamma(g),$$

which can be proved using the fact that $g^{-1}\, dg$ is a one-form. Together these show that modulo some normalization this integral gives a homomorphism $\Pi_{2n+1}(G) \to \mathbb{Z}$, which we denote by $\int \gamma^G$, i.e., $\int \gamma^G(\alpha) \equiv \int_{S^{2n+1}} \gamma(g)$ for $\alpha = [g] \in \Pi_{2n+1}(G)$ and $g: S^{2n+1} \to G$. The immediate consequence is that *$\int \gamma^G(\alpha)$ vanishes for an element of finite order in $\Pi_{2n+1}(G)$*.

Note that the anomaly coefficient is the integral of γ over D^{2n+1}, not over S^{2n+1}.

Now assume that *H is free of perturbative anomalies, i.e., $\mathrm{Tr}\, F^{n+1}|_H = 0$, or $\gamma(h) = 0$ for $h: S^{2n+1} \to H$*. Then, since $\gamma(gh, A, F) = \gamma(g, A, F) + \gamma(h, A^g, F^g)$, we see $\gamma(g)$ is invariant under the gauge transformations of H and depends only on the coset space G/H. If $p: G \to G/H$ is the projection and g: $(D^{2n+1}, S^{2n}) \to (G,H)$, then $\tilde{p}g: (D^{2n+1}, S^{2n}) \to (G/H, *)$ determines $f: S^{2n+1} \to G/H$, since the boundary is collapsed to a point [i.e., there is a smooth map k: $(D^{2n+1}, S^{2n}) \to (S^{2n+1}, *)$] such that

$$\begin{array}{ccc} (D^{2n+1}, S^{2n}) & \xrightarrow{g} & (G,H) \\ \Big\downarrow k & & \Big\downarrow \tilde{p} \\ (S^{2n+1}, *) & \xrightarrow{f} & (G/H, *) \end{array}$$

commutes. Conversely, since $\tilde{p}$ induces an isomorphism $\tilde{p}_*: \Pi_k(G,H) \to \Pi_k(G/H, *)$ a map $f: S^{2n+1} \to G/H$ determines a map $g: (D^{2n+1}, S^{2n}) \to (G,H)$ with $\tilde{p}g$ homotopic to fk.[11] We may set

$$\int_{D^{2n+1}} \gamma(g) = \int_{D^{2n+1}} \gamma(\tilde{p}g) = \int_{D^{2n+1}} \gamma(fk) = \int_{S^{2n+1}} \gamma([f]), \tag{1.5}$$

where $[f] \in \Pi_{2n+1}(G/H)$. We define a map $\int \gamma^{G/H}$: $\Pi_{2n+1}(G/H) \to \mathbb{R}$ by

$$\int \gamma^{G/H}(\beta) = \int_{D^{2n+1}} \gamma(\tilde{p}_*^{-1}\beta) \tag{1.6}$$

and the anomaly for H can be written as

$$A(\tilde{\omega}) = \exp\left[i \int \gamma^{G/H}(\beta)\right] \text{ for some } \beta \in \Pi_{2n+1}(G/H). \tag{1.7}$$

We see immediately the following proposition.[8]

Proposition 1.1: If $\Pi_{2n+1}(G/H)$ is a finite group, then no anomalies exist for H, provided the embedding conditions are satisfied. □

It is now clear that the anomaly coefficient is related to the following exact homotopy sequence[11,13]:

$$\begin{array}{ccccccccc} \Pi_{2n+1}(H) & \xrightarrow{i_*} & \Pi_{2n+1}(G) & \xrightarrow{j_*} & \Pi_{2n+1}(G,H) & \xrightarrow{\partial_*} & \Pi_{2n}(H) & \xrightarrow{i_*} & \Pi_{2n}(G) \\ & & & \searrow^{p_*} & \Big\downarrow \tilde{p}_* & \nearrow_{\Delta_*} & & & \\ & & & & \Pi_{2n+1}(G/H). & & & & \end{array} \tag{1.8}$$

Various facts are known about each homotopy group of this sequence.

(i) All of the homotopy groups are Abelian for $n \geqslant 0$.[13]

(ii) If K is a compact, connected, classical Lie group, then $\Pi_i(K)$ is a finite group with the following exceptions[14]:

$\Pi_{2n-1}(\mathrm{SU}(N))$, $N \geqslant n$;

$\Pi_{4n-1}(\mathrm{Sp}(N))$, $N \geqslant n$;

$\Pi_{4n-1}(\mathrm{SO}(N))$, $N \geqslant 2n+1$;

$\Pi_{4n+1}(\mathrm{SO}(4n+2))$.

(iii) If $\Pi_{2n+1}(H)$ is finite, then *either* both $\Pi_{2n+1}(G)$ and $\Pi_{2n+1}(G/H)$ are finite *or* both contain an infinite cyclic summand $\mathbb{Z}$. (See Lemma A.5 in Appendix A.)

Next we show the following proposition.

Proposition 1.2: If p: $G \to G/H$ is the projection, then

$$\int \gamma^{G/H} \circ p_* = \int \gamma^G. \tag{1.9}$$

Proof: Let k: $(D^{2n+1}, S^{2n}) \to (S^{2n+1}, *)$ be a smooth map collapsing S^{2n} to the point $*$, let j: $(G,1) \to (G,H)$ be the inclusion, and let g_0: $(S^{2n+1}, *) \to (G,1)$. That is,

$$\begin{array}{ccccc} (D^{2n+1}, S^{2n}) & \to & (G,H) & \xrightarrow{\tilde{p}} & (G/H, *) \\ \downarrow k & & \uparrow j & \nearrow p & \\ (S^{2n+1}, *) & \xrightarrow{g_0} & (G,1) & & \end{array}$$

commutes. Then

$$\int \gamma^{G/H}(p_*[g_0]) = \int_{D^{2n+1}} \gamma \tilde{p}_*^{-1} p_*([g_0])$$
$$= \int_{D^{2n+1}} \gamma j_*([g_0]) = \int_{D^{2n+1}} \gamma(j g_0 k)$$
$$= \int_{S^{2n+1}} \gamma(j g_0) = \int_{S^{2n+1}} \gamma(g_0)$$
$$= \int \gamma^G([g_0]). \qquad \square$$

We next express the anomaly coefficient for H in terms of quantities in G. For an arbitrary gauge transformation of H, there is an extension g corresponding to $\tilde{\beta} \in \Pi_{2n+1}(G/H)$. The anomaly coefficient is given by

$$A(\tilde{\omega}) = \exp\left[i \int \gamma^{G/H}(\tilde{\beta}) \right],$$

which can be written as

$$A(\tilde{\omega}) = \exp\left[\frac{i}{m} \int \gamma^{G/H}(m\tilde{\beta}) \right]$$

for an arbitrary m, because of the homomorphic property. Now fix m to be the least common multiple of the orders of elements of $\Pi_{2n}(H)$. Then $\Delta_*(m\tilde{\beta}) = 0$ and thus there exists an $\tilde{\alpha} \in \Pi_{2n+1}(G)$ such that $p_*(\tilde{\alpha}) = m\tilde{\beta}$. Using Proposition 1.2,

$$A(\tilde{\omega}) = \exp\left[\frac{i}{m} \int \gamma^{G/H} \circ p_*(\tilde{\alpha}) \right] = \exp\left[\frac{i}{m} \int \gamma^G(\tilde{\alpha}) \right].$$

Thus we obtain[8] the following proposition.

Proposition 1.3: If $\Pi_{2n+1}(G)$ is finite, then no anomalies exist, provided that the embedding conditions are satisfied. $\square$

Consequently, from now on we consider cases where *both* $\Pi_{2n+1}(G)$ *and* $\Pi_{2n+1}(G/H)$ contain inifinite cyclic subgroups. In the case where $\Pi_{2n+1}(G)$ has rank 1, any element $\tilde{\alpha}$ can be written as $\tilde{\alpha} = b\alpha + \eta$, where b is an integer, α is a generator of $\Pi_{2n+1}(G)$ of infinite order, and η is of finite order. Therefore we have the following proposition.[8]

Proposition 1.4: If $\Pi_{2n+1}(G)$ is of rank 1, then the anomaly coefficient for H is given by

$$A(\tilde{\omega}) = \exp\left[i \frac{b}{m} \int \gamma^G(\alpha) \right], \tag{1.10}$$

where α is a generator of $\Pi_{2n+1}(G)$ of infinite order and m is the least common multiple of the orders of $\Pi_{2n}(H)$. $\square$

If $\Pi_{2n+1}(G)$ is of rank greater than 1, then Eq. (1.10) must be modified.[15]

The calculation of the anomaly is reduced to the calculation of three quantities: b, m and $\int \gamma^G(\alpha)$. As we noted earlier, the last quantity is always $2\pi \times$ (integer). This integer turns out to be a Dynkin index $Q_{n+1}(\tilde{\omega})$,[4,7,8] which can be calculated by group representation theory, but in this paper this fact is not needed. Consequently, the H gauge theory has no anomaly if $(b/2\pi m) \int \gamma^G(\alpha)$ is an integer. In the next section, we show that b/m is expressed only in terms of the map $p_*: \Pi_{2n+1}(G) \to \Pi_{2n+1}(G/H)$, using a certain cohomology sequence. We do not have to know exactly what all the relevant homotopy groups are (it is usually very difficult to calculate homotopy groups).

B. Exact homotopy sequences for various groups

From now on G and H will denote compact connected simple Lie groups.

For $H = \mathrm{SU}(N)$, the nonvanishing homotopy groups $\Pi_{2n}(\mathrm{SU}(N))$ occur only in the unstable range $(N < n+1)$ and $\Pi_{2n+1}(\mathrm{SU}(\tilde{N})) = \mathbb{Z}$ and $\Pi_{2n}(\mathrm{SU}(\tilde{N})) = 0$ in the stable range $(\tilde{N} \geqslant n+1)$, because of the Bott periodicity theorem.[16] From Ref. 8, we know that the embedding condition is satisfied for $G = \mathrm{SU}(\tilde{N})$ with $\tilde{N} > N$. Thus we can choose $G = \mathrm{SU}(\tilde{N})$ with its homotopy in the stable range. Then by using Lemma A.4 of Appendix A the exact homotopy sequence Eq. (1.8) has the form

$$0 \to \mathbb{Z} \to \mathbb{Z} \oplus T' \to T \to 0, \tag{1.11}$$

where T' and T are finite.

For $H = \mathrm{Sp}(N)$, the nonvanishing homotopy groups $\Pi_{2n}(\mathrm{Sp}(N))$ occur in both the unstable $(4N < 2n-1)$ and the stable range $(4N \geqslant 2n-1)$. The embedding condition for the representation is satisfied for $G = \mathrm{Sp}(\tilde{N})$ with $\tilde{N} > N$. The Bott periodicity theorem[16] states that

$$\Pi_k(\mathrm{Sp}) = \begin{cases} \mathbb{Z}, & k \equiv 3,7 \quad (\mathrm{mod}\ 8), \\ \mathbb{Z}_2, & k \equiv 4,5 \quad (\mathrm{mod}\ 8), \\ 0, & k \equiv 0,1,2,6 \quad (\mathrm{mod}\ 8). \end{cases} \tag{1.12}$$

If $D = 4n+2$, we can use $G = \mathrm{Sp}(n+1)$, and we have $\Pi_{4n+2}(G) = 0$ and $\Pi_{4n+3}(G) = \mathbb{Z}$. Furthermore, $\Pi_{4n+2}(H)$ in the unstable range is always finite (see Appendix B) and thus we have the sequence (1.11) for (1.8). If $D = 8n$, we can use $G = \mathrm{Sp}(2n)$ in the stable range and show that no anomalies exist by Proposition 1.3, since $\Pi_{8n+1}(G) = 0$. In dimensions $D = 8n+4$, we cannot use $G = \mathrm{Sp}(\tilde{N})$. It turns out that we can choose $G = \mathrm{SU}(\tilde{N})$ and satisfy the embedding condition, as we will see in Sec. III.

For $H = \mathrm{SO}(N)$, the situation becomes complicated. First, the embedding condition for the representation cannot be always satisfied; for a tensor representation (self-contragredient representation) we can choose any $G = \mathrm{SO}(\tilde{N})$ with $\tilde{N} > N$. However, for a spinor representation or a self-dual tensor represenation, we must choose $G = \mathrm{SO}(2l)$ for $H = \mathrm{SO}(2l-1)$, but for $H = \mathrm{SO}(2l)$ we cannot in general find a representation $\tilde{\omega}$ and G.[8] Second, as we have mentioned earlier, the homotopy of SO(N) in the unstable range can contain $\mathbb{Z}$. In particular,[14,17]

$$\begin{cases} \Pi_{4n-1}(\mathrm{SO}(m)) = \mathbb{Z}\oplus F, & \text{for } m\geqslant 2n+1, \\ \Pi_{4n-1}(\mathrm{SO}(4n)) = \mathbb{Z}\oplus\mathbb{Z}, \\ \Pi_{4n+1}(\mathrm{SO}(4n+2)) = \mathbb{Z}\oplus F', \end{cases}$$

where F,F' are finite groups, and every other $\Pi_k(\mathrm{SO}(m))$ is finite. Because of these two facts, we cannot say in general what the exact homotopy sequence (1.8) is, as in the case of SU(N) and Sp(N). For example, in dimensions $D\equiv 2,6 \pmod 8$, we have for $D>6$

$$0\to\mathbb{Z}\to\mathbb{Z}\oplus\mathbb{Z}\to\mathbb{Z}\to\mathbb{Z}_2\to 0$$

for $H=\mathrm{SO}(D+1)$ and $G=\mathrm{SO}(D+2)$.[8] Note also that because of the Bott periodicity theorem[16]

$$\Pi_k(\mathrm{SO}) = \begin{cases} \mathbb{Z}, & k\equiv 3,7 \pmod 8, \\ \mathbb{Z}_2, & k\equiv 0,1 \pmod 8, \\ 0, & k\equiv 2,4,5,6 \pmod 8. \end{cases} \quad (1.13)$$

We cannot use $G=\mathrm{SO}(\widetilde{N})$ in the stable range in dimensions $D\equiv 0 \pmod 8$.

Note that the complication occurred because of the choice of $G=\mathrm{SO}(\widetilde{N})$. Again, by choosing $G=\mathrm{SU}(\widetilde{N})$, we can circumvent part of the problem. In Sec. III, we do this in dimensions $D\equiv 0 \pmod 8$, where we cannot use $G=\mathrm{SO}(\widetilde{N})$.

For exceptional groups, homotopy groups are not as regular as those of classical groups.

We concentrate on the exact sequence (1.11) in the next section.

II. A COHOMOLOGY SEQUENCE

A. Notation

Let p be a prime integer. If A is a finitely generated Abelian group, let $e_p(A)$ be the smallest power of p such that $e_p(A)\cdot A$ is p-torsion-free. Then, $e(A)=\Pi_p e_p(A)$ is the least positive integer such that $e(A)\cdot A$ is free. Note that $e(A)$ is the order of the element of maximal order in the torsion subgroup of A, the order of every element of the torsion subgroup divides $e(A)$, and $e(A)$ divides the order of the torsion subgroup of A. Let ${}_nA=\{\alpha\in A\,|\,n\alpha=0\}$. Then for $n=e_p(A)$, the group ${}_nA$ is the p-torsion subgroup of A, and for $n=e(A)$, the group ${}_nA$ is the torsion subgroup of A. We will use $|\alpha|$ for the order of $\alpha\in A$, and $|A|$ for the order of A, so that $0\leqslant|\alpha|\leqslant|A|\leqslant\infty$. We use (m,n) to denote the greatest common divisor of the integers m and n. If p is prime and $n=p^a\cdot q$, where $(p,q)=1$, define $v_p(n)=a$.

B. Determination of *b*

As we have mentioned in the previous section, the typical exact sequence for the case where $H=\mathrm{SU}(N)$ in any even dimension and $H=\mathrm{Sp}(N)$ in dimensions $D\equiv 2,6 \pmod 8$ is the following short exact sequence:

$$0\to\mathbb{Z}\xrightarrow{f}\mathbb{Z}\oplus T'\xrightarrow{g}T\to 0, \quad (2.1)$$

where T' and T are finite Abelian groups. Note that g restricted to the summand T' is an isomorphism of T' onto a subgroup of T (See Lemma A.6 in Appendix A). Thus $e(T')$ divides $e(T)$ and $|T'|$ divides $|T|$.

Choose generators $\alpha\in\mathbb{Z}$ and $\beta\in\mathbb{Z}\oplus T'$ such that

$$f(\alpha)=a\beta+\tau, \quad (2.2)$$

where a is a nonvanishing positive integer and $\tau\in T'$. Of course the choice of the generator β is not unique; we may write $f(\alpha)=a(\beta+\sigma)+(\tau-a\sigma)$, where $\sigma\in T'$. There are, however, only finitely many choices for β, since T' is a finite group.

By exactness, $\sigma=g(n\beta+\rho)$ for any $\sigma\in T$. Writing $n=qa+r$ with $0\leqslant r<a$, we see that

$$\begin{aligned}\sigma &= g((qa+r)\beta+\rho) \\ &= g(r\beta)+qag(\beta)+g(\rho) \\ &= g(r\beta-q\tau+\rho),\end{aligned}$$

where we have used $g(f(\alpha))=ag(\beta)+g(\tau)=0$. Thus we may assume $\sigma=g(r\beta+\rho)$ with $0\leqslant r<a$.

Lemma 2.1: If $\sigma\in T$ and $\sigma=g(r\beta+\rho)$ with $0\leqslant r<a$, then r is unique and ρ is unique in T'/rT'.

Proof: If $g(r\beta+\rho)=\sigma=g(r'\beta+\rho')$, then $g((r-r')\beta+\rho-\rho')=0$, so by exactness, $(r-r')\beta+\rho-\rho'=f(m\alpha)=ma\beta+m\tau$. But, then $r-r'=ma$ and $\rho-\rho'=m\tau$. Since $a>|r-r'|=|m|a$, we must have $m=0$, so $r=r'$ and $\rho=\rho'$. If $\beta'=\beta+\eta$ is another choice of generator, then $\sigma=g(r\beta+\rho)=g(r(\beta+\eta)+\rho-r\eta)=g(r\beta'+\rho')$. Thus r is independent of the choice of generator β, and ρ is independent of the choice of generator (mod rT'). □

This Lemma shows that $r=r(\sigma)$ is a function r: $T\to\mathbb{Z}$.

Now, suppose $\sigma\in{}_mT$ so that $0=m\sigma=g(mr(\sigma)\beta+m\rho)$. By exactness,

$$mr(\sigma)\beta=m\rho=f(b_m(\sigma)\cdot\alpha)=a\cdot b_m(\sigma)\beta+b_m(\sigma)\tau \quad (2.3)$$

for a well-determined positive integer $b_m(\sigma)$. Then,

$$m\cdot r(\sigma)=a\cdot b_m(\sigma), \quad m\rho=b_m(\sigma)\cdot\tau. \quad (2.4)$$

Thus b_m is a function b_m: ${}_mT\to\mathbb{Z}$.

Lemma 2.2: The integer $b_m(\sigma)$ is divisible by $m/(m,a)$, and $r(\sigma)$ is divisible by $a/(m,a)$, provided that $r(\sigma)\neq 0$.

Proof: Divide both sides of the equation $mr(\sigma)=ab_m(\sigma)$ by (m,a) and observe that $m/(m,a)$ and $a/(m,a)$ are relatively prime to each other. □

We observe the following: (i) this result is stronger than $b_m(\sigma)$ being divisible by m/a; (ii) since $r(g(\beta+\rho))=1$ and thus $b_m(g(\beta+\rho))=m/a$, it follows that m/a is the smallest value for nonvanishing $b_m(\sigma)$; (iii) we have $b_m(\sigma)=0$ if and only if $r(\sigma)=0$, i.e., if and only if $\sigma=g(\theta)$ for some $\theta\in T'$. That is, the torsion part T' of $\Pi_{2n+1}(G/H)$ does not contribute to the anomaly, confirming Proposition 1.1.

For different values of m, one obtains different information. For example to study σ itself, one might consider $m=|\sigma|$; to study the two-primary part take $m=e_2(T)$; to study the odd torsion, take $m=e(T)/e_2(T)$, etc. The most important choice for our purposes is $m=e(T)$.

Lemma 2.3: The element $g(\beta)$ has order $a|\tau|$.

Proof: Since $f(|\tau|\alpha)=a|\tau|\beta+|\tau||\tau=a|\tau|\beta$, we have $0=g(a|\tau|\beta)=a|\tau|g(\beta)$. If $0=mg(\beta)=g(m\beta)$, then $m\beta=f(n\alpha)=na\beta+n\tau$. But then $m=na$ and $n\tau=0$, so $|\tau|$ divides n. Thus $a|\tau|$ divides m. □

From Eq. (2.4) and the comments after that, we obtain the following proposition.

Proposition 2.4: For all $\sigma \in T$, the integer $b_{e(T)}(\sigma)$ is divisible by $b = e(T)/a$. □

Observe that if $g(\beta) = 0$, then $\beta = f(p\alpha) = pa\beta + p\tau$ for some p. Thus $a = 1$ and $b_{e(T)}(\sigma) = 0$ for all $\sigma \in T$. As we noted earlier, the smallest value for nonvanishing $b_{e(T)}(\sigma)$ is $b_{e(T)}(g(\beta)) = e(T)/a$, which we denote by b. Thus we have

$$b/e(T) = 1/a, \tag{2.5}$$

and we obtain the following proposition.

Proposition 2.5: The anomaly coefficient for the sequence (2.1) is given by an integral power of

$$A(\tilde{\omega})_0 \exp\left[\frac{i}{a}\int \gamma^G(\alpha)\right], \tag{2.6}$$

where a is defined by $f(\alpha) = a\beta + \tau$ and α and β are generators of infinite order of $\Pi_{D+1}(G)$ and $\Pi_{D+1}(G/H)$, respectively. □

The subscript 0 indicates the fact that Eq. (2.6) is the smallest possible nontrivial anomaly.

Remark 2.6: If we write $\bar{b}_m(\sigma)$ for the reduction of $b_m(\sigma)$ (mod m), then $\bar{b}_m: {}_mT \to \mathbf{Z}/m\mathbf{Z}$ is the connecting homomorphism in the long exact cohomology sequence of the group $\mathbf{Z}/m\mathbf{Z} = \mathbf{Z}_m$ with respect to the exact coefficient sequence (2.1). We discuss this aspect in the next subsection.

Because of Proposition 2.5, the calculation of the anomaly reduced to the calculation of a. The next example gives the calculation of a, given special T and T'. In the next section, we discuss a different way of calculating a.

Example 2.7: Consider the exact sequence (Lemma A.6 in Appendix A)

$$0 \to \mathbf{Z} \xrightarrow{f} \mathbf{Z} \oplus \mathbf{Z}_l \xrightarrow{g} \mathbf{Z}_{lm} \to 0.$$

We show that one can choose generators, α, β, γ, δ of $\mathbf{Z}$, $\mathbf{Z} \oplus \mathbf{Z}_l$, $\mathbf{Z}_l$, and $\mathbf{Z}_{lm}$, respectively, such that

$$f(\alpha) = m\beta - q\gamma, \quad g(\beta) = q\delta, g(\gamma) = m\delta. \tag{2.7}$$

Thus we find that $a = m$.

Choose generators $\alpha \in \mathbf{Z}$, $\beta \in \mathbf{Z} \oplus \mathbf{Z}_l$, $\gamma' \in \mathbf{Z}_l$, and $\delta \in \mathbf{Z}_{lm}$ so that $f(\alpha) = a\beta + c'\gamma'$, where a is a positive integer and $g(\beta) = q\delta$, $g(\gamma') = k\delta$. Since $0 = g(l\gamma') = lk\delta$, we have $lk = mlu$ or $k = mu$. We must have $(l,u) = 1$, since $g([l/(l,u)]\cdot\gamma') = ml\cdot[u/(l,u)]\cdot\delta = 0$ and thus γ' has an order less than l if $(l,u) \neq 1$, since g restricted to $\mathbf{Z}_l$ is injective (Lemma A.6 in Appendix A). Let $rl + su = 1$ for integers r and s and let $\gamma = s\gamma'$. Then γ is also a generator of $\mathbf{Z}_l$, since $(s,l) = 1$. We have $g(\gamma) = g(s\gamma') = sg(\gamma) = msu\delta = m(1 - rl)\delta = m\delta$. Thus we write $f(\alpha) = a\beta + c\gamma, g(\beta) = q\delta, g(\gamma) = m\delta$. Since g is onto, for integers r' and s', $\delta = g(r'\beta + s'\gamma) = r'q\delta + s'm\delta$, so $r'q + s'm = 1 + tml$, or $r'q + (s' - tl)m = 1$, and $(q,m) = 1$. By exactness, $0 = g(a\beta + c\gamma) = (qa + cm)\delta$, so $qa + cm = uml$ or $qa = m(ul - c)$. But $(q,m) = 1$, so $a = a'm$. Now $g(m\beta - q\gamma) = 0$, so by exactness $m\beta - q\gamma = f(k\alpha) = ka\beta + kc\gamma$ and thus $m = ka = ka'm$ and $kc = -q + vl$. But, then $ka' = 1$ so $k = a' = 1$ and $c = -q + vl$.

In a sense, this example is the prototype example. If we let $\langle\sigma\rangle$ denote the subgroup generated by σ, then we have a short exact sequence

$$0 \to \mathbf{Z} \to \mathbf{Z} \oplus \langle\tau\rangle \to \langle g(\beta)\rangle \to 0.$$

This exact sequence is not independent of the choice of generators, but $b = e_{\langle g(\beta)\rangle}/a = |\tau|$. By varying the choice of the generator β, one obtains different values of b. We will not pursue this line of development, preferring to present the calculation in the form of cohomology theory which is independent of generator choice.

C. A cohomology sequence

Let m be an integer, and if A is an Abelian group, let $d_m: A \to A$ be the homomorphism $d_m(\gamma) = m\gamma$. Consider the short exact sequence of cochain complexes

$$\begin{array}{ccccccc}
0 \to & \mathbf{Z} & \xrightarrow{f} & \mathbf{Z} \oplus T' & \xrightarrow{g} & T \to 0 \\
& \downarrow{\scriptstyle \bar{d}^0 = d_m} & & \downarrow{\scriptstyle d^0 = d_m} & & \downarrow{\scriptstyle \bar{d}^0 = d_m} \\
0 \to & \mathbf{Z} & \xrightarrow{f} & \mathbf{Z} \oplus T' & \xrightarrow{g} & T \to 0 \\
& \downarrow{\scriptstyle \bar{d}^1} & & \downarrow{\scriptstyle d^1} & & \downarrow{\scriptstyle \bar{d}^1} \\
& 0 & & 0 & & 0
\end{array}.$$

It is important to note that the vertical sequences (the complexes) are *not exact*; the bottom row of zeros is intended to indicate that we are working with a one-dimensional complex, i.e., that all cochains of dimension greater than 1 are trivial. Recall that for a complex of cochains, a cohomology group is $H^k = \text{Ker}\, d^k/\text{Im}\, d^{k-1}$, or just $\text{Ker}\, d^0$ in dimension 0. It is a standard result[18] that a short exact sequence of cochain complexes yields a long exact sequence of cohomology groups. In our case, this results in the six term sequence

$$0 \to H^0(\mathbf{Z}) \xrightarrow{f_*} H^0(\mathbf{Z} \oplus T') \xrightarrow{g_*} H^0(T) \xrightarrow{d_{m*}} H^1(\mathbf{Z})$$
$$\xrightarrow{f_*} H^1(\mathbf{Z} \oplus T') \xrightarrow{g_*} H^1(T) \to 0. \tag{2.8}$$

These cohomology groups are easy to calculate. If A is one of the groups $\mathbf{Z}$, $\mathbf{Z} \oplus T'$, T, then ${}_mA = \text{Ker}\, d_m$ while $mA = \text{Im}\, d_m$. Since the complexes above all have $d^1 = 0$ and ${}_m\mathbf{Z} = 0$, we have the following result.

Proposition 2.8: The long exact sequence of Eq. (2.8) is

$$0 \to {}_mT' \xrightarrow{g_*} {}_mT \xrightarrow{d_{m*}} \mathbf{Z}_m \xrightarrow{f_*} \mathbf{Z}_m \oplus T'/mT' \xrightarrow{g_*} T/mT \to 0. \quad \square$$

We next analyze the maps in this sequence. The map $g_*: {}_mT' \to {}_mT$ is just the restriction of g. If $h: A \to A'$ is a homomorphism, let $\bar{h}: A/mA \to A'/mA'$ be the induced homomorphism, and let $\bar{\gamma} \in A/mA$ denote the reduction of $\gamma \in A$. Then, the two rightmost maps in the sequence are $\bar{f}$ and $\bar{g}$, and with our earlier choices of generators $\bar{f}(\bar{\alpha}) = a\bar{\beta} + \bar{\tau}$. Thus our exact sequence is

$$0 \to {}_mT' \xrightarrow{g} {}_mT \xrightarrow{d_{m*}} \mathbf{Z}_m \xrightarrow{\bar{f}} \mathbf{Z}_m \oplus T'/mT' \xrightarrow{\bar{g}} T/mT \to 0.$$

The homomorphism d_{m*} is just $\bar{b}_m$ as described in Remark 2.6 above.

We further analyze the connecting homomorphism $d_{m*} = \bar{b}_m$. Now Im $\bar{f}$ is a cyclic subgroup of $\mathbb{Z}_m \oplus T'/mT'$ generated by $\bar{f}(\bar{\alpha}) = a\bar{\beta} + \bar{\tau}$, and since this is a direct sum decomposition, the order of $a\bar{\beta} + \bar{\tau}$ is the least common multiple of the orders of $a\bar{\beta}$ and $\bar{\tau}$.

Lemma 2.9: For each integer m,

(i) $|a\bar{\beta}| = m/(m,a)$,

(ii) $|\bar{\tau}| = (m,|\tau|)$.

Proof: For (i), clearly $m/(m,a)\cdot a\bar{\beta} = m[a/(m,a)] \times \bar{\beta} = 0$. Let $rm + sa = (m,a)$ for integers r and s and suppose $ta\bar{\beta} = 0$. Then, $ta = qm$, so $t(m,a) = trm + tsa = (tr + qs)m$ and $m/(a,m)$ divides t.

For (ii) write $(m,|\tau|) = rm + s|\tau|$ for integers r and s. Then $(m,|\tau|)\bar{\tau} = (rm + s|\tau|)\bar{\tau} \equiv 0$. If $t\bar{\tau} \equiv 0$, then $t\tau = mq\tau$, and $(t - mq)\tau = 0$ or $t = mq + k|\tau|$. But, then $(m,|\tau|)$ divides t. □

Corollary 2.10: The least common multiple of the orders of $a\bar{\beta}$ and $\bar{\tau}$ is

$$l = m[(m,|\tau|)/(m,a|\tau|)].$$

Proof: Note the fact that the least common multiple of two positive integers is given by $ab/(a,b)$, and $(m,xy) = (m,(m,x)(m,y))$. □

Our exact sequence has been reduced to

$$0 \to {}_mT' \xrightarrow{g} {}_mT \xrightarrow{d_{m*}} \mathbb{Z}_m \xrightarrow{\bar{f}} \mathbb{Z}_l \to 0,$$

where $l = m[(m,|\tau|)/(m,a|\tau|)]$.

Proposition 2.11: For each integer m

(i) Ker $d_{m*} = {}_mT'$,

(ii) Im $d_{m*} = l\mathbb{Z}_m$.

Proof: By exactness. □

Again, from (i) we see that the torsion part of $\Pi_{2n+1}(G/H)$ does not contribute to the anomaly.

Corollary 2.12: For each m, (i) if m divides $|\tau|$, then Im $d_{m*} = 0$; (ii) if m divides $a|\tau|$, then Im $d_{m*} = (|\tau|,m)\mathbb{Z}_m$; (iii) if $a|\tau|$ divides m, then Im $d_{m*} = (m/a)\mathbb{Z}_m$.

Proof: This is just a calculation of l. □

To calculate the number b, one observes that the reduction of b (mod m) is $b_m(g(\beta))$. In the case we are interested in, we take $m = e(T)$ in which case $a|\tau|$ divides $e(T)$ (Lemma 2.3). We apply part (iii) of Corollary 2.12 to obtain the following proposition.

Proposition 2.13: For the exact sequence Eq. (1.11), we have

$$b/e(T) = 1/a. \quad □$$

This is the same as Eq. (2.5).

D. Determination of *a*

In the previous two subsections, we have shown that the calculation of the global anomaly of H is reduced to the determination of a and $\int \gamma^G$. Now, we give a method of calculating a, by comparing two exact sequences. For this purpose, we discuss a diagram where two exact sequences are connected by homomorphisms.

Suppose we have homomorphisms, h, h', h'' so that

$$\begin{array}{ccccccccc} 0 & \to & \mathbb{Z} & \xrightarrow{f_0} & \mathbb{Z} \oplus T'_0 & \xrightarrow{g_0} & T_0 & \to & 0 \\ & & \downarrow h' & & \downarrow h & & \downarrow h'' & & \\ 0 & \to & \mathbb{Z} & \xrightarrow{f_1} & \mathbb{Z} \oplus T'_1 & \xrightarrow{g_1} & T_1 & \to & 0 \end{array} \tag{2.9}$$

is commutative with exact rows and T_i, T'_i $(i = 0,1)$, are finite Abelian. Generators α_i, β_i, and elements $\tau_i \in T'_i$ may be chosen so that $f_i(\alpha_i) = a_i\beta_i + \tau_i (\alpha_i \neq 0)$ as earlier. Moreover, for integers, c, c' and $\sigma \in T'_1$ we have

$$h'(\alpha_0) = c'\alpha_1, \quad h(\beta_0) = c\beta_1 + \sigma. \tag{2.10}$$

Commutativity of the first square of the diagram yields

$$c'a_1 = ca_0, \quad c'\tau_1 = a_0\sigma + h(\tau_0). \tag{2.11}$$

Proposition 2.14: For the diagram (2.9) above (i) the map h' is either injective or trivial; (ii) if h'' and h' are injective, then h is injective; (iii) if h'' is onto and h' is injective, then $(c, a_1) = 1$ and c' is a multiple of c.

Proof: For part (i), let $\alpha = m\alpha_0$. If $0 = h'(\alpha) = mc'\alpha_1$, then $mc' = 0$. If $c' = 0$, h' is trivial and if $c' \neq 0$, then $m = 0$, so $\alpha = 0$ which means that h' is injective. For part (ii), suppose that $h(\theta) = 0$. Then $0 = g_1h(\theta) = h''g_0(\theta)$, and since h'' is injective $g_0(\theta) = 0$. By exactness, $\theta = f_0(\alpha)$ and $0 = h(\theta) = hf_0(\alpha) = f_1h'(\alpha)$. But f_1h' is injective so $\alpha = 0$ and $\theta = f_0(\alpha) = 0$. Thus h is injective. For part (iii), observe that $g_1(\beta_1) = h''g_0(r\beta_0 + \rho)$ for some integer r and $\rho \in T'_0$, since g_0 and h'' are onto. But then $g_1(\beta_1) = g_1h(r\beta_0 + \rho)$, or $g_1(\beta_1 - h(r\beta_0 + \rho)) = 0$, so that $\beta_1 - rc\beta_1 - r\sigma - h(\rho) = \beta_1 - h(r\beta_0 + \rho) = f_1(s\alpha_1) = sa_1\beta_1 + s\tau_1$. From this, $(1 - rc) = sa_1$ and $-r\sigma - h(\rho) = s\tau_1$. The first of these equations, $1 = rc + sa_1$, shows that $(c,a_1) = 1$. On the other hand, $c'a_1 = ca_0$, so c' is a multiple of c (and a_0 is a multiple of a_1). □

Consequently, using the known sequence at the bottom of the diagram (2.9), we can get information on a of the top sequence. We illustrate the usage of this method for classical groups in the next section. Note that one can also study the diagram (2.9) from a cohomological point of view. We will not do this here.

III. JAMES NUMBERS AND CALCULATION OF *a*

A. Special unitary groups

We consider complex Stiefel manifolds $W_{n,k} = \mathrm{SU}(n)/\mathrm{SU}(n-k)$ and the fibrations

$$\mathrm{SU}(n-k) \to \mathrm{SU}(n+1) \xrightarrow{p} W_{n+1,k+1},$$

$$\mathrm{SU}(n) \to \mathrm{SU}(n+1) \xrightarrow{\bar{q}} S^{2n+1},$$

$$W_{n,k} \to W_{n+1,k+1} \xrightarrow{q} S^{2n+1}.$$

The homotopy sequences fit together in a diagram:

$$
\begin{array}{ccccccc}
 & & & \Pi_{2n+2}(S^{2n+1}) & & \to\Pi_{2n+1}(\mathrm{SU}(n)) \swarrow & \\
 & & & \downarrow \partial_* & & & \\
 & & & \Pi_{2n+1}(W_{n,k}) & & & \\
 & & & \downarrow i_* & \searrow \Delta'_* & & \\
\Pi_{2n+1}(\mathrm{SU}(n-k)) & \to & \Pi_{2n+1}(\mathrm{SU}(n+1)) & \xrightarrow{p_*} \Pi_{2n+1}(W_{n+1,k+1}) & \xrightarrow{\Delta_*} & \Pi_{2n}(\mathrm{SU}(n-k)) & \to 0 \\
\downarrow i'_* & & \downarrow \cong & \downarrow q_* & & \downarrow i'_* & \\
\Pi_{2n+1}(\mathrm{SU}(n)) & \to & \Pi_{2n+1}(\mathrm{SU}(n+1)) & \xrightarrow{\bar q_*} \Pi_{2n+1}(S^{2n+1}) & \xrightarrow{\bar\Delta_*} & \Pi_{2n}(\mathrm{SU}(n)) & \to 0.
\end{array}
$$

Note that $\Pi_{2n+1}(\mathrm{SU}(n-k))$ is of finite order (Appendix B) and has image 0 in $\Pi_{2n+1}(\mathrm{SU}(n+1)) = \mathbb{Z}$ (Lemma A.4 in Appendix A). From the work of Toda,[19] $\mathrm{Im}\, \partial_* = 0$ when n is odd and is at most of order 2 when n is even.

Choose $\iota \in \Pi_{2n+1}(S^{2n+1})$ to be the homotopy class of the identity map. By theorems of Bott[20] and Borel–Hirzebruch,[21] we may choose a generator $\alpha \in \Pi_{2n+1}(\mathrm{SU}(n+1))$ so that $\bar{q}_*(\alpha) = n!\iota$. Now, choose a generator $\beta \in \Pi_{2n+1}(W_{n+1,k+1})$ so that $q_*(\beta) = W\{n+1,k+1\}\cdot\iota$, where $W\{n+1,k+1\}$ is the James number.[22] With these choices, the central exact sequence

$$0 \to \Pi_{2n+1}(\mathrm{SU}(n+1)) \xrightarrow{p_*} \Pi_{2n+1}(W_{n+1,k+1})$$
$$\xrightarrow{\Delta_*} \Pi_{2n}(\mathrm{SU}(n-k)) \to 0$$

is the short exact sequence in Sec. II, where

$$p_*(\alpha) = a(n+1,k+1)\beta + \tau,$$
$$a(n+1,k+1) = n!/W\{n+1,k+1\},$$

and $\tau \in T' = \mathrm{Im}\, i_* \cong \Pi_{2n+1}(W_{n,k})/\mathrm{Im}\, \partial_*$. Therefore, once we know the James number, we can calculate $a(n+1,k+1)$. Using Proposition 2.5, we have the following proposition.

Proposition 3.1: For the special unitary group $\mathrm{SU}(n-k)$, the anomaly is given by

$$A(\tilde{\omega})_0 = \exp\left[i\, \frac{W\{n+1,k+1\}}{n!} \int \gamma^G(\alpha) \right],$$

where $G = \mathrm{SU}(n+1)$. □

The James numbers $W^-\{n+1,k+1\}$ are known for $0 \leqslant k \leqslant 7$ by the works of Sigrist, Ōshima, and Knapp,[23] and for $k = n-2$ and $n-3$ by the work of Walker.[24] A different method used in Refs. 7 and 8 has shown that any group has only a $\mathbb{Z}_2$ anomaly, and any group which has only self-contragredient representations has no anomalies in dimensions $D \equiv 2,6 \pmod 8$, provided that the theory is locally anomaly-free in the sense discussed in Sec. I. Thus the complete knowledge of $\int \gamma^G(\alpha)$ is not needed to examine whether anomalies exist or not. As we see in the following examples, it is sometimes sufficient to investigate whether the integral is even or odd.

Example 3.2: For $k = 0$ [i.e., $H = \mathrm{SU}(n)$], the James number is given by

$$W\{n+1,1\} = 1. \tag{3.1}$$

In Ref. 7, it was shown that $\int \gamma^G(\alpha)/2\pi n!$ is an integer for $n =$ odd and a half-integer for $n =$ even, using the local anomaly-free condition. Thus no anomalies exist for $D \equiv 2 \pmod 4$, but $\mathbb{Z}_2$ anomalies exist for $D \equiv 0 \pmod 4$.

Example 3.3: For $k = 1$ [i.e., $H = \mathrm{SU}(n-1)$], it is known that

$$W\{n+1,2\} = 2/(n+1,2) = (n,2) \quad \text{for } n \geqslant 3. \tag{3.2}$$

Thus if $\int \gamma^G(\alpha)/2\pi$ is divisible by $n!/(n,2)$, then no anomalies exist. It was shown in Ref. 7 that this is indeed the case when the local anomaly-free condition is satisfied.

Example 3.4: For $k = n-2$ [i.e., $H = \mathrm{SU}(2)$], Walker[24] has shown that

$$a(n+1,n-1) = \begin{cases} \text{denom}\, B_{n-1}, & n \equiv 3 \pmod 4, \\ \tfrac{1}{2}\, \text{denom}\, B_{n-1}, & n \equiv 1 \pmod 4, \\ 1, & n \equiv 0 \pmod 4, \\ 2, & n \equiv 2 \pmod 4, \end{cases}$$

where denom B_{n-1} is the denominator of a Bernoulli number and $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$, etc. Note that the denominator of a Bernoulli number B_k is a product of primes, p that satisfy $p-1 | k$ and $\nu_p(B_k) = 0$ or 1. Thus $\nu_p(a(n+1,n-1)) = 0$ or 1. The anomaly is given by

$$A(\tilde{\omega})_0 = \begin{cases} \exp\left[i(1/\text{denom}\, B_{n-1}) \int \gamma^G(\alpha) \right], & n \equiv 3 \pmod 4, \\ \exp\left[i(2/\text{denom}\, B_{n-1}) \int \gamma^G(\alpha) \right], & n \equiv 1 \pmod 4, \\ \exp\left[i \int \gamma^G(\alpha) \right], & n \equiv 0 \pmod 4, \\ \exp\left[i\tfrac{1}{2} \int \gamma^G(\alpha) \right], & n \equiv 2 \pmod 4. \end{cases}$$

Consequently, for $D=2n\equiv 0 \pmod 8$ SU(2) has no anomalies, since $\int\gamma^G(\alpha)$ is 2π times some integer. For $D=2n\equiv 4 \pmod 8$, no anomalies exist only when $\int\gamma^G(\alpha)/2\pi$ is even. In Ref. 8, it was shown that for all irreducible representations of SU(2) except those with dimension $(2J+1)$ with $J=\frac{1}{2}(4l+1)$, this integral is even. For $D\equiv 2,6 \pmod 8$, no anomalies exist and thus we have that $\int\gamma^G(\alpha)/2\pi$ is even for $D\equiv 6 \pmod 8$, since denom $B_{n-1}/2$ contains only odd integers. It is amazing that $\int\gamma^G(\alpha)/2\pi$ has denom $B_{n-1}/2$ as a divisor for $D\equiv 2,6 \pmod 8$.

Example 3.5: For $k=n-3$ [i.e., $H=\mathrm{SU}(3)$], Walker has shown that for n odd, $a(n+1,n-2)=\text{denom } B_{n-1}$. For $D\equiv 2,6 \pmod 8$, SU(3) has no anomalies and thus $\int\gamma^G(\alpha)/2\pi$ is even.

B. Symplectic groups

1. $D\equiv 2 \pmod 4$

By using the quaternionic Stiefel manifolds $X_{n,k}=\mathrm{Sp}(n)/\mathrm{Sp}(n-k)$ and the fibrations analogous to those in the case of SU(n), and by making analogous choices of generators $\alpha\in\Pi_{4n+3}(\mathrm{Sp}(n+1))$ and $\beta\in\Pi_{4n+3}(X_{n+1,k+1})$, one obtains an exact sequence

$$0\to\Pi_{4n+3}(\mathrm{Sp}(n+1))\xrightarrow{p_*}\Pi_{4n+3}(X_{n+1,k+1})\xrightarrow{\Delta_*}\Pi_{4n+2}(\mathrm{Sp}(n-k))\to 0, \tag{3.3}$$

where

$$p_*(\alpha)=a(n+1,k+1)\beta+\tau,\qquad a(n+1,k+1)=\frac{(n+1,2)(2n+1)!}{X\{n+1,k+1\}}. \tag{3.4}$$

Thus we have the following proposition.

Proposition 3.6: In dimensions $D=4n+2$, the anomaly for $H=\mathrm{Sp}(n-k)$ is given by

$$A(\tilde{\omega})_0=\exp\left[i\frac{X\{n+1,k+1\}}{(n+1,2)(2n+1)!}\int\gamma^G(\alpha)\right],$$

where $G=\mathrm{Sp}(n+1)$.

Note that n is defined as $D=4n+2$, not $D=2n$. However, as we mentioned in the previous subsection, for $D\equiv 2 \pmod 4$, symplectic groups do not have anomalies, since they have only self-contragredient representations. Thus $\int\gamma^G(\alpha)$ must be divisible by $2\pi[(n+1,2)(2n+1)!/X\{n+1,k+1\}]$.

The numbers $X\{n+1,k+1\}$ are known for $0\leqslant k\leqslant 4$ from the work of Ohshima.[23] A calculation based on the work of Walker[24,25] establishes the following lemma for $H=\mathrm{Sp}(1)\simeq\mathrm{SU}(2)$.

Lemma 3.7: For $n\geqslant 1$, we have

$$X\{n+1,n\}=W\{2n+2,2n\}=\begin{cases}(2n+1)!/\text{denom } B_{2n}, & n\equiv 1 \pmod 2,\\ 2(2n+1)!/\text{denom } B_{2n}, & n\equiv 0 \pmod 2.\end{cases}$$

□

2. Comparison with special unitary groups

We can improve Proposition 3.6 somewhat by comparing the symplectic groups with special unitary groups.

The standard embeddings $\mathrm{Sp}(n-k)\subset\mathrm{Sp}(n+1)\subset\mathrm{SU}(2n+2)$ lead to the following (horizontal and vertical) fiber bundles, where $Y_{n+1,k+1}\equiv\mathrm{SU}(2n+2)/\mathrm{Sp}(n-k)$ for $k\geqslant -1$:

$$\begin{array}{ccccc}
\mathrm{Sp}(n-k) & \to & \mathrm{Sp}(n+1) & \to & X_{n+1,k+1}\\
\downarrow\cong & & \downarrow & & \downarrow\\
\mathrm{Sp}(n-k) & \to & \mathrm{SU}(2n+2) & \to & Y_{n+1,k+1}\\
 & & \downarrow & & \downarrow\\
 & & Y_{n+1,0} & \xrightarrow{\cong} & Y_{n+1,0}
\end{array}.$$

Passing to homotopy, one obtains a commutative diagram with exact rows and columns:

$$\begin{array}{ccccccccc}
 & & \Pi_{4n+4}(Y_{n+1,0}) & \xrightarrow{\cong} & \Pi_{4n+4}(Y_{n+1,0}) & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \to & \Pi_{4n+3}(\mathrm{Sp}(n+1)) & \to & \Pi_{4n+3}(X_{n+1,k+1}) & \to & \Pi_{4n+2}(\mathrm{Sp}(n-k)) & \to & 0\\
 & & \downarrow & & \downarrow & & \downarrow\cong & & \\
0 & \to & \Pi_{4n+3}(\mathrm{SU}(2n+2)) & \to & \Pi_{4n+3}(Y_{n+1,k+1}) & \to & \Pi_{4n+2}(\mathrm{Sp}(n-k)) & \to & 0\\
 & & \downarrow & & \downarrow & & & & \\
 & & \Pi_{4n+3}(Y_{n+1,0}) & \xrightarrow{\cong} & \Pi_{4n+3}(Y_{n+1,0}) & & & &
\end{array}. \tag{3.5}$$

This yields two sequences of the type we analyze:

$$\begin{array}{ccccccc} & 0 & \rightarrow & 0 & & & \\ & \downarrow & & \downarrow & & & \\ 0\rightarrow & \mathbb{Z} & \xrightarrow{f_0} & \mathbb{Z}\oplus T_0' & \xrightarrow{g_0} & T_0 & \rightarrow 0 \\ & \downarrow h' & & \downarrow h & & \downarrow h'' & \\ 0\rightarrow & \mathbb{Z} & \xrightarrow{f_1} & \mathbb{Z}\oplus T_1' & \xrightarrow{g_1} & T_1 & \rightarrow 0. \\ & \downarrow & & \downarrow & & & \\ & \mathbb{Z}_{(n+1,2)} & \xrightarrow{\cong} & \mathbb{Z}_{(n+1,2)} & & & \\ & \downarrow & & \downarrow & & & \\ & 0 & & 0 & & & \end{array} \tag{3.6}$$

Using the same notation as in Sec. II D, we know that $h'(\alpha_0)=(n+1,2)\alpha_1$. By Proposition 2.14, h is injective and is an isomorphism if n is even. Thus for $n=2m$, we may use the sequence

$$0\rightarrow \Pi_{4n+3}(\mathrm{SU}(2n+2))\xrightarrow{p_*}\Pi_{4n+3}(Y_{n+1,k+1})$$
$$\rightarrow \Pi_{4n+2}(\mathrm{Sp}(n-k))\rightarrow 0, \tag{3.7}$$

where

$$p_*(\alpha)=a(n+1,k+1)\alpha+\tau,$$
$$a(n+1,k+1)=(2n+1)!/X\{n+1,k+1\}. \tag{3.8}$$

The case $n=2m+1$ is more complicated. From the commutative diagram

$$\begin{array}{ccc} \Pi_{8m+7}(\mathrm{Sp}(2m+2)) & \xrightarrow{p'_{0*}} & \Pi_{8m+7}(X_{2m+2,2m+1}) \\ \downarrow p_{0*} & & \downarrow \\ \Pi_{8m+7}(X_{2m+2,k+1}) & \xrightarrow{\cong} & \Pi_{8m+7}(X_{2m+2,k+1}) \end{array}$$

and the fact that

$$p'_{0*}(\alpha)=\frac{2(4m+3)!}{X\{2m+2,2m+1\}}\beta'_0+\tau'_0$$
$$=2(\operatorname{denom} B_{4m+2})\beta'_0+\tau'_0,$$

by Lemma 3.7, it follows that

$$p_{0*}(a)=a_0\beta_0+\tau_0,$$

and a_0 is divisible by 4. Referring to the diagram (3.6) and letting $h(\beta_0)=c\beta_1+\sigma$, we see that $ca_0=2a_1$, so $a_1=ca_0/2$ is even. By Proposition 2.14, $(c,a_1)=1$ and c divides 2. Thus $c=1$. We may use the sequence

$$0\rightarrow \Pi_{8m+7}(\mathrm{SU}(4m+4))\xrightarrow{p_*}\Pi_{8m+7}(Y_{2m+2,k+1})$$
$$\rightarrow \Pi_{8m+6}(\mathrm{Sp}(2m-k))\rightarrow 0,$$

where

$$p_*(\alpha)=a(2m+2,k+1)\beta+\tau,$$
$$a(2m+2,k+1)=(4m+3)!/X\{2m+2,k+1\}.$$

Thus in both cases we may use

$$0\rightarrow \Pi_{4n+3}(\mathrm{SU}(2n+2))\xrightarrow{p_*}\Pi_{4n+3}(Y_{n+1,k+1})$$
$$\rightarrow \Pi_{4n+2}(\mathrm{Sp}(n-k))\rightarrow 0, \tag{3.9}$$

where

$$p_*(\alpha)=a(n+1,k+1)\beta+\tau,$$
$$a(n+1,k+1)=(2n+1)!/X\{n+1,k+1\}. \tag{3.10}$$

Proposition 3.8: For $D=4n+2$, $H=\mathrm{Sp}(n-k)$, and $G=\mathrm{SU}(2n+2)$, the anomaly coefficient is given by

$$A(\tilde{\omega})_0=\exp\left[i\frac{X\{n+1,k+1\}}{(2n+1)!}\int\gamma^G(\alpha)\right]. \qquad \square$$

Note that this improves by a factor of 2 the multiplier of the integral when we take $G=\mathrm{Sp}(n+1)$. Moreover, from the work of Oshima,[23] $X\{n+1,k+1\}=2^rW\{2n+2,2k+2\}$ with $r\geqslant 0$, so that the anomaly coefficient can be written as

$$A(\tilde{\omega})_0=\left(\exp\left[i\frac{W\{2n+2,2k+2\}}{(2n+1)!}\int\gamma^G(\alpha)\right]\right)^{2^r}.$$

Corollary 3.9: In dimension $D=4n+2$, the anomaly coefficient for $H=\mathrm{Sp}(n-k)$ is a 2^rth power of the anomaly coefficient for $H=\mathrm{SU}(2n-2k)$ with the representation ω' obtained from $\tilde{\omega}$ of $\mathrm{SU}(2n+2)$. □

3. D≡0 (mod 4)

As we mentioned in Sec. I, in dimensions $D\equiv 4 \pmod 8$ we cannot use $G=\mathrm{Sp}(\tilde{N})$ for $H=\mathrm{Sp}(N)$. Therefore we examine the choice of $G=\mathrm{SU}(\tilde{N})$ with $\tilde{N}\geqslant 2n+1$ for $H=\mathrm{Sp}(n-k)$ in dimensions $D=4n$.

It follows from the work of Walker[24] that the following lemma holds.[25]

Lemma 3.10: For $1\leqslant k\leqslant 2m$ the inclusion j: $\mathrm{Sp}(2m+1-k)\rightarrow \mathrm{Sp}(2m+1)$ induces j_*: $\Pi_{8m+4}(\mathrm{Sp}(2m+1-k))\rightarrow \Pi_{8m+4}(\mathrm{Sp}(2m+1))$ that is *onto*. □

Now consider the diagram where the bottom sequence is in the stable range:

$$\begin{array}{ccccc} 0\rightarrow \Pi_{4n+1}(\mathrm{SU}(\tilde{N})) & \xrightarrow{p_*} & \Pi_{4n+1}(\mathrm{SU}(\tilde{N})/\mathrm{Sp}(n-k)) & \rightarrow & \Pi_{4n}(\mathrm{Sp}(n-k))\rightarrow 0 \\ \downarrow \cong & & \downarrow q_* & & \downarrow j_* \\ 0\rightarrow \Pi_{4n+1}(\mathrm{SU}(\tilde{N})) & \xrightarrow{\tilde{p}_*} & \Pi_{4n+1}(\mathrm{SU}(\tilde{N})/\mathrm{Sp}(n)) & \rightarrow & \Pi_{4n}(\mathrm{Sp}(n))\rightarrow 0. \end{array}$$

Since j_* is onto [by Lemma 3.10 if n is odd and $\Pi_{4n}(\mathrm{Sp}(n)) = 0$ if n is even], it follows that q_* is onto. For generators $\alpha \in \Pi_{4n+1}(\mathrm{SU}(2n+1))$ and $\gamma \in \Pi_{4n+1}(\mathrm{SU}(2n+1)/\mathrm{Sp}(n)) = \mathbb{Z}$,[26] we have $\bar{p}_*(\alpha) = (n+1,2)\gamma$ (Example 2.7 with $l = 1$ if n odd and Lemma A.3 if n is even). From Lemma A.7 in Appendix A, we can choose a generator $\beta \in \Pi_{4n+1}(\mathrm{SU}(2n+1)/\mathrm{Sp}(n-k))$ with $q_*(\beta) = \gamma$, and then $p_*(\alpha) = (n+1,2)\beta + \tau$ where τ is of finite order. We therefore obtain

$$A(\tilde{\omega})_0 = \exp\left[i \frac{1}{(n+1,2)} \int \gamma^G(\alpha)\right] \tag{3.11}$$

for the anomaly coefficient for $H = \mathrm{Sp}(n-k)$ with $G = \mathrm{SU}(2n+1)$ in dimension $4n$. Therefore in dimension $D \equiv 0 \pmod 8$, no anomaly exists, confirming the analysis of Sec. I. In dimension $D \equiv 4 \pmod 8$, the anomaly is at most of $\mathbb{Z}_2$ type, confirming the differential geometric analysis in Ref. 7.

C. Special orthogonal groups

As mentioned in Sec. I, the representation embedding condition can be satisfied only for tensor representations (excluding self-dual ones) for $H = \mathrm{SO}(N)$ with $G = \mathrm{SO}(\tilde{N})$. We discuss only tensor representations, excluding self-dual ones and spinors in the case where $G = \mathrm{SO}(\tilde{N})$. The Bott periodicity theorem [Eq. (1.14)] tells us that in $D \equiv 4 \pmod 8$ no anomalies exist for tensor representations of $\mathrm{SO}(N)$ in the unstable range.

1. $D \equiv 2 \pmod 4$

For $D \equiv 2,6 \pmod 8$, we have $\Pi_{4n+2}(\mathrm{SO}) = 0$ and $\Pi_{4n+3}(\mathrm{SO}) = \mathbb{Z}$. However, the situation differs from the cases of $\mathrm{SU}(N)$ and $\mathrm{Sp}(N)$ in that the homotopy of $\mathrm{SO}(N)$ in the unstable region can be of positive rank.

Since $\Pi_k(S^n)$ is finite except for $k = n$ and $k = 2m-1$, $n = 2m$, one knows that $i_*: \Pi_{4n+3}(\mathrm{SO}(4n+5-j)) \to \Pi_{4n+3}(\mathrm{SO}(4n+5)) = Z$ is nonzero for $0 \leqslant j \leqslant 2n+2$, and hence that $\Pi_{4n+3}(V_{4n+5,j})$ is finite, where $V_{4n+3,j}$ is the real Stiefel manifold. Using Proposition 1.1, we have the following proposition.

Proposition 3.11: For $D = 4n+2$ and $m \geqslant 2n+3$, no anomalies exist for $H = \mathrm{SO}(m)$ with tensor representations (except for self-dual representations). □

Thus we have derived part of the more general theorem obtained in Ref. 7 that self-contragredient representations have no anomalies in $D \equiv 2 \pmod 4$.

For other cases, we must do an analysis similar to $\mathrm{SU}(N)$ and $\mathrm{Sp}(N)$. There is a (highly technical) homotopy construction called *localization* that enables one to construct for each space X a related space $X_{\mathscr{S}}$ and a map $l_{\mathscr{S}}: X \to X_{\mathscr{S}}$, where $\mathscr{S}$ is a set of primes. The map $l_{\mathscr{S}}$ induces homomorphisms in homotopy that preserve all p-primary information for $p \in \mathscr{S}$ and delete all q-primary information for q a prime not in $\mathscr{S}$. If one lets $\mathscr{S}$ be the set of odd primes, there is a map h: $\mathrm{SO}(2n+1)_{\mathscr{S}} \to \mathrm{Sp}(n)_{\mathscr{S}}$ that is a homomorphism up to homotopy and is a homotopy equivalence. One has a commutative diagram

$$\begin{array}{ccccccc} 0 \to & \Pi_{4n+3}(\mathrm{SO}(2n+3)_{\mathscr{S}}) & \to & \Pi_{4n+3}(\mathrm{SO}(2n+3)_{\mathscr{S}}, \mathrm{SO}(m)_{\mathscr{S}}) & \to & \Pi_{4n+2}(\mathrm{SO}(m)_{\mathscr{S}}) & \to 0 \\ & \downarrow & & \downarrow & & \downarrow & \\ 0 \to & \Pi_{4n+3}(\mathrm{Sp}(n+1)_{\mathscr{S}}) & \to & \Pi_{4n+3}(\mathrm{Sp}(n+1)_{\mathscr{S}}, \mathrm{Sp}(\tilde{m})_{\mathscr{S}}) & \to & \Pi_{4n+2}(\mathrm{Sp}(\tilde{m})_{\mathscr{S}}) & \to 0 \end{array}, \tag{3.12}$$

where $m = 2(n-k)+1$ and $\tilde{m} = n-k$. Thus, for odd primes, the computation of b_p for $\mathrm{SO}(2(n-k)+1)$ is the same as for $\mathrm{Sp}(n-k)$. Unfortunately, very little is known about the behavior at prime 2.

2. $D \equiv 0 \pmod 8$

We choose $G = \mathrm{SU}(\tilde{N})$ with $\tilde{N} \geqslant 8m+3$ for $H = \mathrm{SO}(8m+3-k)$. In dimensions $D = 8m$, we can use the result that the map $\Pi_{8m}(\mathrm{SO}(8)) \to \Pi_{8m}(\mathrm{SO}(8m+2)) = \mathbb{Z}_2$ is onto.[25,28] Together with the fiber bundles

$$\mathrm{SO}(8m+3-k) \to \mathrm{SU}(8m+3) \to \mathrm{SU}(8m+3)/\mathrm{SO}(8m+3-k)$$

we obtain a homotopy diagram for $1 < k \leqslant 8m-6$:

$$\begin{array}{ccccccc} 0 \to & \Pi_{8m+1}(\mathrm{SU}(\tilde{N})) & \xrightarrow{p_*} & \Pi_{8m+1}(\mathrm{SU}(\tilde{N})/\mathrm{SO}(M-k)) & \to & \Pi_{8m}(\mathrm{SO}(m-k)) & \to 0 \\ & \downarrow \cong & & \downarrow h_* & & \downarrow \text{onto} & \\ 0 \to & \Pi_{8m+1}(\mathrm{SU}(\tilde{N})) & \to & \Pi_{8m+1}(\mathrm{SU}(\tilde{N})/\mathrm{SO}(M)) & \to & \Pi_{8m}(\mathrm{SO}(M)) & \to 0 \end{array}, \tag{3.13}$$

where $M = 8m+2$. The reason for the condition $\tilde{N} \geqslant 8m+3$ is because $\Pi_{8m+1}(\mathrm{SO}(8m+2)) = \mathbb{Z}$. Note that $\Pi_{8m+1}(\mathrm{SU}(8m+3)) = \mathbb{Z} \oplus \mathbb{Z}_2$, $\Pi_{8m+1}(\mathrm{SU}(8m+3))/\mathrm{SO}(8m+3)) = \mathbb{Z}$,[26] and $\Pi_{8m}(\mathrm{SO}(8m+3)) = \mathbb{Z}_2$. It follows from Proposition 2.8 that h_* is onto, and hence that in the top sequence

$$p_*(\alpha) = 2\beta + \tau.$$

Proposition 3.12: In dimension $D = 8m$, an anomaly coefficient for $H = \mathrm{SO}(8m+3-k)$, $1 < k \leqslant 8m-6$, is given by

$$A(\tilde{\omega})_0 = \exp\left[i \frac{1}{2} \int \gamma^G(\alpha)\right],$$

where $G = \mathrm{SU}(8m+3)$. □

Again we have derived part of the theorem derived in Ref. 7 that the only global anomaly for any group is of $\mathbb{Z}_2$ type. Note that two methods are completely different.

3. Other dimensions

In dimension $D = 8n + 4$, if $m \geqslant 8n + 6$, then $\Pi_{8n+4}(\mathrm{SO}(m)) = 0$. If $m \leqslant 8n + 6$ the comparison diagram

$$\begin{array}{ccccccccc} 0 \to & \Pi_{8n+5}(\mathrm{SU}(8n+6)) & \to & \Pi_{8n+5}(\mathrm{SU}(8n+6)/\mathrm{SO}(m)) & \to & \Pi_{8n+4}(\mathrm{SO}(m)) & \to 0 \\ & \downarrow \cong & & \downarrow & & \downarrow & \\ 0 \to & \Pi_{8n+5}(\mathrm{SU}(8n+6)) & \xrightarrow{\cong} & \Pi_{8n+5}(\mathrm{SU}(8n+6)/\mathrm{SO}(8n+6)) & \to & \Pi_{8n+4}(\mathrm{SO}(8n+6)) & \to 0 \\ & & & & & \| & \\ & & & & & 0 & \end{array}$$

shows that $a = 1$. Thus in dimension $8n + 4$ there are no anomalies for tensor representations of $\mathrm{SO}(m)$, excluding self-dual ones.

However, the similar diagrams for $D \equiv 2 \pmod 4$ cannot yield any information, since[16]

$$\begin{array}{ccccc} ? \to \mathbb{Z} \to & ? & \to T \to 0 \\ \downarrow & \downarrow & \downarrow \\ \mathbb{Z} \to \mathbb{Z} \to 0 \text{ or } \mathbb{Z}_2 & \to 0 & \to 0, \end{array}$$

where ? denotes the fact that many possibilities exist.

ACKNOWLEDGMENTS

Y.T. would like to thank Professor Okubo for discussions.

The work of Y.T. was supported in part by the U.S. Department of Energy Contract No. DE-AC02-86ER40253.

APPENDIX A: VARIOUS FACTS ABOUT AN ABELIAN EXACT SEQUENCE

In this Appendix, we collect various facts about an Abelian exact sequence. We use the notation that $T, T', T'', \ldots$ denote finite groups and Z denotes the infinite cyclic group. Here Z_a denotes the finite cyclic group of order a. The first three lemmas can be found in any books on homology, and thus we omit the proofs.

Lemma A.1: For the exact sequence of the form

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} 0$$

the map α is onto.

Lemma A.2: For the exact sequence of the form

$$0 \xrightarrow{\alpha} B \xrightarrow{\beta} C$$

the map β is one to one.

Lemma A.3: If the exact sequence is the following:

$$0 \to A \to B \to 0$$

then A is isomorphic to B.

Lemma A.4: For $T \xrightarrow{\alpha} \mathbb{Z} \oplus T'$, $\quad \mathrm{Im}\, \alpha = \{0 \oplus *\}$.

Let us assume that a generator of T, δ, maps into $p\beta + \tau$ where β is a generator of infinite order of $\mathbb{Z} \oplus T'$ and p is some integer. Then, $n\delta$ is mapped into $np\beta + n\tau$. If $n = e(T)$, then $np\beta + n\tau = 0$. Since $n \neq 0$, we have $p = 0$. □

Lemma A.5: There is no exact sequence of the form

$$T \xrightarrow{\alpha} \mathbb{Z} \oplus T' \xrightarrow{\beta} T''.$$

By Lemma A.4, Im α is in the torsion part. However, Ker β always contains an element of infinite order. Thus it is impossible to have Im α = Ker β. □

Lemma A.6: For the exact sequence of the form

$$T \to \mathbb{Z} \xrightarrow{p_*} \mathbb{Z} \oplus \mathbb{Z}_a \xrightarrow{\Delta_*} \mathbb{Z}_b \to 0,$$

a divides b.

By Lemma A.4, the image of T is 0. Then by Lemmas 1 and 2, the map p_* is one to one and Δ_* is onto. Let us take $x \in \mathbb{Z}_a$ such that $\Delta_*(x) = 0$. Let us denote the generators of $\mathbb{Z}$ and $\mathbb{Z} \oplus \mathbb{Z}_a$ of infinite order by α and β, respectively. Then we have

$$\alpha \xrightarrow{p_*} q\beta + \tau,$$

where $q \neq 0$, since otherwise p_* is not one to one. Because of the exactness, there exists an integer v such that

$$p_*(v\alpha) = x,$$

while

$$p_*(v\alpha) = vq\beta + v\tau.$$

Since $x \in Z_a$, we obtain $v = 0$. That is, the map Δ_* restricted to $\mathbb{Z}_a$ is a one-to-one map. Thus $\mathbb{Z}_b$ contains a subgroup $\mathbb{Z}_a$. Therefore a divides b. □

Lemma A.7: If the homomorphism $f: \mathbb{Z} \oplus T \to \mathbb{Z}$ is onto, then $f(T) = 0$ and generators $\beta \in \mathbb{Z} \oplus T$ and $\gamma \in \mathbb{Z}$ can be chosen to satisfy $f(\beta) = \gamma$.

For $\tau \in T$, suppose that $f(\tau) = d\gamma$ for some integer d. Then $0 = f(m\tau) = md\gamma$ for m being the least common multiple of orders of T. Thus $d = 0$. Since f is onto, we can find some integer r such that $f(r\beta) = \gamma$. But $f(\beta) = c\gamma$ for some integer c, so that $f(r\beta) = cr\gamma = \gamma$. Thus $cr = 1$. □

APPENDIX B: Π_{2n+1} (SU(m)) IS FINITE FOR $m < n$

Using the fibration

$$\mathrm{SU}(m-1) \to \mathrm{SU}(m) \to S^{2m-1},$$

we obtain the exact sequence

$$\Pi_{2n+2}(S^{2m-2}) \to \Pi_{2n+1}(\mathrm{SU}(m-1)) \to \Pi_{2n+1}(\mathrm{SU}(m)) \to \Pi_{2n+1}(S^{2m-1}).$$

We use from Lemma A.2 that $\Pi_q(S^{2m-1})$ is finite for $q > 2m-1$ (see, for example, Ref. 29). Thus, for $m \leqslant n$, both $\Pi_{2n+1}(S^{2m-1})$ and $\Pi_{2n+2}(S^{2m-1})$ are finite. Using Lemma A.5, *either* both $\Pi_{2n+1}(\mathrm{SU}(m-1))$ and $\Pi_{2n+1}(\mathrm{SU}(m))$ are finite *or* both of them contain an infinite part. However, for $m = 3 \leqslant n$ (i.e., $n \geqslant 3$), we know that $\Pi_{2n+1}(\mathrm{SU}(2)) = \Pi_{2n+1}(S^3)$, which is finite. Thus $\Pi_{2n+1}(\mathrm{SU}(m))$ is finite for $m \leqslant n$. A similar lemma can be proved for Sp(N), that $\Pi_{2n+1}(\mathrm{Sp}(m))$ for $m \leqslant \frac{1}{2}n$ is finite, using the fibration $\mathrm{Sp}(m-1) \to \mathrm{Sp}(m) \to S^{4m-1}$. Note that in both SU($m$) and Sp($m$), the dimension of the sphere is odd. However, for SO(m), we cannot use this method, since the fibration is $\mathrm{SO}(m) \to \mathrm{SO}(m+1) \to S^m$ and $\Pi_q(S^m)$ contains an infinite part not only when $q = m$ but also when $q = 2m-1$ with m = even. In fact, some of the unstable homotopy groups of SO(m) contain an infinite part.

APPENDIX C: TABLE OF STABLE HOMOTOPY GROUPS FOR CLASSICAL GROUPS AND CERTAIN COSET SPACES

In this Appendix, we summarize the Bott periodicity theorem[16] as

k (mod 8)	U	O	Sp	U/O	U/Sp	Sp/U	O/U
0	0	2	0	0	0	0	2
1	∞	2	0	∞	∞	0	0
2	0	0	0	2	0	∞	∞
3	∞	∞	∞	2	0	2	0
4	0	0	2	0	0	2	0
5	∞	0	2	∞	∞	0	0
6	0	0	0	0	2	∞	∞
7	∞	∞	∞	0	2	0	2

,

where ∞ (Ref. 2) denotes $\mathbb{Z}$ ($\mathbb{Z}_2$). Note the following relations:

$$\Pi_k(\mathrm{U/Sp}) = \Pi_{k+2}(\mathrm{O}), \quad \Pi_k(\mathrm{U/O}) = \Pi_{k+2}(\mathrm{Sp}),$$
$$\Pi_k(\mathrm{Sp/U}) = \Pi_{k+1}(\mathrm{Sp}), \quad \Pi_k(\mathrm{O/U}) = \Pi_{k+1}(\mathrm{O}).$$

[1] E. Witten, Phys. Lett. B **117**, 324 (1982).
[2] E. Witten, Commun. Math. Phys. **100**, 197 (1985).
[3] J.-M. Bismut and D. S. Freed, Commun. Math. Phys. **106**, 159 (1986); **107**, 103 (1986). See, also, D. S. Freed, *ibid.* **107**, 483 (1986); S. Della Pietra, V. Della Pietra, and L. Alvarez-Gaume, *ibid.* **109**, 691 (1987); **110**, 573 (1987); D. S. Freed and C. Vafa, *ibid.* **110**, 349 (1987).
[4] E. Witten, Nucl. Phys. **223**, 422, 433 (1983); S. Elitzur and V. P. Nair, *ibid.* B **243**, 205 (1984).
[5] Note that for a "simple" topology for space-time, we do not have to worry about gravitational and mixed anomalies, since for a constant curvature space, the curvature two-form takes the form $R^a_b = k\theta^a \wedge \theta^b$, which implies Tr R^n vanishes identically.
[6] R. Holman and T. W. Kephart, Phys. Lett. B **167**, 417 (1986); E. Kiritsis, *ibid.* B **178**, 53 (1986); **181**, 416(E) (1986); H. W. Braden, Univ. of North Carolina preprint IFP-296-UNC, 1987; L. N. Chang and Y. Liang, Commun. Math. Phys. **108**, 139 (1987).
[7] S. Okubo, H. Zhang, Y. Tosa, and R. E. Marshak, Phys. Rev. D **37**, 1655 (1988).
[8] H. Zhang, S. Okubo, and Y. Tosa, Phys. Rev. D **37**, 2946 (1988).
[9] See, for example, S. B. Treiman, R. Jackiw, B. Zumino, and E. Witten, *Current Algebra and Anomalies* (Princeton U.P., Princeton, NJ, 1985).
[10] G. W. Whitehead, *Elements of Homotopy Theory* (Springer, New York, 1978), p. 159.
[11] Reference 10, p. 187 since $H \to G \xrightarrow{p} G/H$ is a fibration (p. 675).
[12] For example, M. Spivak, *A Comprehensive Introduction to Differential Geometry* (Publish or Perish, Berkeley, 1979), Vol. I, p. 376.
[13] Reference 10, p. 125.
[14] J.-P. Serre, Ann. Math. **54**, 425 (1951); **58**, 258 (1953); B. Harris, *ibid.* **74**, 407 (1961).
[15] In the case where $\Pi_{2n+1}(G)$ is of rank >1, the anomaly coefficient is given by (see Ref. 8)

$$A(\tilde{\omega}) = \exp\left[\frac{i}{e(T)} \int \mathcal{S}\gamma^G\left(\sum_j b_j \alpha_j\right)\right],$$

where the summation j goes over all infinite cyclic summands of $\Pi_{2n+1}(G)$ and α_j denote generators. This occurs for $G = \mathrm{SO}(2n+2)$ [$n \equiv +$ or 3 (mod 4) and $n \geqslant 3$], where $\Pi_{2n+1}(G)$ is of rank 2 (Ref. 17).
[16] R. Bott, Ann. Math. **70**, 313 (1959). See also, for example, J. Milnor, *Morse Theory* (Princeton U.P., Princeton, NJ, 1973).
[17] M. A. Kervaire, Ill. J. Math. **4**, 161 (1960).
[18] For example, see Chap. 11 of Ref. 10.
[19] H. Toda, Mem. Coll. Sci. Kyoto Imp. Univ. **32**, 109 (1959).
[20] R. Bott, Michigan Math. J. **5**, 35 (1958).
[21] A. Borel and F. Hirzebruch, Am. J. Math. **81**, 315 (1959).
[22] I. M. James, Proc. London Math. Soc. **8**, 536 (1958).
[23] F. Sigrist, Comments Math. Helv. **43**, 121 (1968); H. Ōshima, Math. Proc. Cambridge Philos. Soc. **92**, 139 (1982); Osaka J. Math. **21**, 765 (1984); K. Knapp, *Habilitationsschrift* (Bonn, 1979).
[24] G. Walker, Quart. J. Math. Oxford **32**, 467 (1981).
[25] A. T. Lundell, University of Colorado preprint, 1987.
[26] R. Bott, Ann. Math. **70**, 313 (1959); B. Harris, *ibid.* **76**, 295 (1962); M. Mimura, C. R. Acad. Sci. Paris **262**, 20 (1966).
[27] M. G. Barratt and M. E. Mahowald, Bull. Am. Math. Soc. **70**, 758 (1964).
[28] A. T. Lundell, Topology **8**, 209 (1969). In Lemmas 4.2 and 4.3 of this paper, SO($4n+1$) should read SO($4(n+1)$).
[29] S. Hu, *Homotopy Theory* (Academic, New York, 1959), p. 318.

Maximal subalgebra associated with a first integral of a system possessing sl(3,$\mathfrak{R}$) algebra

P. G. L. Leach and F. M. Mahomed
The Centre for Nonlinear Studies and Department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, P.O. WITS 2050, South Africa

(Received 26 August 1987; accepted for publication 6 April 1988)

Conventionally the symmetries of a dynamical system are used to determine the first integrals associated with the system. In this paper the reverse procedure is adopted for one-dimensional systems possessing SL(3,$\mathfrak{R}$) symmetry. The equivalence of such systems enables the free particle equation to be used as a vehicle. It is observed that for three particular first integrals, each gives rise to a triplet of generators having isomorphic algebras. It is then shown how the knowledge of a first integral and its associated triplet enables one to obtain the remaining integrals and triplets.

I. INTRODUCTION

In this paper we shall use the Hamiltonian version of the Lie method.[1,2] Suppose a one-parameter symmetry group has the generator

$$Y(q,p,t) = \xi(q,t)\frac{\partial}{\partial t} + \eta(q,t)\frac{\partial}{\partial q} + \zeta(q,p,t)\frac{\partial}{\partial p}. \tag{1.1}$$

As a consequence of its property of transforming solution curves in (q,p,t) space into solution curves, the first extension of $Y(q,p,t)$,

$$Y^{(1)}(q,p,t) = Y(q,p,t) + \eta^{(1)}(q,p,t)\frac{\partial}{\partial \dot q} + \zeta^{(1)}(q,p,t)\frac{\partial}{\partial \dot p}, \tag{1.2}$$

where

$$\eta^{(1)} = \dot\eta - \dot\xi\frac{\partial H}{\partial p}, \quad \zeta^{(1)} = \zeta + \dot\xi\frac{\partial H}{\partial q}, \tag{1.3}$$

preserves the form of Hamilton's equations

$$\dot q - \frac{\partial H}{\partial p} = 0, \quad \dot p + \frac{\partial H}{\partial q} = 0. \tag{1.4}$$

The action of $Y^{(1)}(q,p,t)$ on (1.4) produces the equations

$$\eta^{(1)} - Y\frac{\partial H}{\partial p} = 0, \quad \zeta^{(1)} + Y\frac{\partial H}{\partial q} = 0, \tag{1.5}$$

which may be manipulated to obtain Y (this being possible due to ξ and η being free of the variable p).

Assuming the existence of solutions for Y, these constitute the complete symmetry group for the dynamical system. Once the generators are found the associated first integrals may also be found.

Here we work from a different viewpoint. Given a constant of the motion we find the set of one-parameter symmetry groups with which it is associated. It is known already that in some cases the same constant arises from different generators. As an example, the simple harmonic oscillator with

$$H = \tfrac{1}{2}(p^2 + q^2) \tag{1.6}$$

has three generators,

$$G_6 = q\frac{\partial}{\partial q},$$
$$G_7 = q\sin t\frac{\partial}{\partial t} + q^2\cos t\frac{\partial}{\partial q}, \tag{1.7}$$
$$G_8 = q\cos t\frac{\partial}{\partial t} - q^2\sin t\frac{\partial}{\partial q}$$

(in the usual notation[3,4]), all of which give rise to the same first integral,

$$I(q,p,t) = (q\cos t - p\sin t)/(q\sin t + p\cos t). \tag{1.8}$$

At this stage we should note that the functional form given to a first integral is a matter of choice. Another form for I(1.8) (Ref. 5) is

$$I'(q,p,t) = t - \arctan(q/p). \tag{1.9}$$

What we shall do is show that for a one-dimensional system possessing SL(3,$\mathfrak{R}$) symmetry, all of the generators of the complete symmetry group may be found from three constants only. We look at the relationship between combinations of generators and various different constants of the motion. The common features of the action of the set of generators on the constants mentioned above are noted. Finally we take a new look at the commutation properties of the Lie algebra which may be written in a simplified form. The vehicle for this discussion is the free particle. The results are true for all one-dimensional systems having sl(3,$\mathfrak{R}$) algebra.

II. GENERATOR FROM FIRST INTEGRAL

In the usual discussion of Lie symmetry groups the transformations act on solution curves in (q,t) space, transforming solution curves into solution curves. In the Hamiltonian context, which we adopt here, we have trajectories in (q,p,t) space. The advantage of considering (q,p,t) space is that the motion, being of the form $I(q,p,t) = C$, represents invariant surfaces in this space, and we may consider infinitesimal transformations that transform solutions into solutions on the same invariant surface.

Let $Y(q,p,t)$ be the generator of such an infinitesimal transformation (i.e., $YI = 0$). Referring to Fig. 1, suppose that Y transforms the trajectory (1) into trajectory (2) and, in particular, that P is transformed to P'. Let the vectors τ_1

 0022-2488/88/081807-07$02.50

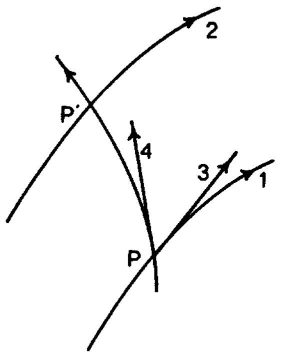


FIG. 1. Curves (1) and (2) are solution curves in (q,p,t) space lying on an invariant surface. At P, (3) is tangent to the solution curve (1) and is represented by τ_1. The transformation path in the invariant surface from P to P' is (4) and is represented by the vector τ_2.

and τ_2 be tangential to the solution curve and the transformation curve at P, respectively. The vectors have the following properties:

$$\tau_1 \times \tau_2 \propto \nabla I (\neq 0), \quad \tau_1 \cdot \nabla I = 0, \quad \tau_2 \cdot \nabla I = 0, \tag{2.1}$$

where ∇ is the gradient operator in (q,p,t) space. The three properties listed in (2.1) are not independent, and we consider the first, which essentially is the rule for permutation of solutions, and the third, which is that the infinitesimal transformation is in the invariant surface.

Writing

$$\tau_1 = 1\hat{\mathbf{t}} + \frac{\partial H}{\partial p}\hat{\mathbf{q}} - \frac{\partial H}{\partial q}\hat{\mathbf{p}},$$

$$\tau_1 = \xi\hat{\mathbf{t}} + \eta\hat{\mathbf{q}} + \zeta\hat{\mathbf{p}}, \tag{2.2}$$

$$\nabla = \hat{\mathbf{t}}\frac{\partial}{\partial t} + \hat{\mathbf{q}}\frac{\partial}{\partial q} + \hat{\mathbf{p}}\frac{\partial}{\partial p},$$

where $\hat{\mathbf{q}}$, $\hat{\mathbf{p}}$, and $\hat{\mathbf{t}}$ are unit vectors in the direction of q, p, and t increasing, respectively, and letting the proportionality in the first of (2.1) be $\gamma(q,p,t)$, we have

$$\begin{bmatrix} 0 & \partial I/\partial p & \partial I/\partial q & \partial I/\partial t \\ -\partial I/\partial p & 0 & 1 & -\partial H/\partial p \\ -\partial I/\partial q & -1 & 0 & -\partial H/\partial q \\ -\partial I/\partial t & \partial H/\partial p & \partial H/\partial q & 0 \end{bmatrix} \begin{bmatrix} \gamma \\ \zeta \\ \eta \\ \xi \end{bmatrix} = 0. \tag{2.3}$$

The condition that the transformation be canonical, viz.,

$$[\bar{q},\bar{p}]_{PB_{q,p}} = 1, \tag{2.4}$$

reduces in the infinitesimal case to

$$\frac{\partial \eta}{\partial q} + \frac{\partial \zeta}{\partial p} = 0. \tag{2.5}$$

From (2.3) we see that this is equivalent to

$$[\gamma, I]_{PB_{q,p}} + [\xi, H]_{PB_{q,p}} = 0. \tag{2.6}$$

In addition to the first integral satisfying (2.3), it also satisfies (1.5). In particular, taking the first of each set we have

$$\zeta\frac{\partial I}{\partial p} + \eta\frac{\partial I}{\partial q} + \xi\frac{\partial I}{\partial t} = 0, \tag{2.7}$$

$$\eta^{(1)} - \zeta\frac{\partial^2 H}{\partial p^2} - \eta\frac{\partial^2 H}{\partial q\,\partial p} - \xi\frac{\partial^2 H}{\partial t\,\partial p} = 0. \tag{2.8}$$

Eliminating $\zeta(q,p,t)$ between (2.7) and (2.8) and insisting that ξ and η be independent of p, we may determine those ξ and η permitted by the particular invariant I. The expression for ζ is obtained from (2.7) or (2.8) and that for γ from one of (2.3). We observe that the rank of the coefficient matrix in (2.3) is always 2.

III. SOME EXAMPLES

To illustrate the method, we consider a few examples. The examples appended are linear as well as nonlinear.

Example (a): As is customary nowadays, we begin by considering the simple harmonic oscillator that has Hamiltonian (1.6). There exist six standard constants of the motion, which are listed in Table I together with the generators for which they are invariants. The order of generators follows the usage of Lutzky.[3] For simplicity the generators are written in unextended form. Starting from the six linearly independent constants listed in Table I we may work backwards to find the generator(s) of the transformation(s) under which each remains invariant. We illustrate the method with J_3. Substituting into (2.7) and (2.8) we have

$$\xi\cos t + \eta\sin t + (q\cos t - p\sin t)\xi = 0, \tag{3.1}$$

$$\eta^{(1)} - \zeta = 0. \tag{3.2}$$

We eliminate ζ between (3.1) and (3.2) and then impose the additional constraint that η and ξ be independent of p to obtain

TABLE I. Symmetry generators and associated invariants for the one-dimensional harmonic oscillator. The generators are given in unextended form for the purpose of simplicity.

Generator	Invariant
$G_1 = \sin 2t\frac{\partial}{\partial t} + q\cos 2t\frac{\partial}{\partial q}$	$J_1 = \frac{1}{2}(p^2 - q^2)\sin 2t - qp\cos 2t$
$G_2 = \cos 2t\frac{\partial}{\partial t} - q\sin 2t\frac{\partial}{\partial q}$	$J_2 = \frac{1}{2}(p^2 - q^2)\cos 2t + qp\sin 2t$
$G_3 = \cos t\frac{\partial}{\partial q}$	$J_3 = q\sin t + p\cos t$
$G_4 = \sin t\frac{\partial}{\partial q}$	$J_4 = q\cos t - p\sin t$
$G_5 = \frac{\partial}{\partial t}$	$J_5 = \frac{1}{2}(p^2 + q^2)$
$G_6 = q\frac{\partial}{\partial q}$	$J_6 = (q\cos t - p\sin t)/(q\sin t + p\cos t)$
$G_7 = q\sin t\frac{\partial}{\partial t} + q^2\cos t\frac{\partial}{\partial q}$	$J_7 = (q\cos t - p\sin t)/(q\sin t + p\cos t)$
$G_8 = q\cos t\frac{\partial}{\partial t} - q^2\sin t\frac{\partial}{\partial q}$	$J_8 = (q\cos t - p\sin t)/(q\sin t + p\cos t)$

$$\frac{\partial\eta}{\partial t}+p\frac{\partial\eta}{\partial q}-p\left(\frac{\partial\xi}{\partial t}+p\frac{\partial\xi}{\partial q}\right)$$
$$+\eta\tan t+(q-p\tan t)\xi=0. \tag{3.3}$$

Equating coefficients of independent powers of p to zero we see that

$$\xi(q,t)=a(t),\quad \eta(q,t)=(\dot a+a\tan t)q+b(t), \tag{3.4}$$

where $a(t)$ and $b(t)$ are solutions of

$$d^2(a/\cos t)/dt^2+a/\cos t=0,\quad \dot b+b\tan t=0. \tag{3.5}$$

Thus we have

$$a(t)=A\cos^2 t+C\cos t\sin t,\quad b(t)=B\cos t, \tag{3.6}$$

from which it follows that

$$Y(q,p,t)$$
$$=(A\cos^2 t+C\sin t\cos t)\frac{\partial}{\partial t}$$
$$+\{(-A\sin t\cos t+C\cos^2 t)q+B\cos t\}\frac{\partial}{\partial q}$$
$$+\{(A\sin t\cos t+C\sin^2 t)p-[A(\cos^2 t-\sin^2 t)$$
$$+2C\sin t\cos t]q-B\cos t\}\frac{\partial}{\partial p}. \tag{3.7}$$

Similar calculations may be performed for the other constants of the motion. It eventuates that, for J_1, J_2, and J_5, only one generator is obtained in each case and it is the one listed in Table I. For J_4 a three-parameter solution is again obtained and, not surprisingly, J_6, J_7, and J_8 give rise to a common three-parameter solution consisting of a linear combination of $G_6^{(1)}(q,p,t)$, $G_7^{(1)}(q,p,t)$, and $G_8^{(1)}(q,p,t)$.

We emphasize that the first extensions are written in terms of the momentum p and not the velocity $\dot q$. It makes no difference to the functional form of the operator in this case, but does, say, for the damped free particle where $p=p(\dot q,t)$. The three triplets of generators are of interest to us and we list them in unextended form in Table II. To mark the departure from the usual expressions used we now employ the symbol $X(q,t)$ to refer to an operator in (q,t) space (i.e., the "unextended" operator). The corresponding operator in (q,p,t) space (i.e., the "extended" operator) is written as $Y(q,p,t)$. In Table II the generators are listed against their invariants (now relabeled I_1,I_2,I_3) and the corresponding γ's are also given.

We defer discussion of the algebraic properties of the generators listed in Table II to the next section. In this section we compare the usual forms for the generators and first integrals given in Table I with those adopted here, show how higher-order integrals are constructed from linear combinations of the generators, and observe the pattern of the action of all the generators on all the integrals in the Table II format. In terms of the X's, the G's, which are usually used to determine the commutator relations that identify the algebra of the SL$(3,\Re)$ group, are given by

$$G_1=X_{22}-X_{33},\quad G_2=-X_{23}-X_{32},\quad G_3=X_{21},$$
$$G_4=X_{31},\quad G_5=-X_{23}+X_{32}, \tag{3.8}$$
$$G_6=-X_{11}=X_{22}+X_{33},\quad G_7=-X_{12},\quad G_8=X_{13}.$$

The linear dependence of the X's is seen in the expressions for G_6, which are equivalent to the linear relation

$$X_{11}+X_{22}+X_{33}=0. \tag{3.9}$$

We further observe that the X's are related among each other as follows:

$$qX_{i1}=\cos tX_{i2}+\sin tX_{i3},\quad i=1,3, \tag{3.10}$$
$$X_{1i}+q\cos tX_{2i}+q\sin tX_{3i}=0,\quad i=1,3. \tag{3.11}$$

TABLE II. The three invariants with associated generators and constants of proportionality. Again, X's instead of Y's are listed for reasons of clarity.

Invariant	Generator	Constant of proportionality
$I_1=\dfrac{q\cos t-p\sin t}{q\sin t+p\cos t}$	$X_{11}=-q\dfrac{\partial}{\partial q}$	$\gamma_{11}=(q\sin t+p\cos t)^2$
	$X_{12}=-q\sin t\dfrac{\partial}{\partial t}-q^2\cos t\dfrac{\partial}{\partial q}$	$\gamma_{12}=(q\sin t+p\cos t)^2 \times(q\cos t-p\sin t)$
	$X_{13}=q\cos t\dfrac{\partial}{\partial t}-q^2\sin t\dfrac{\partial}{\partial q}$	$\gamma_{13}=(q\sin t+p\cos t)^3$
$I_2=q\sin t+p\cos t$	$X_{21}=\cos t\dfrac{\partial}{\partial q}$	$\gamma_{21}=1$
	$X_{22}=\sin t\cos t\dfrac{\partial}{\partial t}+q\cos^2 t\dfrac{\partial}{\partial q}$	$\gamma_{22}=q\cos t-p\sin t$
	$X_{23}=-\cos^2 t\dfrac{\partial}{\partial t}+q\sin t\cos t\dfrac{\partial}{\partial q}$	$\gamma_{23}=q\sin t+p\cos t$
$I_3=q\cos t-p\sin t$	$X_{31}=\sin t\dfrac{\partial}{\partial q}$	$\gamma_{31}=-1$
	$X_{32}=\sin^2 t\dfrac{\partial}{\partial t}+q\sin t\cos t\dfrac{\partial}{\partial q}$	$\gamma_{32}=-q\cos t+p\sin t$
	$X_{33}=-\sin t\cos t\dfrac{\partial}{\partial t}+q\sin^2 t\dfrac{\partial}{\partial q}$	$\gamma_{33}=-q\sin t+p\cos t$

The γ's, which are also listed in Table II, are constants, which perhaps is not unexpected. We note that in each case,

$$\gamma_{i2} = I_3\gamma_{i1}, \quad \gamma_{i3} = I_2\gamma_{i1}, \quad i = 1,3. \tag{3.12}$$

By inspection, the relations between the constants listed in Tables I and II are

$$J_1 = -\tfrac{1}{2}I_2I_3, \quad J_2 = \tfrac{1}{2}(I_3^2 - I_2^2), \quad J_3 = I_2, \quad J_4 = I_3, \quad J_5 = \tfrac{1}{2}(I_2^2 + I_3^2), \quad J_6 = J_1(=J_7 = J_8). \tag{3.13}$$

The members of each of the three classes of first integrals—linear, quadratic, and quotient—constitute a complete set of each class. In the case of the quotient integral, I_1, it will be appreciated that the theory involved in determining the integral does not distinguish between I_1 and its reciprocal.

Considering the relations between the I's and J's and the X's and G's, it is instructive to calculate the integral that results from a linear combination of the X's. As an example of the calculation we take

$$\widetilde{X}_1 = mX_{21} + nX_{31}. \tag{3.14}$$

The associated integral is a solution of

$$\widetilde{Y}_1\widetilde{I}_1 = 0, \quad \frac{d(\widetilde{I}_1)}{dt} = 0. \tag{3.15}$$

The first of (3.15) is

$$\frac{dt}{0} = \frac{dq}{m\cos t + n\sin t} = \frac{dp}{-m\sin t + n\cos t}, \tag{3.16}$$

from which is follows that $\widetilde{I}_1$ is a function of u_1 and u_2, where

$$u_1 = t, \quad u_2 = m(q\sin t + p\cos t) - n(q\cos t - p\sin t). \tag{3.17}$$

As a matter of terminology we point out that here we use u_1 and u_2 to emphasize the traditional equivalence of the canonical coordinates in Hamiltonian mechanics. Were we using q, $\dot{q}$, and t as variables, the usual u and v where v represents the *differential* invariant would be appropriate. Using the second of (3.15) in the standard fashion,

$$\frac{du_2}{du_1} = \frac{du_2}{dt}\left(\frac{du_1}{dt}\right)^{-1} = 0. \tag{3.18}$$

Hence the invariant is

$$\widetilde{I}_1 = mI_2 - nI_3, \tag{3.19}$$

i.e., simply a linear combination.

Combining X_{23} and X_{32} in the same manner we obtain

$$\widetilde{I}_2 = (m+n)J_2 + (m-n)J_5, \tag{3.20}$$

i.e., a linear combination of two of the quadratic integrals. The choices $m = -1 = n$ and $m = -1$, $n = 1$ yield J_2 and J_5, respectively (up to a constant multiplicative factor there is an arbitrariness in how an integral is written). These two choices coincide with the combinations given in (3.8). Similarly the combination of X_{22} and X_{33} gives

$$\widetilde{I}_3 = (p\sin t - q\cos)^m(p\cos t + q\sin t)^{-n}, \tag{3.21}$$

yielding J_1 and J_6 for the choices $m = 1$, $n = -1$ and $m = 1$, $n = 1$, respectively.

We note that in the three examples of integrals considered, the first gave simply a linear combination of the two linear integrals, the second a linear combination of the quadratic integrals, whereas the third provided an integral of arbitrary order depending only upon our choice of m and n. We do not wish to labor the matter of combinations of generators, but we point out that the relationship between the combination of generators and the combination of the associated integrals is not simple. The reason for the diversity of the results in (3.19), (3.20), and (3.21) possibly lies in the commutation relations between generators. The effect of the commutation relations would become obvious if exponentiation were used due to the Campbell–Baker–Hausdorff formula.

To conclude this example we observe the action of each of the Y's on the three integrals I_1, I_2, and I_3. These are given in Table III. The table may be summarized as follows. If

$$Y_{i1}I_j = A_{ij}, \tag{3.22}$$

where A_{ij} is some combination of I_1, I_2, and I_3, then

$$Y_{i2}I_j = I_3A_{ij}, \quad Y_{i3}I_j = I_2A_{ij}. \tag{3.23}$$

The connection with (3.12) is obvious.

We have treated the simple harmonic oscillator in great detail. For the remaining examples, we merely give a sketch.

Example (b): Let us now consider the damped free particle, which has the Hamiltonian

$$H = \tfrac{1}{2}p^2e^{-kt}, \quad p = \dot{q}e^{kt}. \tag{3.24}$$

The appropriate first integrals are

$$I_1 = q/p - (1 - e^{-kt})/k, \quad I_2 = p, \quad I_3 = q - p(1 - e^{-kt})/k. \tag{3.25}$$

The triplets of generators corresponding to each first integral are

$$X_{11} = -q\frac{\partial}{\partial q}, \quad X_{12} = \frac{q(1 - e^{kt})}{k}\frac{\partial}{\partial t} - q^2\frac{\partial}{\partial q}, \quad X_{13} = qe^{kt}\frac{\partial}{\partial t}; \tag{3.26}$$

$$X_{21} = \frac{\partial}{\partial q}, \quad X_{22} = -\frac{(1 - e^{kt})}{k}\frac{\partial}{\partial t} + q\frac{\partial}{\partial q}, \tag{3.27}$$

TABLE III. The effect of each Y acting on each invariant. The expressions in parentheses illustrate the effect more explicitly. A Y_{2i} operator has the effect of dividing by I_3 and a Y_{3i} operator by I_2, in contrast to the multiplying effect shown by Y_{i2} and Y_{i3}.

	I_1	I_2	I_3
Y_{11}	0	$-I_2$	$-I_3$
Y_{12}	0	$-I_3I_2$	$-I_3^2$
Y_{13}	0	$-I_2^2$	$-I_2I_3$
Y_{21}	$1/I_2$ $(=I_1/I_3)$	0	1 $(=I_3/I_3)$
Y_{22}	I_1 $(=I_3I_1/I_3)$	0	I_3 $(=I_3I_3/I_3)$
Y_{23}	1 $(=I_2I_1/I_3)$	0	I_2 $(=I_2I_3/I_3)$
Y_{31}	$-I_1/I_2$	1 $(=I_2/I_2)$	0
Y_{32}	$-I_1^2$ $(=-I_3I_1/I_2)$	I_3 $(=I_3I_2/I_2)$	0
Y_{33}	$-I_1$ $(=-I_2I_1/I_2)$	I_2 $(=I_2I_2/I_2)$	0

$$X_{23} = -e^{kt}\frac{\partial}{\partial t};$$

$$X_{31} = \frac{(1-e^{-kt})}{k}\frac{\partial}{\partial q},$$

$$X_{32} = \frac{(e^{kt}+e^{-kt}-2)}{k^2}\frac{\partial}{\partial t} + \frac{q(1-e^{-kt})}{k}\frac{\partial}{\partial q}, \qquad (3.28)$$

$$X_{33} = \frac{(1-e^{kt})}{k}\frac{\partial}{\partial t}.$$

The linear dependence relation is again (3.9) and the following relations amongst the generators are observed:

$$qX_{i1} = e^{kt}X_{i2} + [(e^{kt}-1)/k]X_{i3}, \quad i = 1,3,$$
$$X_{1i} + qe^{kt}X_{2i} - qke^{kt}X_{3i} = 0. \qquad (3.29)$$

It may be verified by direct calculation that the first integrals and generators have the properties summarized in Tables III and IV.

We next treat two familiar examples of nonlinear systems.

Example (c): The differential equation

$$\ddot{q} + 3q\dot{q} + q^3 = 0 \qquad (3.30)$$

occurs in the investigation of univalued functions defined by second-order differential equations and in the study of the modified Emden equation.[6,7] Recently it has also been treated in Ref. 8. Equation (3.30) has the Hamiltonian

$$H = t/q - pq^2 - \tfrac{1}{2}(t^2 - 2pq^2)^{1/2},$$
$$p = \frac{1}{2}\frac{t^2}{q^2} - \frac{1}{2}\frac{1}{(q^2+\dot{q})^2}. \qquad (3.31)$$

The first integrals are (cf. Ref. 7)

$$I_1 = -t + (t^2 - 2pq^2)^{1/2},$$

$$I_2 = t^2/2 + (1/q)(1-tq)(t^2-2pq^2)^{1/2}, \qquad (3.32)$$

$$I_3 = \frac{t^2q + 2(1-tq)(t^2-2pq^2)^{1/2}}{-2tq + 2q(t^2-2pq^2)^{1/2}}.$$

The triplets of generators corresponding to each first integral are

$$X_{11} = \frac{1}{2}t^2q\frac{\partial}{\partial t} + \left(tq^2 - \frac{1}{2}t^2q^3 - q\right)\frac{\partial}{\partial q},$$

$$X_{12} = tq\frac{\partial}{\partial t} + (q^2 - tq^3)\frac{\partial}{\partial q}, \qquad (3.33)$$

$$X_{13} = q\frac{\partial}{\partial t} - q^3\frac{\partial}{\partial q};$$

$$X_{21} = \left(\frac{1}{2}t^3q - \frac{1}{2}t^2\right)\frac{\partial}{\partial t} + \left(1 - 2tq - \frac{1}{2}t^3q^3 + \frac{3}{2}t^2q^2\right)\frac{\partial}{\partial q},$$

$$X_{22} = (t^2q - t)\frac{\partial}{\partial t} + (2tq^2 - t^2q^3 - q)\frac{\partial}{\partial q}, \qquad (3.34)$$

$$X_{23} = (tq-1)\frac{\partial}{\partial t} + (q^2 - tq^3)\frac{\partial}{\partial q};$$

$$X_{31} = \left(\frac{1}{2}t^3 - \frac{1}{4}qt^4\right)\frac{\partial}{\partial t} + \left(-t + \frac{3}{2}t^2q + \frac{1}{4}t^4q^3 - t^3q^2\right)\frac{\partial}{\partial q},$$

$$X_{32} = \left(-\frac{1}{2}t^3q + t^2\right)\frac{\partial}{\partial t} + \left(tq + \frac{1}{2}t^3q^3 - \frac{3}{2}t^2q^2\right)\frac{\partial}{\partial q}, \qquad (3.35)$$

$$X_{33} = \left(-\frac{1}{2}t^2q + t\right)\frac{\partial}{\partial t} + \left(\frac{1}{2}t^2q^3 - tq^2\right)\frac{\partial}{\partial q}.$$

The linear dependence relation is given by

$$X_{11} = X_{22} + X_{33}, \qquad (3.36)$$

and the relations among the operators are

$$(t - 1/q)X_{1i} = X_{2i}, \quad i = 1,3,$$
$$(t/q) - t^2/2)X_{1i} = X_{3i}. \qquad (3.37)$$

Example (d): The equation

$$t\ddot{q} = \dot{q}^3 + \dot{q} \qquad (3.38)$$

was considered in Ref. 9. It has the Hamiltonian

$$H = -4t(q^3p/2)^{1/2} - 4tq, \quad p = 2q^2 - 2t^2/\dot{q}^2, \qquad (3.39)$$

and first integrals

$$I_1 = 2(q^2 - p/2)^{1/2} + 2q,$$

$$I_2 = t^2 - q^2 - 2q(q^2 - p/2)^{1/2}, \qquad (3.40)$$

$$I_3 = \frac{t^2 - q^2 - 2q(q^2-p/2)^{1/2}}{2q + 2(q^2 - p/2)^{1/2}}.$$

The triplets of generators associated with each first integral are

TABLE IV. Commutation relations between the X's. The entry is [*X column, X row*]. The relations for the Y's are, naturally, the same.

	X_{11}	X_{12}	X_{13}	X_{21}	X_{22}	X_{23}	X_{31}	X_{32}	X_{33}
X_{11}	0	$-X_{12}$	$-X_{13}$	X_{21}	0	0	X_{31}	0	0
X_{12}	X_{12}	0	0	$X_{22}-X_{11}$	$-X_{12}$	$-X_{13}$	X_{32}	0	0
X_{13}	X_{13}	0	0	X_{23}	0	0	$X_{33}-X_{11}$	$-X_{12}$	$-X_{13}$
X_{21}	$-X_{21}$	$X_{11}-X_{22}$	$-X_{23}$	0	X_{21}	0	0	X_{31}	0
X_{22}	0	X_{12}	0	$-X_{21}$	0	$-X_{23}$	0	X_{32}	0
X_{23}	0	X_{13}	0	0	X_{23}	0	$-X_{21}$	$X_{33}-X_{22}$	$-X_{23}$
X_{31}	$-X_{31}$	$-X_{32}$	$X_{11}-X_{33}$	0	0	X_{21}	0	0	X_{31}
X_{32}	0	0	X_{12}	$-X_{31}$	$-X_{32}$	$X_{22}-X_{33}$	0	0	X_{32}
X_{33}	0	0	X_{13}	0	0	X_{23}	$-X_{31}$	$-X_{32}$	0

$$X_{11}=\left(\frac{t}{2}-\frac{1}{2}t^{-1}q^2\right)\frac{\partial}{\partial t}+q\frac{\partial}{\partial q},$$
$$X_{12}=-t^{-1}q\frac{\partial}{\partial t}+\frac{\partial}{\partial q}, \tag{3.41}$$
$$X_{13}=\frac{1}{2}t^{-1}\frac{\partial}{\partial t};$$
$$X_{21}=\left(\frac{1}{2}tq-\frac{1}{2}t^{-1}q^3\right)\frac{\partial}{\partial t}+q^2\frac{\partial}{\partial q},$$
$$X_{22}=-t^{-1}q^2\frac{\partial}{\partial t}+q\frac{\partial}{\partial q}, \tag{3.42}$$
$$X_{23}=\frac{1}{2}t^{-1}q\frac{\partial}{\partial t};$$
$$X_{31}=\left(\frac{t^3}{2}-\frac{1}{2}t^{-1}q^4\right)\frac{\partial}{\partial t}+(t^2q+q^3)\frac{\partial}{\partial q},$$
$$X_{32}=-(tq+t^{-1}q^3)\frac{\partial}{\partial t}+(t^2+q^2)\frac{\partial}{\partial q}, \tag{3.43}$$
$$X_{33}=\left(\frac{1}{2}t+\frac{1}{2}t^{-1}q^2\right)\frac{\partial}{\partial t}.$$

The linear dependence relation is again (3.36) and the relations among the generators are

$$qX_{1i}=X_{2i},\quad i=1,3,$$
$$(t^2+q^2)X_{1i}=X_{3i}. \tag{3.44}$$

IV. COMMUTATION RELATIONS

Each of the triplets $\{X_{1i}\}$, $\{X_{2i}\}$, and $\{X_{3i}\}$ constitutes a Lie subalgebra. The commutation relations are

$$[X_{11},X_{12}]=-X_{12},\quad [X_{11},X_{13}]=-X_{13},$$
$$[X_{12},X_{13}]=0, \tag{4.1}$$
$$[X_{21},X_{22}]=\pm X_{21},\quad [X_{21},X_{23}]=0,$$
$$[X_{22},X_{23}]=\mp X_{23}, \tag{4.2}$$
$$[X_{31},X_{32}]=0,\quad [X_{31},X_{33}]=\pm X_{31},$$
$$[X_{32},X_{33}]=\pm X_{32}. \tag{4.3}$$

Clearly each of the three sets of commutation relations given above can be written in the form

$$[Z_1,Z_2]=0,\quad [Z_1,Z_3]=Z_1,\quad [Z_2,Z_3]=Z_2, \tag{4.4}$$

and so the algebraic properties of the triplets are identical.

The Lie algebra represented by the commutation relations (4.4) is denoted by $A_{3,3}$ (see Refs. 10 and 11). The question now arises as to whether the three triplets of generators, having isomorphic algebras (4.1), (4.2), and (4.3), associated with each of the examples treated in the previous section, can be transformed into a canonical triplet of generators by a point transformation. To answer this we prove the following result.

Proposition: In order that a differential equation $\ddot{q}=N(\dot{q},q,t)$ possess $sl(3,\mathfrak{R})$ algebra it is necessary and sufficient that it have the algebra $A_{3,3}$.

Proof: To prove sufficiency, suppose that $\ddot{q}=N(\dot{q},q,t)$ has the algebra $A_{3,3}$. The generators of symmetry G_i $(i=1,3)$ then satisfy the $A_{3,3}$ commutation relations. Assume first that $G_2=\rho(t,q)G_1$ and $G_3=\psi(t,q)G_1$ for suitable nonconstant functions ρ and ψ. This certainly is a strong condition. There is, however, no less generality in assuming the conjunction than there is in assuming the disjunctive statement, since $G_2=\rho G_1$ implies $G_3=\psi G_1$ and vice versa. Invoking $A_{3,3}$ we obtain $G_1\rho=0$ and $G_1\psi=1$. Transforming G_1 to $\overline{G}_1=\partial/\partial Q$ and taking the simplest solutions of the aforementioned equations we arrive at the generators

$$\overline{G}_1=\frac{\partial}{\partial Q},\quad \overline{G}_2=T\frac{\partial}{\partial Q},\quad \overline{G}_3=Q\frac{\partial}{\partial Q}. \tag{4.5}$$

Expressing the invariance of the differential equation for Q with respect to (4.5) results in the free particle equation $Q''=0$.

By assuming that $G_2\neq\phi(t,q)G_1$ for any function ϕ, we straightforwardly obtain

$$\overline{G}_1=\frac{\partial}{\partial T},\quad \overline{G}_2=\frac{\partial}{\partial Q},\quad \overline{G}_3=T\frac{\partial}{\partial T}+Q\frac{\partial}{\partial Q}, \tag{4.6}$$

the differential equation once more being the free particle equation. Proof of the case for which the generators G_1 and G_3 are unconnected is omitted as it also yields (4.6).

Thus it follows that the equation for q has the $sl(3,\mathfrak{R})$ algebra. Proof of the necessity is trivial since $A_{3,3}$ is a subalgebra of $sl(3,\mathfrak{R})$. ◊

We have two realizations for the algebra $A_{3,3}$, viz., (4.5) and (4.6). On examining each of the triplets of generators obtained in the examples we notice that they are unconnected. For example, $X_{12}\neq\rho X_{11}$, $X_{13}\neq\psi X_{11}$, and $X_{12}\neq\phi X_{13}$ for any functions ρ, ψ, and ϕ. Therefore each of the triplets of the examples can be reduced to the standard form (4.6) by means of a point transformation that transforms the original equation to the free particle equation $Q''=0$. It is thus instructive to look at the free particle equation. The Hamiltonian is

$$H=\tfrac{1}{2}P^2,\quad P=Q',$$

and the first integrals are

$$I_1=P,\quad I_2=Q-TP,\quad I_3=(Q-TP)/P. \tag{4.7}$$

Here I_1 has associated the triplet [cf. (4.6)]

$$\overline{X}_{11}=T\frac{\partial}{\partial T}+Q\frac{\partial}{\partial Q},\quad \overline{X}_{12}=\frac{\partial}{\partial T},\quad \overline{X}_{13}=\frac{\partial}{\partial Q}.$$

For I_2, we have

$$\overline{X}_{12}=T^2\frac{\partial}{\partial T}+TQ\frac{\partial}{\partial Q},\quad \overline{X}_{22}=T\frac{\partial}{\partial T},\quad \overline{X}_{23}=T\frac{\partial}{\partial Q},$$

and for I_3,

$$\overline{X}_{31}=TQ\frac{\partial}{\partial T}+Q^2\frac{\partial}{\partial Q},\quad \overline{X}_{32}=Q\frac{\partial}{\partial T},\quad \overline{X}_{33}=Q\frac{\partial}{\partial Q},$$

We immediately observe these relations among the generators [cf. Examples (c) and (d)]:

$$T\overline{X}_{1i}=\overline{X}_{2i},\quad i=1,3,$$
$$Q\overline{X}_{1i}=\overline{X}_{3i}. \tag{4.8}$$

By inspection we observe that the I_2 generators are equivalent to the I_3 generators under the interchange transformation

$$\overline{T}=Q,\quad \overline{Q}=T. \tag{4.9}$$

The I_1 generators are form-invariant under (4.9). It is easily verified that the I_2 generators transform into the I_1 generators via the transformation

$$\bar{T} = -1/T, \quad \bar{Q} = Q/T. \tag{4.10}$$

Hence using (4.9) together with (4.10) we deduce that

$$\bar{T} = -1/Q, \quad \bar{Q} = T/Q, \tag{4.11}$$

reduces the I_3 generators to the I_1 generators. Clearly the above transformations leave the free particle equation invariant.

Hence the knowledge of a first integral J together with its associated triplet $\{G_i: i = 1,3\}$ enables an equation to be reduced to the free particle equation by a point transformation that transforms the G_i's to the standard form (4.6). Accordingly one can obtain the remaining first integrals from (4.7) and the associated triplets from (4.8). The foregoing examples can easily be shown to illustrate this. However, we consider the time-dependent oscillator as a further example. The equation is

$$\ddot{q} + \omega^2(t)q = 0, \tag{4.12}$$

with the Hamiltonian

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2 q^2, \quad p = \dot{q}.$$

A first integral is given by

$$I = 1/(q\dot{C} - pC),$$

where C is a particular solution of (4.12). The triplet of generators associated with I are obtained by using (2.7) and (2.8). They are

$$\begin{aligned} X_{11} &= C^2 \int C^{-2}\,dt \frac{\partial}{\partial t} + q\left(1 + C\dot{C}\int C^{-2}\,dt\right)\frac{\partial}{\partial q}, \\ X_{12} &= C^2 \frac{\partial}{\partial t} + C\dot{C}q\frac{\partial}{\partial q}, \\ X_{13} &= C\frac{\partial}{\partial q}. \end{aligned} \tag{4.13}$$

The transformation which reduces (4.13) to (4.6) is given by

$$T = \int C^{-2}\,dt, \quad Q = q/C.$$

It is now a simple matter to obtain the remaining first integrals and triplets by using (4.7) and (4.8) together with the above transformation. The first integrals are

$$\begin{aligned} J &= \frac{q}{C} - (pC - q\dot{C})\int C^{-2}\,dt, \\ K &= \frac{q}{C(pC - q\dot{C})} - \int C^{-2}\,dt, \end{aligned} \tag{4.14}$$

and the triplets are given by

$$\begin{aligned} &\int C^{-2}\,dt\, X_{1i} = X_{2i}, \quad i = 1,3, \\ &(q/C)X_{1i} = X_{3i}, \quad i = 1,3. \end{aligned} \tag{4.15}$$

ACKNOWLEDGMENT

F.M.M. thanks the Council of Scientific and Industrial Research of South Africa and the University of the Witwatersrand for financial support.

[1]G. E. Prince and P. G. L. Leach, Hadronic J. **3**, 941 (1980).
[2]P. G. L. Leach, J. Austral. Math. Soc. Series B **23**, 173 (1981).
[3]M. Lutzky, J. Phys. A **11**, 249 (1978).
[4]P. G. L. Leach, J. Math. Phys. **21**, 300 (1980).
[5]W. Sarlet, J. Phys. A **14**, 2227 (1981).
[6]P. G. L. Leach, J. Math. Phys. **26**, 2510 (1985).
[7]F. M. Mahomed and P. G. L. Leach, Quaestiones Math. **8**, 241 (1985).
[8]L. G. S. Duarte, S. E. S. Duarte, and I. C. Moreira, J. Phys. A **20**, L701 (1987).
[9]F. M. Mahomed and P. G. L. Leach, University of the Witwatersrand preprint, Department of Computational and Applied Mathematics.
[10]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[11]J. Patera and P. Winternitz, J. Math. Phys. **18**, 1449 (1977).

Detailed behavior of the phase-integral approximations at zeros and singularities of the square of the base function

Andrzej A. Skorupski
Soltan Institute for Nuclear Studies, Hoża 69, 00-681 Warsaw, Poland

(Received 9 June 1987; accepted for publication 30 March 1988)

Approximate solutions to the one-dimensional time independent wave equation, called the phase-integral approximations, are analyzed in the vicinity of characteristic points. The approximations are of arbitrary order and are generated from an unspecified base function. The general theory is illustrated by examples involving the power and/or the exponential behavior of the square of the base function. In these cases simple estimates are derived for the integrals which define the accuracy of the phase-integral approximation, and the optimum approximation order is determined.

I. INTRODUCTION

This paper deals with the phase-integral approximations, which are approximate solutions to the one-dimensional time independent wave equation

$$\frac{d^2\psi}{dz^2}+R(z)\psi=0,\quad z=x+iy. \tag{1.1}$$

Their early version[1] was related to the ordinary WKB approximation. They were then generalized so as to include an unspecified base function.[2,3] This more general version, which is the subject of our analysis, can be written

$$\psi(z)\simeq f_{1,2}(z)\equiv q^{-1/2}(z)\exp\left[\pm i\int q(z)dz\right], \tag{1.2}$$

where, for the approximation of order $2N+1$, $N\geqslant 0$,

$$q(z)=Q(z)\sum_{n=0}^{N}Y_{2n}(z),\quad Y_0(z)\equiv 1, \tag{1.3}$$

where $Q(z)$ is the base function, and the $Y_{2n}(z)$, $n\geqslant 1$, are the higher-order corrections; they can be expressed uniquely in terms of $R(z)$, $Q^2(z)$, and the derivatives d/dz of these functions.[4] The central role in the theory is played by two dimensionless quantities,

$$\zeta=\int Q(z)dz, \tag{1.4}$$

and

$$\varepsilon_0(z)=Q^{-2}(z)\left[R(z)-Q^2(z)+Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right], \tag{1.5a}$$

where the differential term in (1.5a) can be expressed uniquely in terms of $Q^2(z)$:

$$Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\equiv\frac{5}{16}\left(Q^{-2}\frac{dQ^2}{dz}\right)^2-\frac{1}{4}Q^{-2}\frac{d^2Q^2}{dz^2}. \tag{1.5b}$$

Thus all corrections Y_{2n} are polynomials in ε_0 and the derivatives $d/d\zeta$ of ε_0, e.g.,

$$Y_2=\tfrac{1}{2}\varepsilon_0,\quad Y_4=-\tfrac{1}{8}(\varepsilon_0^2+\varepsilon_2),$$
$$Y_6=\tfrac{1}{32}(2\varepsilon_0^3+6\varepsilon_0\varepsilon_2+5\varepsilon_1^2+\varepsilon_4),\ldots, \tag{1.6a}$$

where

$$\varepsilon_k=\frac{d^k\varepsilon_0}{d\zeta^k},\quad k=0,1,2,\ldots. \tag{1.6b}$$

In principle all corrections Y_{2n} can be determined successively from the recurrence relations derived in Refs. 1–4, but one soon encounters significant algebraic complications.[4] For $4<n\leqslant 10$ the problem was handled by using a computer to perform the algebra.[5] For our purposes the following general result will be useful[4]:

$$Y_{2n}=\frac{(-1)^{n+1}}{2^{2n-1}}\left[\frac{(2n)!\varepsilon_0^n}{2(2n-1)(n!)^2}+\cdots+\varepsilon_{2n-2}\right], \tag{1.6c}$$

where the first term in brackets contains no derivatives of ε_0, while the last term contains the highest derivative.

An important point is that the phase-integral approximations belong to the class of approximations (1.2) in any order, and for any choice of the base function. This often simplifies generalization of the first-order results to higher orders in contrast to the ordinary WKB approximation.[6,7] Also the general theory of the approximations (1.2) developed in Ref. 8 is applicable. The special role of the functions $f_1(z)$ and $f_2(z)$ defined by (1.2) in representing approximate solutions of Eq. (1.1) follows from the fact that these functions are exact solutions of Eq. (1.1), for any $q(z)$, if $R(z)$ is defined in terms of $q(z)$ as

$$R(z)=q^2(z)-q^{1/2}\frac{d^2q^{-1/2}}{dz^2}. \tag{1.7}$$

Therefore the $f_{1,2}(z)$ exhibit these general features of the exact solutions of Eq. (1.1) which are independent of the details of $R(z)$, e.g., constancy of the Wronskian [for any $q(z)$], or the quantum mechanical current conservation [if $R(x)$ defined by Eq. (1.7) is real and nonsingular, i.e., if $q^2(x)$ is real, nonsingular, and nonzero, see Eq. (1.5b)].

The base function $Q(z)$ is often chosen so that either

$$Q^2(z)=R(z), \tag{1.8}$$

or $R(z)-Q^2(z)$ is small in comparison to the differential term in Eq. (1.5a) as z tends to a zero or singularity of $Q^2(z)$. Such situations are discussed in Sec. IV. Another possibility is to choose $Q(z)$ so as to optimize the approximation at a given characteristic point (Sec. V). In any case our aim is to derive the leading terms of the $Y_{2n}(z)$ as z tends to a zero or a

 0022-2488/88/081814-10$02.50

singularity of $Q^2(z)$. General aspects of these derivations are described in Sec. III. The leading terms of the $Y_{2n}(z)$ are then used to derive simple estimates for the integrals defining the accuracy of the phase-integral approximation (Sec. VI). The dependence of these integrals on the approximation order is discussed in Sec. VII. Section II gives the transformation properties of the $Y_{2n}(z)$, which are used in the following derivations. They also clarify some earlier results.

II. TRANSFORMATION PROPERTIES OF THE $Y_{2n}(z)$

The wave equation (1.1) can be mapped into another differential equation of the same form by the well-known[8] simultaneous transformation of the independent variable, $z \to \tilde{z}$, and the unknown function, $\psi \to \tilde{\psi}$:

$$\frac{d^2\tilde{\psi}}{d\tilde{z}^2} + \tilde{R}(\tilde{z})\tilde{\psi} = 0\,, \tag{2.1a}$$

where

$$\tilde{\psi} = \left(\frac{d\tilde{z}}{dz}\right)^{1/2}\psi\,, \tag{2.1b}$$

$$\tilde{R} = \left(\frac{d\tilde{z}}{dz}\right)^{-2}\left[R + \left(\frac{d\tilde{z}}{dz}\right)^{1/2}\frac{d^2(d\tilde{z}/dz)^{-1/2}}{dz^2}\right]. \tag{2.1c}$$

Specializing Eqs. (1.2)–(1.6b) to the $\tilde{z}$ plane, and choosing some base function $\tilde{Q}(\tilde{z}) \equiv d\tilde{\zeta}/d\tilde{z}$, Eq. (2.1a) can then be solved with the phase-integral approximation. To see how this approximation is related to the original phase-integral approximation pertaining to Eq. (1.1) we introduce three auxiliary transformations:

$$z \to \zeta(z), \quad \tilde{z} \to \tilde{\zeta}(\tilde{z}), \quad \zeta \to \tilde{\zeta}(\zeta)\,. \tag{2.2}$$

It can be seen by using Eqs. (2.1c) and (1.5a), appropriately specialized, that the R functions after the first two transformations (2.2) are $1 + \varepsilon_0$, and $1 + \tilde{\varepsilon}_0$, respectively. Hence Eq. (2.1c) specialized to the transformation $\zeta \to \tilde{\zeta}(\zeta)$ gives

$$1 + \tilde{\varepsilon}_0 = \left(\frac{d\tilde{\zeta}}{d\zeta}\right)^{-2}\left[1 + \varepsilon_0 + \left(\frac{d\tilde{\zeta}}{d\zeta}\right)^{1/2}\frac{d^2(d\tilde{\zeta}/d\zeta)^{-1/2}}{d\zeta^2}\right]. \tag{2.3}$$

If the base function $\tilde{Q}(\tilde{z})$ is chosen so as to keep invariant the ζ variable (1.4), $d\tilde{\zeta}/d\zeta = 1$, Eq. (2.3) indicates that ε_0 will also be invariant, $\tilde{\varepsilon}_0(\tilde{z}) \equiv \varepsilon_0(z)$. This in turn implies invariance of all corrections Y_{2n}, in view of Eqs. (1.6a), (1.6b),

$$\tilde{Y}_{2n}(\tilde{z}) \equiv Y_{2n}(z), \quad n = 1,2,\ldots\,. \tag{2.4}$$

In that case, on transforming the phase-integral approximation pertaining to Eq. (2.1a) back to the z plane, according to Eq. (2.1b), one arrives at the original approximation (1.2)–(1.6b). In the derivation of Eqs. (1.2)–(1.6b) given in Ref. 3 the above mentioned property of the phase-integral approximation was postulated, along with the invariance of the ζ variable [see Eqs. (14) and (14′) in Ref. 3], leading in conclusion to the invariance (2.4). Our arguments based on Eq. (2.3) relate invariance of the Y_{2n} to invariance of ζ, and indicate that these features are just an option. The usefulness of other options will be demonstrated later in this section. Recalling the fact that the $Y_{2n}(z)$ can be expressed in terms of $Q^2(z)$ rather than $Q(z)$ (Sec. I), the invariance (2.4) will be valid also if we change the sign of $\tilde{Q}(\tilde{z})$, i.e., the sign of $\tilde{\zeta}$. In other words $\tilde{Q}^2(\tilde{z})$ can be chosen so that $(d\tilde{\zeta}/d\zeta)^2 = 1$, leading to

$$\tilde{Q}^2 = \left(\frac{d\tilde{z}}{dz}\right)^{-2}Q^2\,. \tag{2.5}$$

Equation (2.5) is the transformation rule for $Q^2(z)$ which keeps invariant the corrections $Y_{2n}(z)$.

Transformations for which the differential term in brackets in Eq. (2.1c) is identically zero play a special role, as in that case the transformation rules for $R(z)$ [Eq. (2.1c)] and for $Q^2(z)$ [Eq. (2.5)] are the same. The most general form of such transformations is

$$\tilde{z} = (az + b)/(cz + d), \quad ad - bc \neq 0\,, \tag{2.6}$$

but we can exclude the trivial cases of $c = 0$. With an appropriate choice of the origin in the z plane, and the origin, the unit length, and the real axis in the $\tilde{z}$ plane, Eq. (2.6) can be written ($c \neq 0$)

$$\tilde{z} = L^2/z, \quad L > 0\,, \tag{2.7}$$

where L is a characteristic length. Equations (2.5) and (2.7) lead to

$$\tilde{z}^2\tilde{Q}^2(\tilde{z}) \equiv z^2Q^2(z), \quad z\tilde{z} = L^2\,. \tag{2.8}$$

The same transformation rule is obtained for $R(z)$, but also for the differential term in Eq. (1.5a) due to the invariance of ε_0. Therefore the ratio of $R(z) - Q^2(z)$ to the differential term in Eq. (1.5a) is also invariant. This never happens with other transformations if Eqs. (2.5) and (2.4) are fulfilled.

In Ref. 9 both Eqs. (1.1) and (2.1a) were solved with the phase-integral approximation satisfying (1.8), and the transformation $\tilde{z}(z)$ was not of the type (2.7). Therefore Eq. (2.5) was violated, and the invariance of the Y_{2n} was broken. Nevertheless the functions $Y_{2n}(z)$ and $\tilde{Y}_{2n}(\tilde{z})$ [not satisfying (2.4)] turned out useful. They were used to construct the double phase-integral approximations,[9] the efficiency of which was demonstrated in Ref. 10.

Equations (2.8) and (2.4) will be used in the following sections to simplify the derivations as follows. If $Q^2(z)$ has a zero or a singularity at $z = 0$, Eq. (2.8) indicates that $\tilde{Q}^2(\tilde{z})$ also has some zero or singularity at $\tilde{z} = \infty$, and *vice versa.* Using Eq. (2.4) one can relate the behavior of the $Y_{2n}(z)$ at $z = 0$, a zero or singularity of $Q^2(z)$, to the behavior of the $\tilde{Y}_{2n}(\tilde{z})$ at $\tilde{z} = \infty$, the corresponding zero or singularity of $\tilde{Q}^2(\tilde{z})$.

III. LEADING TERMS OF THE $Y_{2n}(z)$ FROM SIMPLE MODELS

The analysis given in this and the following two sections aims at the leading terms of the $Y_{2n}(z)$ in the vicinity of characteristic points. To define the $Y_{2n}(z)$, both the $R(z)$ and $Q^2(z)$ must be specified, and the most convenient approach will be to associate various $R(z)$ with a given typical $Q^2(z)$. This will be shown later to be helpful also in defining $Q^2(z)$ for given $R(z)$ (Secs. V and VI).

We assume that $Q^2(z)$ has a zero or singularity at $z = 0$, and can be approximated in the vicinity of $z = 0$ by a simple model $Q^2_M(z)$:

$$Q^2(z) = Q^2_M(z)[1 + d(z)]\,, \tag{3.1}$$

where $d(z) \to 0$ as $z \to 0$. We write $\varepsilon_0(z)$ given by (1.5a) as

$$\varepsilon_0(z) = (\varepsilon_0)_L[1 + d_0(z)] . \tag{3.2}$$

Here and in what follows the subscript L denotes the leading contribution as $z\to 0$, which means that the function $d(z)$ with an appropriate subscript tends to zero as $z\to 0$; if this d function has no definite limit as $z\to 0$, but it tends to zero as $x\to 0\pm$, the analysis will be restricted to the real axis [e.g., if the d function has an essential singularity at $z=0$]. Equations (3.1) and (1.5b) lead to

$$\begin{aligned} Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2} &= Q_M^{1/2}\frac{d^2Q_M^{-1/2}}{dz^2} + \frac{1}{8}\frac{d'(z)}{1+d(z)}\frac{d}{dz}\ln(Q_M^2) \\ &\quad - \frac{1}{4}\frac{d''(z)}{1+d(z)} + \frac{5}{16}\left[\frac{d'(z)}{1+d(z)}\right]^2 \\ &\equiv \left(Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right)_L [1 + d_{1/2}(z)] . \end{aligned} \tag{3.3}$$

Inserting Eq. (3.3) into (1.5a) we can determine $(\varepsilon_0)_L$ $[=2(Y_2)_L]$. The leading terms $(Y_{2n})_L$, $n>1$, can then be obtained by differentiating $d/d\zeta$, Eq. (3.2), and using the results in Eqs. (1.6). This procedure can often be simplified by noticing that the leading term and the d function relevant to ε_{k+1} can be written

$$(\varepsilon_{k+1})_L = \frac{d(\varepsilon_k)_L}{d\zeta_M}, \tag{3.4a}$$

and

$$d_{k+1}(z) = d_k(z) + d'_k(z)\left[\frac{d}{dz}\ln(\varepsilon_k)_L\right]^{-1} + O[d(z)], \tag{3.4b}$$

if

$$\lim_{z\to 0} d'_k(z)\left[\frac{d}{dz}\ln(\varepsilon_k)_L\right]^{-1} = 0, \quad k\geqslant 0; \tag{3.5}$$

ζ_M is defined as

$$\zeta_M = \int Q_M(z)dz, \tag{3.6}$$

with any convenient choice of the integration constant, and the O term is a power series in $d(z)$. Equation (3.4a) yields

$$(\varepsilon_k)_L = \frac{d^k(\varepsilon_0)_L}{d\zeta_M^k}, \quad k=0,1,2,\ldots . \tag{3.7}$$

Thus, if condition (3.5) is fulfilled for successive k (which can be checked *a posteriori*), $(\varepsilon_k)_L$ can be determined by differentiating $(\varepsilon_0)_L$ with respect to ζ_M.

The leading terms in the z plane ($z\to 0$) can be transformed to the $\tilde{z}$ plane ($\tilde{z}\to\infty$) by using Eqs. (2.4) and (2.7),

$$[\tilde{Y}_{2n}(\tilde{z})]_L \equiv [Y_{2n}(z)]_L, \quad z\tilde{z} = L^2 . \tag{3.8}$$

At the same time the functions $\tilde{Q}^2(\tilde{z})$ and $\tilde{R}(\tilde{z})$ are related to $Q^2(z)$ and $R(z)$ by Eq. (2.8). If also the model $Q_M^2(z)$ is transformed according to (2.8), the form of Eq. (3.1) will be invariant:

$$\tilde{Q}^2(\tilde{z}) = \tilde{Q}_M^2(\tilde{z})[1 + \tilde{d}(\tilde{z})], \tag{3.9a}$$

where

$$\tilde{d}(\tilde{z}) \equiv d(z), \quad z\tilde{z} = L^2 . \tag{3.9b}$$

This implies the same invariance for the remaining Eqs. (3.2)–(3.7). Therefore it is only a matter of convenience whether to perform our analysis for $Q^2(z)$ given by Eq. (3.1) with $z\to 0$, or $\tilde{Q}^2(\tilde{z})$ given by (3.9a) with $\tilde{z}\to\infty$. In any case the final results can easily be transformed to the other limit by using Eqs. (3.8) and (2.8). To simplify the notation used in the following sections, a tilde will be omitted when referring to the invariant quantities, such as ζ, ζ_M, ε_k, Y_{2n}, or the d functions.

IV. COMMON BEHAVIOR OF THE PHASE-INTEGRAL APPROXIMATIONS

In this section we examine the phase-integral approximations in some vicinity of a zero or a singularity of $Q^2(z)$ (at $z=0$) under the assumption that the leading contribution to $\varepsilon_0(z)$ comes from the differential term in Eq. (1.5a). This requirement can be written as

$$\lim_{z\to 0} D(z) = 0, \quad D(z) = \frac{a[R(z) - Q^2(z)]}{(Q^{1/2}d^2Q^{-1/2}/dz^2)_L}, \tag{4.1}$$

where $a=$ const. Condition (4.1) is obviously fulfilled if $Q^2(z) = R(z)$, but in practice it usually holds also with the other choice of the base function. Therefore this behavior will be referred to as common. Examples of a different behavior will be given in Sec. V. Recalling the fact that the ratio of $R(z) - Q^2(z)$ to the differential term in (1.5a) is invariant under the transformation (2.8) it follows that condition (4.1) is fulfilled if and only if an analogous condition is fulfilled in the $\tilde{z}$ plane, as $\tilde{z}\to\infty$. The constants a and $\tilde{a}$ can be chosen in any convenient way. Equations (1.5a), (3.1), (3.2), and (4.1) lead to

$$(\varepsilon_0)_L = Q_M^{-2}(z)\left(Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right)_L, \tag{4.2a}$$

$$d_0(z) = D(z)/a + d_{1/2}(z) + O[d(z)], \tag{4.2b}$$

where $d_{1/2}(z)$ is defined in (3.3), and the O term is a power series in $d(z)$. Equations (4.2) have the same form in the $\tilde{z}$ plane.

The general theory developed so far will now be illustrated by two groups of models that cover many practically interesting cases.

A. The power models

If $Q^2(z)$ exhibits a power behavior as $z\to 0$, i.e.,

$$Q_M^2(z) = cz^m, \quad m\neq 0, \tag{4.3a}$$

the same type of behavior is obtained in the $\tilde{z}$ plane,

$$\tilde{Q}_M^2(\tilde{z}) = \tilde{c}\tilde{z}^{\tilde{m}}, \quad \tilde{m}\neq -4 \tag{4.3b}$$

[$\tilde{c} = cL^{2(m+2)}$, $\tilde{m} = -(m+4)$, see Eq. (2.8)]. In our analysis the constant c can be complex (nonzero), whereas m will be assumed to be real. In practice m is usually an integer, in which case $Q^2(z)$ has either a zero or a pole at $z=0$, but this restriction is not necessary in our discussion.

For $Q_M^2(z)$ given by Eq. (4.3a) we obtain

$$Q_M^{1/2}\frac{d^2Q_M^{-1/2}}{dz^2} = \frac{m(m+4)}{16z^2}, \quad \frac{d}{dz}\ln(Q_M^2) = \frac{m}{z}, \tag{4.4}$$

and

$$\zeta_M = c^{1/2}\times\begin{cases}[2/(m+2)]z^{(m+2)/2}, & m\neq -2, \quad (4.5a)\\ \ln(z), & m=-2. \quad (4.5b)\end{cases}$$

The following analysis for the power models assumes that the function $d(z)$ in Eq. (3.1) satisfies

$$\lim_{z\to 0}[d(z), z\,d'(z), z^2 d''(z),...] = 0. \quad (4.6)$$

Using Eqs. (4.4), (4.6), and (3.3) it can be seen that $Q_M^{1/2} d^2 Q_M^{-1/2}/dz^2$ is the leading term in (3.3), and $d_{1/2}(z)$ also satisfies condition (4.6). Therefore condition (4.1) can be written (for $m\neq 0, -4$)

$$\lim_{z\to 0} D(z) = 0, \quad D(z) = [R(z) - Q^2(z)]z^2, \quad (4.7)$$

and Eqs. (4.2a), (4.4), and (4.5a) yield

$$(\varepsilon_0)_L = \frac{m(m+4)}{16cz^{m+2}} = \begin{cases} b\zeta_M^{-2}, & m\neq 0,-2,-4, \quad (4.8a)\\ -(4c)^{-1}, & m=-2, \quad (4.8b)\end{cases}$$

where

$$b\equiv b(m) = \frac{m(m+4)}{4(m+2)^2} = \frac{1}{4} - \frac{1}{(m+2)^2}. \quad (4.9)$$

Equations (3.7), (4.8a), and (4.5a) lead to

$$(\varepsilon_k)_L = (-1)^k b(k+1)!\zeta_M^{-(k+2)}, \quad k\geqslant 0, \quad m\neq 0,-2,-4, \quad (4.10)$$

and

$$\left[\frac{d}{dz}\ln(\varepsilon_k)_L\right]^{-1} = -2[(m+2)(k+2)]^{-1}z. \quad (4.11)$$

Using Eq. (4.11) it can be seen that the validity condition (3.5) for (4.10) will be fulfilled if the functions $d(z)$ and $D(z)$ [Eq. (4.7)] both satisfy condition (4.6). (In that case a repeated use of the operator zd/dz on the function $d_0(z)$ [Eq. (4.2b)] produces a result tending to zero as $z\to 0$.) The same relations (4.4)–(4.11) and the same validity conditions are obtained for quantities with a tilde (as $\tilde{z}\to\infty$). Equation (4.6) is obviously fulfilled if $d(z)$ and $D(z)$ are regular at $z=0$, but it allows also for some singular behavior, e.g.,

$$d(z), D(z) \sim z^p \ln^q(z), \quad p>0, \quad q \text{ real}, \quad (4.12)$$

and similarly for quantities with a tilde ($\tilde{p} = -p < 0$).

For $m=-2$ the $(\varepsilon_k)_L$, $k>0$, can be determined by a direct differentiation $d/d\zeta$ of Eq. (3.2) in which $(\varepsilon_0)_L = -(4c)^{-1}$. Using the identity

$$\frac{d}{d\zeta} = c^{-1/2}[1+d(z)]^{-1/2} z\frac{d}{dz}, \quad m=-2, \quad (4.13)$$

and assuming again that $d(z)$ and $D(z)$ satisfy condition (4.6) we obtain $(\varepsilon_k)_L\to 0$ as $z\to 0$ $(k>0)$. Therefore all but the first term in brackets in (1.6c) tend to zero as $z\to 0$, leading to $(\tilde{c}=c)$

$$(Y_{2n})_L = -(2n)![2^{4n}(2n-1)(n!)^2 c^n]^{-1}, \quad \text{for } m=\tilde{m}=-2. \quad (4.14)$$

The problem is more complicated for $(\varepsilon_k)_L$ given by (4.10), as in that case all terms in (1.6c) contribute to the $(Y_{2n})_L$. Each term is proportional to ζ_M^{-2n}, leading to

$$(Y_{2n})_L = P_n(b)\zeta_M^{-2n}, \quad m\neq 0,-2,-4, \quad (4.15)$$

where $P_n(b)$ is a polynomial of order n in b. However, if $|b|\lesssim 1$, the dominant contribution to $P_n(b)$ comes from the last term in Eq. (1.6c), which contains the largest factorial, $(2n-1)!$, and the lowest power of b. Thus we can write

$$(Y_{2n})_L = \beta_n(-1)^{n+1}2b(2n-1)!(2\zeta_M)^{-2n}, \quad m\neq 0,-2,-4, \quad (4.16)$$

where $\beta_n\equiv\beta_n(m)$, the ratio of the $P_n(b)$ to the contribution coming from the last term in (1.6c), is close to unity. This representation was first introduced in Ref. 4 for integer m; β_n was shown to be slightly increasing (for $|m+2|>2$) or decreasing (for $|m+2|<2$) from $\beta_1(m)\equiv 1$ to $\beta_\infty(m)$, see Table I. [Note that both $b(m)$ and $\beta_n(m)$ are functions of $|m+2|$.] The arguments given in Ref. 4 can easily be shown to be generally valid if $|b|\lesssim 1$ [$-\frac{3}{4}\leqslant b<\frac{1}{4}$ if $m=$ integer $\neq -2$]. Table I indicates that $\beta_n(m)$ varies slowly as a function of n and is not much different from unity if $|m+2|>\frac{2}{3}$, in which case $-2<b<\frac{1}{4}$. If $|m+2|=\frac{2}{3}$, one obtains $\beta_n(m)\simeq 0$ for $n\geqslant 8$, and if $|m+2|<\frac{2}{3}$, $\beta_n(m)$ changes sign and is no longer monotone as a function of n. In that case no simple analytical representation of the $(Y_{2n})_L$ can be given. Equations (4.15) and (4.16) have the same form in the $\tilde{z}$ plane; ζ_M is defined by Eq. (4.5a). We recall

TABLE I. Typical values of the coefficient $\beta_n(m)$, defined in Eq. (4.16), for $n>2$ and $|m+2|\geqslant\frac{2}{3}$ [$\beta_1(m)\equiv 1$]. The estimates for $\beta_\infty(m)$ were calculated from the recurrence relation (B5) in Ref. 4 [where now m can be real, and d_m is denoted by $b(m)$].

$\|m+2\|$	$b(m)$	β_2	β_3	β_4	β_{10}	β_∞
0.667	−2.000	0.667	0.133	0.013	-4×10^{-7}	
0.756	−1.500	0.750	0.337	0.206	0.133	0.104
0.894	−1.000	0.833	0.550	0.434	0.332	0.283
1.000	−0.750	0.875	0.659	0.562	0.462	0.410
1.155	−0.500	0.917	0.771	0.699	0.615	0.568
1.500	−0.194	0.968	0.910	0.878	0.837	0.811
2.500	0.090	1.015	1.042	1.058	1.082	1.097
3.000	0.139	1.023	1.065	1.090	1.128	1.153
3.500	0.168	1.028	1.079	1.110	1.157	1.188
4.000	0.187	1.031	1.088	1.123	1.176	1.211
6.000	0.222	1.037	1.105	1.146	1.210	1.253
∞	0.250	1.042	1.118	1.165	1.238	1.288

that Eqs. (4.14)–(4.16) are valid if both $d(z)$ and $D(z)$ satisfy condition (4.6).

In the remaining cases of $m=-4$ or $\tilde{m}=0$ the leading term in Eq. (3.3), used in either the z or $\tilde{z}$ plane, depends on the d function. The analysis is simpler in the $\tilde{z}$ plane, where both terms given by (4.4) are zero. Two groups of models for the d functions will be considered: the power models,

$$d=g_p z^p[1+d_p(z)]\equiv\tilde{g}_p\tilde{z}^{\tilde{p}}[1+d_p(\tilde{z})], \qquad (4.17a)$$

$$p>0, \quad \tilde{p}=-p, \quad \tilde{g}_p=g_p L^{2p}, \qquad (4.17b)$$

and the exponential models,

$$d=\gamma_e \exp(-\eta L/x)[1+d_e(x)], \quad L/x=\tilde{x}/L, \qquad (4.18)$$

where $|\eta|=1$, and the phase of η must be such that $\mathrm{Re}(\eta\tilde{x})>0$ as $\tilde{x}\to\pm\infty$, i.e., $\mathrm{Re}(\eta/x)>0$ as $x\to 0\pm$; $d_p(z)$ and $d_e(x)$ tend to zero as $z\to 0$ or $x\to 0\pm$. In either case a leading term in Eq. (3.3) is that containing $d''(\tilde{z})$, and the leading contribution to $Y_{2n}(\tilde{z})$ comes from the last term in Eq. (1.6c); $(\tilde{\varepsilon}_k)_L$ can be calculated from Eq. (3.7). For the d function given by Eqs. (4.17) the final results are as follows (written for convenience in the z plane). Condition (4.1) reads

$$\lim_{z\to 0} D(z)=0, \quad D(z)=[R(z)-Q^2(z)]z^{2-p}, \qquad (4.19)$$

and

$$(Y_{2n})_L=(-1)^n|p|(|p|+1)\cdots(|p|+2n-1)\gamma^{|p|}\times g_p L^p 2^{-(2n+1)}\zeta_M^{-(2n+|p|)}, \qquad (4.20)$$

where

$$\gamma=c^{1/2}/L=\tilde{c}^{1/2}L, \qquad (4.21)$$

$$\zeta_M=\gamma L/z\equiv\gamma\tilde{z}/L. \qquad (4.22)$$

Equations (4.19) and (4.20) have the same form in the $\tilde{z}$ plane, $\tilde{z}\to\infty$. The corresponding results for the d function given by Eq. (4.18) are

$$\lim_{x\to 0\pm} D(x)=0, \quad D(x)=[R(x)-Q^2(x)]x^4\exp(\eta L/x), \qquad (4.23a)$$

or

$$\lim_{\tilde{x}\to\pm\infty} \tilde{D}(\tilde{x})=0, \quad \tilde{D}(\tilde{x})=[\tilde{R}(\tilde{x})-\tilde{Q}^2(\tilde{x})]\exp(\eta\tilde{x}/L), \qquad (4.23b)$$

and

$$(Y_{2n})_L=\tfrac{1}{2}[-\eta^2/(4\gamma^2)]^n\gamma_e\exp(-\eta\zeta_M/\gamma), \qquad (4.24)$$

where γ and ζ_M/γ are given by Eq. (4.21), and Eq. (4.22) specialized to the real axis. Conditions (4.19) and (4.23) are stronger than (4.7). The validity condition (3.5) for Eq. (4.20) is fulfilled if $d_p(z)$ in Eq. (4.17a) and $D(z)$ [Eq. (4.19)] satisfy condition (4.6), and the same in the $\tilde{z}$ plane, $\tilde{z}\to\infty$. The corresponding condition for Eq. (4.24) is weaker. It takes a simple form in the $\tilde{z}$ plane, where the left-hand side of Eq. (4.11) is constant. Thus the function $d_e(x)$ or $d_e(\tilde{x})$ in Eq. (4.18) is required to satisfy

$$\lim_{x\to 0\pm}[d_e(x),x^2d_e'(x),x^4d_e''(x),\ldots]=0, \qquad (4.25a)$$

or

$$\lim_{\tilde{x}\to\pm\infty}[d_e(\tilde{x}),d_e'(\tilde{x}),d_e''(\tilde{x}),\ldots]=0, \qquad (4.25b)$$

and the same for $D(x)$ or $\tilde{D}(\tilde{x})$ in Eqs. (4.23).

B. The exponential models

The second group of models is given by

$$Q_M^2(z)=\gamma^2L^2z^{-4}\exp(\eta L/z), \qquad (4.26a)$$

which corresponds to the purely exponential behavior in the $\tilde{z}$ plane,

$$\tilde{Q}_M^2(\tilde{z})=\gamma^2L^{-2}\exp(\eta\tilde{z}/L), \qquad (4.26b)$$

where γ is a dimensionless parameter, L (>0) is a characteristic length, and η is the phase factor, $|\eta|=1$; γ and η can be complex. The analysis is simpler in the $\tilde{z}$ plane, where we immediately find

$$\tilde{Q}_M^{1/2}\frac{d^2\tilde{Q}_M^{-1/2}}{d\tilde{z}^2}=\frac{\eta^2}{16L^2}, \quad \frac{d}{d\tilde{z}}\ln(\tilde{Q}_M^2)=\frac{\eta}{L}, \qquad (4.27)$$

and

$$\zeta_M=2\gamma\eta^{-1}\exp(\eta\tilde{z}/2L), \quad \tilde{z}/L=L/z. \qquad (4.28)$$

Condition (4.6) for the d function can now be replaced by a weaker condition of the form (4.25a) in the z plane ($z\to 0$) or (4.25b) in the $\tilde{z}$ plane ($\tilde{z}\to\infty$). This condition and Eqs. (4.27) indicate that the leading term in Eq. (3.3) (specialized to the $\tilde{z}$ plane) is given by the first of Eqs. (4.27). Hence condition (4.1) can be written

$$\lim_{z\to 0} D(z)=0, \quad D(z)=[R(z)-Q^2(z)]z^4, \qquad (4.29a)$$

or

$$\lim_{\tilde{z}\to\infty}\tilde{D}(\tilde{z})=0, \quad \tilde{D}(\tilde{z})=\tilde{R}(\tilde{z})-\tilde{Q}^2(\tilde{z}), \qquad (4.29b)$$

and Eq. (4.2) leads to

$$(\varepsilon_0)_L=\tfrac{1}{4}\zeta_M^{-2}. \qquad (4.30)$$

Conditions (4.29) are weaker than the weakest conditions for the power models, see Eq. (4.7) used in either the z or $\tilde{z}$ plane. The following discussion can be considerably simplified by noticing that the ζ_M dependence of $(\varepsilon_0)_L$ given by Eq. (4.30) is the same as that in Eq. (4.8a) if $m\to\infty$ [$b(\infty)=\frac{1}{4}$]. Therefore $(\varepsilon_k)_L$ and $(Y_{2n})_L$ can be immediately obtained as functions of ζ_M by using Eqs. (4.10) and (4.16) in the limit $m\to\infty$. This leads to

$$(Y_{2n})_L=\beta_n(\infty)(-1)^{n+1}\tfrac{1}{2}(2n-1)!(2\zeta_M)^{-2n}, \qquad (4.31)$$

where $\beta_n(\infty)$ is close to unity (see Table I) and ζ_M is given by Eq. (4.28). It can be shown, using Eq. (4.10) with $m\to\infty$, and Eq. (4.28), that the validity condition (3.5) is fulfilled if both $d(z)$ and $D(z)$ [Eqs. (4.29)] satisfy conditions of the form (4.25). Note also that in contrast to the exponential d functions (4.18) (which must tend to zero), the analysis given here for the exponential models of $\tilde{Q}^2(\tilde{z})$ is not restricted to the real axis.

V. THE OPTIMIZED PHASE-INTEGRAL APPROXIMATIONS

The phase-integral approximation will be called optimized at a given zero or singularity of $Q^2(z)$ (at $z=0$), if the leading contributions to $R(z) - Q^2(z)$ and to the differential term in Eq. (1.5a) as $z \to 0$ cancel each other. This implies the following relation between $Q^2(z)$ and $R(z)$:

$$Q^2(z) = R(z) + \left(Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right)_L [1 + d_{\rm opt}(z)]$$
$$\equiv Q_M^2(z)[1 + d(z)], \qquad (5.1)$$

where $d_{\rm opt}(z) \to 0$ as $z \to 0$. Equations (1.5a) and (5.1) lead to

$$\varepsilon_0(z) = Q_M^{-2}(z)\left(Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right)_L [d_{1/2}(z) - d_{\rm opt}(z)]$$
$$\times [1 + d(z)]^{-1} \equiv (\varepsilon_0)_L [1 + d_0(z)]. \qquad (5.2)$$

The following analysis will be restricted only to such models $Q_M^2(z)$ for which

$$\left(Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right)_L = Q_M^{1/2}\frac{d^2Q_M^{-1/2}}{dz^2}, \qquad (5.3)$$

for any admissible d function [see Eq. (3.3)]. In that case the $(\varepsilon_0)_L$ defined by Eq. (5.2) depends on $Q_M^2(z)$ and on the leading term of $d(z)$ only, a simple enough situation to allow for a general treatment. Equation (5.3) eliminates the power models with $m = 0, \ -4$ [see Eq. (4.4)]. Furthermore we impose the requirement that

$$\left| Q_M^{-2}(z)\left(Q^{1/2}\frac{d^2Q^{-1/2}}{dz^2}\right)_L \right| < \infty, \quad \text{as } z \to 0, \qquad (5.4)$$

according to which the $\varepsilon_0(z)$ given by Eq. (5.2) tends to zero as $z \to 0$ for any choice of $d(z)$ and $d_{\rm opt}(z)$. Condition (5.4) eliminates the models which for the common behavior of the phase-integral approximation lead to infinite values of $\varepsilon_0(z)$ as $z \to 0$ [see Eq. (4.2a)].

For a given model $Q_M^2(z)$ satisfying the requirements (5.3) and (5.4), Eq. (5.1), with arbitrary functions $d_{\rm opt}(z)$ and $d(z)$, defines a class of the admissible $R(z)$. Choosing the $R(z)$ to belong to this class, Eq. (5.1) can then be used to define $Q^2(z)$ in terms of $R(z)$. If the left-hand side of Eq. (5.4) tends to zero, Eq. (5.1) leads to $R(z)/Q^2(z) \to 1$ as $z \to 0$. Hence in that case both $Q^2(z)$ and $R(z)$ are approximated by the same model:

$$R(z) = Q_M^2(z)[1 + d_R(z)], \qquad (5.5)$$

where $d_R(z) \to 0$ as $z \to 0$. This statement is true also for the common behavior of the phase-integral approximation, where the difference $R(z) - Q^2(z)$ is negligible in comparison to that defined by Eq. (5.1).

All arguments given here in the z plane, $z \to 0$, are directly applicable to the $\tilde{z}$ plane, $\tilde{z} \to \infty$, leading again to Eqs. (5.1)–(5.5), but for quantities with a tilde. The phase-integral approximation is optimized at $z = 0$ if and only if it is optimized at $\tilde{z} = \infty$, see Eq. (2.8) and the following comment.

The theory given in this section will now be illustrated by the simple models introduced in Sec. IV.

A. The power models

For the power models (4.3) Eqs. (5.1) and (4.4) lead to

$$Q^2(z) = R(z) + [m(m+4)/16z^2][1 + d_{\rm opt}(z)]$$
$$\equiv cz^m[1 + d(z)], \qquad (5.6)$$

which has the same form in the $\tilde{z}$ plane. Conditions (5.3) and (5.4) are fulfilled if either

$$-4 \neq m < -2, \quad \text{i.e.,} \quad -2 < \tilde{m} \neq 0, \qquad (5.7)$$

or $m = \tilde{m} = -2$. If Eq. (5.7) holds, the left-hand side of Eq. (5.4) tends to zero, and the admissible $R(z)$ or $\tilde{R}(z)$ is given by the rightmost member of Eq. (5.6) [where the d function should be replaced by some function $d_R(z)$ or $d_R(\tilde{z})$].

For $m = -2$ or $\tilde{m} = -2$ the left-hand side of Eq. (5.4) tends to a constant, and Eq. (5.6) reduces to the well-known definition[2-4,8]

$$Q^2(z) = R(z) - (1/4z^2)[1 + d_{\rm opt}(z)] \equiv cz^{-2}[1 + d(z)]. \qquad (5.8)$$

It can be used if either

$$\lim_{z \to 0} z^2 R(z) = 0, \qquad (5.9)$$

$(c = -\frac{1}{4})$, or

$$R(z) = c_R z^{-2}[1 + d_R(z)] \qquad (5.10)$$

$(c = c_R - \frac{1}{4})$. Equations (5.8)–(5.10) have the same form in the $\tilde{z}$ plane, $\tilde{z} \to \infty$.

If $Q^2(z)$ is defined by Eq. (5.6) [or (5.8)], the function $d(z)$ is often of the power type [see Eq. (4.17a)]. In that case Eqs. (3.3) and (4.4) lead to

$$[d_{1/2}(z)]_L = [2p(m + 2 - 2p)/m(m+4)]g_p z^p \equiv c_p z^p. \qquad (5.11)$$

Assuming for simplicity that $d_{\rm opt}(z)$ is negligible [$d_{\rm opt}(z) \equiv 0$, or $d_{\rm opt}(z)/d_{1/2}(z) \to 0$ as $z \to 0$], we find from Eqs. (5.2) and (5.11)

$$(\varepsilon_0)_L = \begin{cases} \gamma_\varepsilon \zeta_M^{-2(1+\alpha)}, & -4 \neq m < -2, \quad (5.12a) \\ -(4c)^{-1} c_p z^p, & m = -2, \quad (5.12b) \end{cases}$$

where

$$\alpha = -p/(m+2) > 0, \quad \gamma_\varepsilon = b\,[4c/(m+2)^2]^\alpha c_p; \qquad (5.13)$$

ζ_M, b, and c_p are defined by Eqs. (4.5a), (4.9), and (5.11). [Equations (5.12a)–(5.13), and the following Eqs. (5.14) and (5.15), are generally valid if $d_{\rm opt}(z)$ exhibits a power behavior, but the meaning of p and c_p may be different.]

If $Q^2(z)$ is given by Eq. (5.8) $(m = -2)$, Eqs. (5.12b), (5.2), and (4.13) lead to $(\varepsilon_k)_L = \text{const} \times z^p$. Therefore the $(Y_{2n})_L$ comes from the last term in brackets in Eq. (1.6c),

$$(Y_{2n})_L = \tfrac{1}{2}(-p^2/4c)^n (c_p/p^2) z^p, \qquad (5.14)$$

where $c_p/p^2 = g_p$ if c_p is given by Eq. (5.11).

For $Q^2(z)$ given by Eq. (5.6) with $m \neq -2$, we calculate $(\varepsilon_k)_L$ from Eqs. (3.7) and (5.12a). Again the $(Y_{2n})_L$ comes from the last term in Eq. (1.6c),

$$(Y_{2n})_L = (-1)^{n+1} 2^{-(2n-1)} \gamma_\varepsilon$$
$$\times [(2 + 2\alpha) \cdots (2n - 1 + 2\alpha)] \zeta_M^{-2(n+\alpha)}. \qquad (5.15)$$

The expression in the brackets in Eq. (5.15) must be replaced by unity if $n = 1$. Equations (5.14) and (5.15) are valid if $d_p(z)$ in Eq. (4.17a) satisfies condition (4.6). These results have the same form in the $\tilde{z}$ plane. [Note that α and γ_ε are invariant under the transformation (2.8).]

B. The exponential models

Specializing Eqs. (4.26) to the real axis, and using (4.27) and (5.1), we obtain

$$Q^2(x) = R(x) + (\eta^2 L^2/16)x^{-4}[1 + d_{\rm opt}(x)]$$
$$\equiv \gamma^2 L^2 x^{-4} \exp(\eta L/x)[1 + d(x)], \quad (5.16a)$$
$$\tilde{Q}^2(\tilde{x}) = \tilde{R}(\tilde{x}) + (\eta^2 L^{-2}/16)[1 + d_{\rm opt}(\tilde{x})]$$
$$\equiv \gamma^2 L^{-2} \exp(\eta\tilde{x}/L)[1 + d(\tilde{x})]. \quad (5.16b)$$

Conditions (5.3) and (5.4) are fulfilled if either

$$\mathrm{Re}(\eta/x), \mathrm{Re}(\eta\tilde{x}) > 0, \quad (5.17)$$

or $\eta = \pm i$. The last two cases are of less general importance and will not be discussed here. If Eq. (5.17) holds, the left-hand side of Eq. (5.4) tends to zero as $x \to 0\pm$ or $\tilde{x} \to \pm\infty$. Therefore the admissible $R(x)$ or $\tilde{R}(\tilde{x})$ are given by the rightmost members of Eqs. (5.16), where again the d functions should be replaced by some d_R functions. To illustrate our theory we chose the $d_{\rm opt}$ function in (5.16) to be zero, and assume that the d function exhibits the exponential behavior ($\alpha > 0$),

$$d = \gamma_e \exp(-\alpha\eta L/x)[1 + d_e(x)], \quad L/x = \tilde{x}/L, \quad (5.18)$$

where the d_e function tends to zero as $x \to 0\pm$ or $\tilde{x} \to \pm\infty$. In that case $(\varepsilon_0)_L$ is again given by Eq. (5.12a) if one defines

$$\gamma_\varepsilon = -\tfrac{1}{2}\gamma_e \alpha(2\alpha + 1)(2\gamma/\eta)^{2\alpha}, \quad (5.19)$$

and calculates ζ_m from Eq. (4.28) specialized to the real axis. With this new meaning of α, γ_ε, and ζ_M, Eq. (5.15) is valid for the $Q^2(x)$ or $\tilde{Q}^2(\tilde{x})$ given by Eqs. (5.16). The validity condition (3.5) is fulfilled if the d_e function in Eq. (5.18) satisfies conditions (4.25). Our assumptions $d_{\rm opt}(\tilde{x}) \equiv 0$ and (5.18) imply the following admissible form of $d_R(\tilde{x})$:

$$d_R(\tilde{x}) = \gamma_e \exp(-\alpha\eta\tilde{x}/L)[1 + d_e(\tilde{x})]$$
$$- (\eta/4\gamma)^2 \exp(-\eta\tilde{x}/L). \quad (5.20)$$

The leading term of the $\tilde{R}(\tilde{x})$ fixes the parameters γ^2, L, and η [see Eq. (5.16b)] whereas γ_e, α, and the function $d_e(\tilde{x})$ remain free. Hence the $d_R(\tilde{x})$ can be an arbitrary function tending to zero faster than $\exp(-\eta\tilde{x}/L)$ [$\alpha = 1$, $\gamma_e = (\eta/4\gamma)^2$], or behaving like $\exp(-\eta\tilde{x}/L)$ [$\alpha = 1$, $\gamma_e \neq (\eta/4\gamma)^2$], or like $\exp(-\alpha\eta\tilde{x}/L)$, $0 < \alpha < 1$. For $\alpha > 1$, the leading term of the $d_R(\tilde{x})$ is fixed by $\tilde{Q}^2_M(\tilde{x})$, and therefore these cases are of less general importance. The admissible $d_R(x)$ is obtained on replacing in Eq. (5.20) $\tilde{x}/L \to L/x$.

VI. INTEGRALS DEFINING ACCURACY OF THE PHASE-INTEGRAL APPROXIMATION

Linear combinations of the functions $f_1(z)$ and $f_2(z)$ given by Eq. (1.2) can be used to trace approximately the wave function $\psi(z)$ from a given point z_0 to other points z. If z_0 and z can be joined by a path Λ along which the modulus of the exponential in (1.2) changes monotonically, accurate error estimates for this tracing can be given.[11] They involve the "μ integral" calculated along Λ,[8] and two additional "ν integrals."[11] If the $q(z)$ in Eq. (1.2) is given by (1.3), and $|Y_{2n}| \ll 1$ along the integration path, $1 \leqslant n \leqslant N$, the integrals in question can be expressed in terms of the correction Y_{2N+2} [see Eqs. (2.24)–(2.25b) in Ref. 11]. Our aim here is to derive simple analytical approximations to these integrals, related to the models introduced in Secs. IV and V. Choosing the integration path Λ to belong to some vicinity of $z = 0$ or $\tilde{z} = \infty$, we can approximate in the relevant integrands

$$Y_{2N+2}(\zeta_1)d\zeta_1 \simeq (Y_{2N+2})_L d\zeta_{M1},$$

where ζ_M is defined by Eq. (3.6). In this approximation we obtain

$$\mu_{2N+1} = \int_{\zeta_{M0}}^{\zeta_M} |d\zeta_{M1}\, 2(Y_{2N+2})_L|, \quad (6.1)$$

$$\nu^{\rm dec}_{2N+1} = \left| \int_{\zeta_{M0}}^{\zeta_M} d\zeta_{M1}\, 2(Y_{2N+2})_L \times \exp[-2i(\zeta_{M1} - \zeta_{M0})] \right|, \quad (6.2a)$$

$$\nu^{\rm inc}_{2N+1} = \left| \int_{\zeta_{M0}}^{\zeta_M} d\zeta_{M1}\, 2(Y_{2N+2})_L \times \exp[2i(\zeta_{M1} - \zeta_M)] \right|. \quad (6.2b)$$

The integration path Λ and the sign of ζ_M must be chosen so that $|\exp(i\zeta_{M1})|$ increases or is constant as one moves along Λ from z_0 to z.

In the following analysis Eqs. (6.1) and (6.2) will only be used if the μ integral (6.1) is convergent at $z = 0$ or $\tilde{z} = \infty$ (for arbitrary N, which implies the convergence of the ν integrals), while the ζ integral (1.4) is divergent,

$$\zeta \to \infty, \quad \text{as } z \to 0 \text{ or } \tilde{z} \to \infty. \quad (6.3)$$

With these two assumptions we obtain (assuming the existence of the limits below)

$$\lim\left(Y_{2n}, \frac{d}{d\zeta}Y_{2n}, \frac{d^2}{d\zeta^2}Y_{2n}, \ldots\right) = 0, \quad z \to 0 \text{ or } \tilde{z} \to \infty, \quad (6.4)$$

$n = 1, 2, \ldots$. Equation (6.4) means that the approximation given by Eqs. (1.2), (1.3) tends to an exact solution of Eq. (1.1) as $z \to 0$ or $\tilde{z} \to \infty$. This fact will be used in Ref. 11 to demonstrate the uniqueness of the reflection, transmission, and absorption coefficients defined in terms of the phase-integral approximation. Furthermore, if Eq. (6.4) holds, the validity condition for Eqs. (6.1)–(6.2b), $|Y_{2n}| \ll 1$ along Λ, can always be fulfilled by choosing Λ close enough to $z = 0$ or $\tilde{z} = \infty$.

The divergence of ζ means that

$$\lim_{z\to 0}|z^2 Q^2(z)| \equiv \lim_{\tilde{z}\to\infty}|\tilde{z}^2\tilde{Q}^2(\tilde{z})| = \begin{cases} c \neq 0, & (6.5a) \\ \infty. & (6.5b) \end{cases}$$

Condition (6.5a) corresponds to the power models with $m = -2$ or $\tilde{m} = -2$. In that case the μ integral (6.1) is logarithmically divergent as $z \to 0$ or $\tilde{z} \to \infty$, if the phase-integral approximation is commonly behaving [$d\zeta_M = c^{1/2}\, dz/z$, and $Y_{2N+2} \to {\rm const} \neq 0$ [see Eq. (4.14)]). Improving the approximation in comparison to this "marginal" situation one can expect the convergence of the μ integral. Hence in the case (6.5a) the convergence of μ can be expected if the approximation is optimized ($Y_{2N+2} \to 0$), which is confirmed by Eq. (5.14). In the case (6.5b) the convergence of μ can be expected for the commonly behaving approximations, which implies the convergence for the optimized approximations. This is suggested by Eq. (4.2a) which indicates that $|\varepsilon_0(z)Q(z)|$ decreases when increasing $|Q^2(z)|$. By the same arguments in the case (6.5b) the left-hand side of Eq. (5.4) should tend to zero, and so both $Q^2(z)$ and $R(z)$ should be approximated by the same model [see Eq. (5.5)]. These predictions were confirmed by our models discussed in Secs. IV and V. In particular Eqs. (4.15), (4.16), (4.20), (4.31), and (5.15), which have the common form,

$$(Y_{2n})_L = C_n \zeta_M^{-2(n+\alpha)}, \quad \alpha \geqslant 0, \tag{6.6}$$

and Eq. (4.24), indicate that the integral (6.1) is convergent as $\zeta_m \to \infty$.

The ν integrals corresponding to Eq. (6.6) are nonelementary, but simple approximations can be obtained by using their asymptotic expansions for large integration limits ($k>0$),

$$\int_{\zeta_{M0}}^{\zeta_M} d\zeta\, \zeta^{-k} \exp(\pm 2\, i\zeta) = \mp \tfrac{1}{2} i\zeta^{-k} \exp(\pm 2\, i\zeta)[1 + O(\zeta^{-1})]\Big|_{\zeta_{M0}}^{\zeta_M}. \tag{6.7}$$

All other integrals pertaining to our models are elementary.

In applications described in Ref. 11 one of the integration limits in Eqs. (6.1), (6.2) is always infinite; the results depend on the other, finite, integration limit $\zeta_{MB} \equiv \zeta_M(z_B)$, where z_B (the "boundary" point) denotes either z_0 or z. Such results will be denoted by the subscript B. If the integration path is not an anti-Stokes line, i.e., Im $\zeta_{M1} \neq$ const along Λ, all our model results lead to

$$\lim_{\zeta_M \to \infty} \nu_{2N+1}^{\rm inc} = \lim_{\zeta_{M0} \to \infty} \nu_{2N+1}^{\rm dec} = 0. \tag{6.8}$$

This behavior seems to be quite general. The nonvanishing integrals (6.1)–(6.2b) (as $\zeta_{M0} \to \infty$ or $\zeta_M \to \infty$) related to our models are listed below in the form convenient for applications. For the $(Y_{2n})_L$ given by Eq. (6.6) the ν integrals were calculated by using Eq. (6.7) and neglecting the final contributions from the O terms. This leads to

$$(\nu_B)_{2N+1} = |C_{N+1}||\zeta_{MB}|^{-2(N+1+\alpha)}, \tag{6.9}$$

which denotes either $(\nu_B^{\rm dec})_{2N+1} \neq 0$, or $(\nu_B^{\rm inc})_{2N+1} \neq 0$. Equation (6.9) can be used if $|\zeta_{MB}| > N+1+\alpha$. The μ integrals were calculated for Λ being a straight line emerging from $z=0$; in that case $(\mu_B)_{2N+1}$ is a function of $|\zeta_{MB}|$.

A. Results from the power models

For the power models (4.3) a simultaneous convergence of the μ integral and divergence of ζ corresponds to $m \leqslant -2$ or $\tilde{m} \geqslant -2$.

If $m=-2$ or $\tilde{m}=-2$, one can only work with the optimized phase-integral approximation [case (6.5a)]. The admissible $R(z)$ is given by Eq. (5.9) or (5.10), and $Q^2(z)$ is related to $R(z)$ by Eq. (5.8). The results depend on the d function in Eq. (5.8). If this function is of the power type [Eq. (4.17a)], the $(Y_{2n})_L$ is given by Eq. (5.14), which leads to

$$(\mu_B)_{2N+1} = \tfrac{1}{2}|g_p z_B^p|\,|\Gamma|^{-(2N+1)}, \tag{6.10a}$$

$$(\nu_B)_{2N+1} = (\mu_B)_{2N+1}|1+i\Gamma|^{-1}, \tag{6.10b}$$

where

$$\Gamma = 2c^{1/2}|p|^{-1}. \tag{6.10c}$$

Equation (6.10a) assumes the validity of the definition (5.11) for c_p ($m=-2$). Equations (5.8) and (6.10) have the same form in the $\tilde{z}$ plane.

If $m<-2$ or $\tilde{m}>-2$, one can work with either the commonly behaving approximation or the optimized approximation [case (6.5b)]. In both situations the admissible $R(z)$ is

$$R(z) = c\, z^m[1 + d_R(z)], \tag{6.11}$$

which has the same form in the $\tilde{z}$ plane, and $Q^2(z)$ is given by the rightmost member of Eq. (5.6).

We first assume that $m \neq -4$ or $\tilde{m} \neq 0$. In that case the leading term $(Y_{2n})_L$ for the commonly behaving approximation is independent of the d function. In particular if

$$-4 \neq m < -(2+\tfrac{2}{3}) \quad \text{or} \quad -(2+\tfrac{2}{3}) < \tilde{m} \neq 0, \tag{6.12}$$

the $(Y_{2n})_L$ is given by Eq. (4.16), leading to

$$(\mu_B)_{2N+1} = 2\beta_{N+1}|b|(2N)!(2|\zeta_{MB}|)^{-(2N+1)}, \tag{6.13a}$$

$$(\nu_B)_{2N+1} = (\mu_B)_{2N+1}(2N+1)/(2|\zeta_{MB}|), \tag{6.13b}$$

where $\zeta_{MB} \equiv \zeta_M(z_B)$; β_{N+1} is close to unity (see Table I).

For the optimized approximation simple results are obtained if the d function in Eq. (5.6) is of the power type [Eqs. (4.17)]. In that case the $(Y_{2n})_L$ is given by Eq. (5.15), leading to

$$(\mu_B)_{2N+1} = G\alpha[(1+2\alpha)(2+2\alpha)\cdots(2N+2\alpha)] \times (2|\zeta_{MB}|)^{-(2N+1+2\alpha)}, \tag{6.14a}$$

$$(\nu_B)_{2N+1} = (\mu_B)_{2N+1}(2N+1+2\alpha)/(2|\zeta_{MB}|), \tag{6.14b}$$

where

$$G = |g_p||16c(m+2)^{-2}|^{\alpha}, \quad \alpha = -p/(m+2), \tag{6.15}$$

and ζ_{MB} has the same meaning as that in Eqs. (6.13). The expression in brackets in Eq. (6.14a) must be replaced by unity if $N=0$. Equation (6.15) assumes the validity of the definition (5.11) for c_p, and has the same form in the $\tilde{z}$ plane.

The last group of cases is related to the power models with $m=-4$ or $\tilde{m}=0$ (commonly behaving approximations only). The results depend on the d function. For the d functions discussed in Sec. IV [Eqs. (4.17) and (4.18)] the admissible d_R functions in (6.11) are obtained if one replaces in Eq. (4.17a) or (4.18) the function $d_p(z)$ or $d_e(x)$ by some function $d_{pR}(z)$ or $d_{eR}(x)$ of the same general properties. The validity conditions (4.19) and (4.23) can then be written as

$$\lim_{z\to 0}[d_p(z) - d_{pR}(z)]z^{-2} \equiv \lim_{\tilde{z}\to\infty}[d_p(\tilde{z}) - d_{pR}(\tilde{z})]\tilde{z}^2/L^4 = 0, \tag{6.16}$$

and

$$\lim_{x\to 0\pm}[d_e(x) - d_{eR}(x)] \equiv \lim_{\tilde{x}\to\pm\infty}[d_e(\tilde{x}) - d_{eR}(\tilde{x})] = 0. \tag{6.17}$$

With these conditions fulfilled the phase-integral approximation has a common behavior, and the $(Y_{2n})_L$ is given by Eq. (4.20) or (4.24). For the power d functions [Eqs. (4.17)] the integrals $(\mu_B)_{2N+1}$ and $(\nu_B)_{2N+1}$ are again given by Eqs. (6.14a)–(6.15), if these equations are specialized to $m = -4$ or $\tilde{m} = 0$ [$\alpha = p/2 = |\tilde{p}/2|$, and ζ_{MB} is given by Eq. (4.22)]. The corresponding results for the exponential d functions [Eq. (4.18)] are

$$(\mu_B)_{2N+1} = \tfrac{1}{2}|\mathrm{Re}\,\eta|^{-1}|\gamma_\epsilon \exp(-\eta L/x_B)|\,|2\gamma|^{-(2N+1)}, \tag{6.18a}$$

$$(\nu_B^{\rm dec})_{2N+1} = (\mu_B)_{2N+1}|\mathrm{Re}\,\eta|\,|\eta + i2\gamma|^{-1}, \tag{6.18b}$$

$$(\nu_B^{\rm inc})_{2N+1} = (\mu_B)_{2N+1}|\mathrm{Re}\,\eta|\,|\eta - i2\gamma|^{-1}, \tag{6.18c}$$

where γ is given by Eq. (4.21), and $L/x_B = \tilde{x}_B/L$.

B. Results from the exponential models

If the exponential models (4.26) are specialized to the real axis, a simultaneous convergence of the μ integral and divergence of ζ corresponds to

$$\mathrm{Re}(\eta/x), \mathrm{Re}(\eta\tilde{x}) > 0 \tag{6.19}$$

[case (6.5b) with $x \to 0\pm$ or $\tilde{x} \to \pm\infty$]. The admissible $R(x)$ or $\tilde{R}(\tilde{x})$ are given by the rightmost members of Eqs. (5.16) if the d functions are replaced by some d_R functions. Choosing $Q^2(x)$ or $\tilde{Q}^2(\tilde{x})$ so as to satisfy condition (4.29a) as $x \to 0\pm$, or (4.29b) as $\tilde{x} \to \pm\infty$, the approximation will be commonly behaving. In that case the $(Y_{2n})_L$ is given by Eq. (4.31), and the integrals $(\mu_B)_{2N+1}$ and $(\nu_B)_{2N+1}$ can be determined from Eqs. (6.13) in the limit $m \to \infty$ [$b(\infty) = \frac{1}{4}$, and $\beta_{N+1}(\infty)$ is close to unity (see Table I)]; ζ_{MB} in (6.13) is defined by Eq. (4.28) specialized to the real axis ($z = x_B$ or $\tilde{z} = \tilde{x}_B$). This definition of ζ_{MB} is applicable also to the optimized phase-integral approximation, for which $Q^2(x)$ or $\tilde{Q}^2(\tilde{x})$ are defined by Eqs. (5.16). In that case the results depend on the d function. In our example of the exponential d function (5.18) the admissible d_R function is given by Eq. (5.20), and the $(Y_{2n})_L$ is defined by (5.15). The integrals $(\mu_B)_{2N+1}$ and $(\nu_B)_{2N+1}$ can be determined from Eqs. (6.14), where now α refers to Eq. (5.18), γ_ϵ is given by (5.19), and

$$G = |\gamma_\epsilon|\,|16\gamma^2|^\alpha. \tag{6.20}$$

VII. CONCLUSIONS

If the requirement of simultaneous convergence of the μ integral and divergence of ζ is fulfilled (Sec. VI), one can always make the integrals $(\mu_B)_{2N+1}$ and $(\nu_B)_{2N+1}$, and the corrections $Y_{2n}(z)$, $1 \leqslant n \leqslant N$, along the integration path Λ, much smaller than unity by choosing the point z_B or $\tilde{z}_B$ close enough to $z = 0$ or $\tilde{z} = \infty$. In that case all validity conditions for the error estimates derived in Ref. 11 will be fulfilled if

$$\delta(z) \equiv \left|\frac{1}{2}\frac{d}{dz}Q^{-1}\right| \ll 1, \quad \text{along } \Lambda, \tag{7.1}$$

or if the condition of the same form holds in the $\tilde{z}$ plane.

Our first point is to show that the admissible $R(z)$ or $\tilde{R}(\tilde{z})$ must satisfy the same condition as required for the $Q^2(z)$ or $\tilde{Q}^2(\tilde{z})$ [Eqs. (6.5)] if condition (7.1) is imposed. Indeed, the only admissible $R(z)$ or $\tilde{R}(\tilde{z})$ for which condition (6.5) is not fulfilled is given by Eq. (5.9). It corresponds to the power model with $m = -2$ or $\tilde{m} = -2$ [see Eq. (5.8)] for which

$$(\delta)_L = \tfrac{1}{2}|c|^{-1/2} = \text{const.} \tag{7.2}$$

Equations (7.1) and (7.2) lead to $|c| \gg \frac{1}{4}$ which eliminates the $R(z)$ given by Eq. (5.9) ($c = -\frac{1}{4}$). Note also that this requirement for c imposes a constraint on the $R(z)$ given by Eq. (5.10):

$$|c_R| \gg \tfrac{1}{4}. \tag{7.3}$$

Equations (7.2) and (7.3) have the same form in the $\tilde{z}$ plane. In the remaining admissible situations condition (6.5b) is fulfilled, and one can expect $\delta, \tilde{\delta} \to 0$ as $z \to 0$ or $\tilde{z} \to \infty$ [$\delta \to \text{const} \neq 0$ if condition (6.5a) is fulfilled, and δ decreases as one increases $|Q^2|$]. With this behavior, which was confirmed by our models, condition (7.1) creates no problems.

The integrals $(\mu_B)_{2N+1}$ given in Sec. VI satisfy

$$(\mu_B)_{2N+3}/(\mu_B)_{2N+1} = (N + \tfrac{1}{2} + \alpha)(N + 1 + \alpha)|\zeta_{MB}|^{-2}, \quad |\Gamma|^{-2}, \quad |2\gamma|^{-2}; \tag{7.4}$$

here and in the following Eqs. (7.5) and (7.7) the first option on the right-hand side refers to Eq. (6.13a) ($\alpha = 0$) or (6.14a); the second option refers to Eq. (6.10a), and the third one to Eq. (6.18a). If for some $N \geqslant 0$ the parameters are chosen so that

$$(N + 1 + \alpha)^{-1}|\zeta_{MB}|, |\Gamma|, |2\gamma| \gg 1, \tag{7.5}$$

Eq. (7.4) and the relevant results for $(\mu_B)_{2N+1}$ and $(\nu_B)_{2N+1}$ indicate that

$$(\mu_B)_{2N+3} \ll (\mu_B)_{2N+1} \ll \cdots \ll (\mu_B)_1 \ll 1, \tag{7.6}$$

and

$$(\nu_B)_{2N+3}/(\mu_B)_{2N+3} = (N + \tfrac{3}{2} + \alpha)|\zeta_{MB}|^{-1}, \quad |\Gamma|^{-1}, \quad |2\gamma|^{-1}|\mathrm{Re}\,\eta| \ll 1, \tag{7.7}$$

i.e., $(\nu_B)_{2N+3} \ll (\mu_B)_{2N+3}, \ldots, (\nu_B)_1 \ll (\mu_B)_1$. This means that in the parameter range (7.5) one can use the phase-integral approximations of order $1, 3, \ldots, 2N+3$, to solve Eq. (1.1) approximately in some vicinity of $z = 0$ or $\tilde{z} = \infty$; the accuracy, defined by the μ and the ν integrals,[11] will improve when increasing the approximation order. For the last two options in Eq. (7.4) this improvement is unlimited, i.e.,

$$\lim_{N \to \infty}[(\mu_B)_{2N+1}, (\nu_B)_{2N+1}] = 0 \tag{7.8}$$

(which is valid for any $|\Gamma|, |2\gamma| > 1$). For the first option the right-hand side of Eq. (7.4) increases with N, and approaches unity at $N = N_0 \gg 1$ [$N_0 + 1 + \alpha \simeq |\zeta_{MB}|$, where $|\zeta_{MB}| \gg 1$ in view of (7.5)]. Hence in that case $(\mu_B)_{2N+1}$ reaches a minimum at $N = N_0$. Expressing the product

$$(1 + 2\alpha)(2 + 2\alpha)\cdots(2N + 2\alpha), \quad \alpha \geqslant 0,$$

in terms of the Γ function we easily find (with a slight overestimation)

$$(\mu_B)_{2N_0+1} = (\mu_B)_{2N+1}|_{\min} = C(2|\zeta_{MB}|)^{-1/2}\exp(-2|\zeta_{MB}|), \tag{7.9}$$

where

$$C = \begin{cases} 2e\beta_\infty|b|, & e \equiv \exp(1), \\ G\alpha(1 + 2\alpha)^{-(1/2 + 2\alpha)}\exp(1 + 2\alpha), \end{cases} \tag{7.10}$$

the upper line corresponding to Eq. (6.13a), and the lower line to Eq. (6.14a). Equation (7.7) but with $N = N_0 - 1$ and $|\zeta_{MB}| = N_0 + 1 + \alpha$ gives us $(\nu_B)_{2N_0+1} \simeq (\mu_B)_{2N_0+1}$, which should be correct as far as the order of magnitude of the ν integral is concerned [see Eq. (6.9)]. This means that Eq. (7.9) can be used as an estimate for both the μ and the ν integrals in question, thereby defining the best obtainable accuracy of the phase-integral approximation. Equations (7.8) and (7.9) may seem impractical ($N \to \infty$ or $N_0 \gg 1$) but they will be used in Ref. 11 to derive useful formulas pertaining to the bound state calculations.

[1]N. Fröman, Ark. Fys. **32**, 541 (1966).

[2]N. Fröman and P. O. Fröman, Ann. Phys. (NY) **83**, 103 (1974).

[3]N. Fröman and P. O. Fröman, Nuovo Cimento B **20**, 121 (1974).

[4]A. A. Skorupski, Rep. Math. Phys. **17**, 161 (1980).

[5]J. A. Campbell, J. Comput. Phys. **10**, 308 (1972).

[6]Ö. Dammert and P. O. Fröman, J. Math. Phys. **21**, 1683 (1980).

[7]N. Fröman and P. O. Fröman, "Arbitrary-order phase-integral approximation generated from an unspecified base function," in *Méthodes Semi-Classiques en Mécanique Quantique*, edited by B. Helffer and D. Robert, Colloque du 10 au 15 septembre 1984, CIRM, Luminy (Publications de l'Université de Nantes, Institut de Mathématiques et d'Informatiques, Nantes, 1985), p. 47.

[8]N. Fröman and P. O. Fröman, *JWKB-Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965).

[9]A. A. Skorupski, J. Math. Phys. **25**, 2651 (1984).

[10]A. A. Skorupski, J. Math. Phys. **25**, 2655 (1984).

[11]A. A. Skorupski, J. Math. Phys. **29**, 1824 (1988).

Efficient integration of the one-dimensional time independent wave equation for bound states and for wave propagation

Andrzej A. Skorupski
Soltan Institute for Nuclear Studies, Hoża 69, 00-681 Warsaw, Poland

(Received 9 June 1987; accepted for publication 30 March 1988)

The wave equation is solved with the phase-integral approximation in the asymptotic regions, and integrated numerically elsewhere. In this application the phase-integral approximation is relatively simple in higher orders, and its accuracy can be precisely controlled with the derived error estimates. For the bound states the analytical formulas defining the termination error in the wave function are derived. For the wave propagation an efficient formalism is given for calculating the reflection, transmission, and absorption coefficients.

I. INTRODUCTION

The aim of this paper is to optimize the integration of the wave equation

$$\frac{d^2\psi}{dx^2} + R(x)\psi = 0, \quad a<x<b, \tag{1.1}$$

by using the recent results concerning the phase-integral approximations.[1] We assume that the function $R(x)$ is sufficiently large in the vicinity of the boundary points B ($=a,b$),

$$|R(x)|_{x\to B} \geqslant \begin{cases} |\gamma(x-B)^{-2}|, & B\neq\pm\infty, \\ |\gamma x^{-2}|, & B=\pm\infty, \end{cases} \tag{1.2}$$

where $\gamma\neq 0$, and furthermore $|\gamma|\gg\frac{1}{4}$ in the case of equality. This behavior of $R(x)$ is often encountered, for example when calculating the reflection, transmission, and absorption coefficients for the electromagnetic wave propagation in a nonhomogeneous medium [where $R(x)$ is the dielectric permeability of the medium, and $-\infty<x<\infty$]. Equation (1.2) is also often fulfilled in quantum mechanical applications. For the radial Schrödinger equation we obtain

$$R(x) = E - V(x) - l(l+1)x^{-2}, \quad 0<x<\infty, \tag{1.3}$$

and Eq. (1.2) is fulfilled for the singular potentials, or the sufficiently large angular momentum quantum number l.

If Eq. (1.2) is fulfilled, one can find the phase-integral approximation which tends to an exact solution of Eq. (1.1) as $x\to B$, and improves as one increases the approximation order for $x\neq B$.[1] This approximation can be used to transform the boundary condition from the actual boundary B to some shifted boundary x_B ($=x_a,x_b$), where $a<x_a$ and $x_b<b$. Equation (1.1) can then be integrated for $x_a\leqslant x\leqslant x_b$ with any convenient numerical technique. Application of this program to bound states is given in Sec. IV, and to wave propagation in Sec. V. For the bound states it will be compared with a more standard approach in which the wave function is assumed to vanish at some shifted rather than the actual boundaries, $\bar{x}_B$. The resulting error in the base function due to this shift, called *termination error*,[2] will be shown to be related to the best obtainable accuracy of the phase-integral approximation. Thus the phase-integral approximation theory in this application cannot offer anything better than the standard approach, but it will be used to optimize the choice of $\bar{x}_B$.

In Sec. V we derive approximate expressions for the reflection, transmission, and absorption coefficients, and realistic error estimates for these approximations. The approximate expressions involve both the numerically determined quantities, and those pertaining to the phase-integral approximation. They take the simplest form in the first order, but the complications arising in higher orders are not serious. At the same time the interval for the numerical integration (x_a,x_b) shrinks as one increases the approximation order for given accuracy. This opens new possibilities for very accurate calculations. The general theory is illustrated by two examples.

In Sec. II we derive accurate error estimates for the phase-integral approximations of an arbitrary order. With these results and the analytical estimates for the relevant integrals,[1] the shifted boundaries $\bar{x}_B$ or x_B can be precisely matched to the required accuracy. The results of Sec. II can also be used in other applications of the phase-integral approximation.

II. THE ERROR ESTIMATES FOR THE PHASE-INTEGRAL APPROXIMATION

This section deals with the phase-integral approximation in its most general form.[3] Thus Eq. (1.1) is continued to the complex z plane ($z=x+iy$), and its solution $\psi(z)$ is approximated by linear combinations of the functions

$$f_{1,2}(z) = q^{-1/2}(z)\exp(\pm iw), \tag{2.1}$$

where

$$w(z) = \text{“phase integral”} = \int_{z_l}^{z} q(z_1)dz_1, \tag{2.2}$$

with any convenient choice of the function $q(z)$ and the lower limit of integration z_l. Our aim is to derive realistic error estimates for these approximations. The results will be then specialized to the phase-integral approximations discussed in Ref. 1, for which (in order $2N+1$, $N\geqslant 0$)

$$q(z) = Q(z)\sum_{n=0}^{N} Y_{2n}(z), \tag{2.3}$$

where $Q(z)$ is the base function, $Y_0(z)\equiv 1$, and $Y_{2n}(z)$, $n>0$, are the higher-order corrections.

 0022-2488/88/081824-08$02.50

A. General theory

In Ref. 3 the exact solution of Eq. (1.1) and its derivative d/dz are written as

$$\psi(z) = a_1(z)f_1(z) + a_2(z)f_2(z) , \quad \psi'(z) = a_1(z)f_1'(z) + a_2(z)f_2'(z) . \tag{2.4}$$

The "coefficients" $a_{1,2}(z)$ are examined by introducing the column vector $\mathbf{a}(z)$ with components $a_{1,2}(z)$, and the matrix $\mathbf{F}(z,z_0)$ which propagates the $\mathbf{a}(z)$ from a given point z_0,

$$\mathbf{a}(z) = \mathbf{F}(z,z_0)\mathbf{a}(z_0), \quad \det \mathbf{F}(z,z_0) = 1 . \tag{2.5}$$

Reference 3 gives the matrix elements $F_{11}(z,z_0)$, etc., in a form of the convergent series, which are then used to derive simple estimates for these elements [see Eqs. (3.22a)–(3.22d) and (4.3a)–(4.3d) in Ref. 3]. These estimates involve the "μ integral,"

$$\mu = \int_{w_0}^{w} |dw_1\, \epsilon(w_1)| , \tag{2.6a}$$

where

$$\epsilon(z) = q^{-2}(z)\left[R(z) - q^2(z) + q^{1/2}\frac{d^2 q^{-1/2}}{dz^2}\right] \tag{2.6b}$$

[$w_0 = w(z_0)$, $dw_1 = q(z_1)dz_1$], and are valid if $|\exp(iw_1)|$ increases (or is constant) as z_1 moves along the integration path Λ from z_0 to z.

We start by deriving more realistic estimates for the matrix elements $F_{12}(z,z_0)$ and $F_{21}(z,z_0)$, valid if

$$\mu \ll 1 . \tag{2.7}$$

These estimates will be given in terms of two integrals,

$$\nu^{\rm dec} = \left| \int_{w_0}^{w} dw_1\, \epsilon(w_1)\exp[-2i(w_1 - w_0)] \right| < \mu , \tag{2.8a}$$

$$\nu^{\rm inc} = \left| \int_{w_0}^{w} dw_1\, \epsilon(w_1)\exp[2i(w_1 - w)] \right| < \mu , \tag{2.8b}$$

where the superscripts refer to the behavior of the corresponding exponential factors along the integration path Λ (decreasing or increasing). The modulus of the first term in the series defining $F_{12}(z,z_0)\exp(2iw_0)$ is equal to $\nu^{\rm dec}/2$, and that for $F_{21}(z,z_0)\exp(-2iw)$ is equal to $\nu^{\rm inc}/2$ [see Eqs. (3.22b) and (3.22c) in Ref. 3]. If the remaining contributions in these two series are estimated in terms of the μ integral in the same way as was done in Ref. 3, the result (involving an overestimation) in both cases is equal to $O(\mu^2)M/4$, where $M \leqslant 1$ [M is defined by Eq. (4.1) in Ref. 3]. Hence if the ν integrals are greater than or comparable to μ^2, e.g.,

$$\nu^{\rm dec}, \nu^{\rm inc} \geqslant \mu^2/2 , \tag{2.9}$$

we obtain

$$F_{12}(z,z_0) = \exp(-2iw_0)\tfrac{1}{2}O_{12}(\nu^{\rm dec}) , \tag{2.10a}$$

$$F_{21}(z,z_0) = \exp(2iw)\tfrac{1}{2}O_{21}(\nu^{\rm inc}) . \tag{2.10b}$$

Here and in what follows the O symbols are used in a nonstandard sense, which is convenient for our purposes (accurate error estimates). Thus, in general, for any complex c, the $O(c)$ denotes a complex quantity satisfying

$$|O(c)| \leqslant |c|(1 + d), \quad 0 < d \ll 1 . \tag{2.11}$$

[$O(c)$ is "practically" bounded by $|c|$.]

It should be pointed out that in typical situations discussed in Ref. 1 (Sec. VI) the μ integrals and the nonvanishing ν integrals satisfy the requirement (2.9). (The ν integrals which vanish in the limit $x \to B$ never appear in our analysis given in Secs. IV and V.) If condition (2.9) is not fulfilled, the estimates (2.10) require the replacement $\nu^{\rm dec} \to \mu^2/2$ or $\nu^{\rm inc} \to \mu^2/2$.

In examples given in Ref. 1 one obtains $\nu^{\rm dec}, \nu^{\rm inc} \ll \mu$ if $\mu \ll 1$. This indicates that the relevant estimates of Ref. 3, which can be obtained from Eqs. (2.10) by replacing $\nu^{\rm dec}, \nu^{\rm inc} \to \mu$, involve a noticeable overestimation. The reason for this is somewhat different for the Stokes lines ($\mathrm{Re}\, w_1 = \text{const}$), and the anti-Stokes lines ($\mathrm{Im}\, w_1 = \text{const}$). If Λ coincides with an anti-Stokes line, the exponential factors in Eqs. (2.8) are of modulus one but they are oscillating. For a Stokes line these factors are real and positive but exponentially small except for an immediate vicinity of z_0 or z.

For the remaining matrix elements F_{11} and F_{22} we can use the estimates of Ref. 3, which for $\mu \ll 1$ take the form

$$F_{11}(z,z_0) = 1 + \tfrac{1}{2}O_{11}(\mu) , \tag{2.12a}$$

$$F_{22}(z,z_0) = 1 + \tfrac{1}{2}O_{22}(\mu) + \tfrac{1}{4}M \exp[2i(w - w_0)]O_2(\mu^2) , \tag{2.12b}$$

where $M \leqslant 1$. These estimates are realistic except for the last term in Eq. (2.12b) which is obtained if the exponential factors analogous to those in Eqs. (2.8) are replaced by unity. However, the resultant overestimation is of little relevance to our analysis. If Λ is an anti-Stokes line (Sec. V), the O_2 term is insignificant [$\mathrm{Im}(w - w_0) = 0$]. For the bound states (Sec. IV) it will appear explicitly but only in the situation in which the nonvanishing ν integral is comparable to the μ integral [for N in Eq. (2.3) corresponding to the optimum approximation order[1]]. In that case the overestimation in the O_2 term should not be serious.

To discuss the consequences of the estimates (2.10) and (2.12) it is convenient to write

$$f_{1,2}'(z) = D_{1,2}(z)f_{1,2}(z) , \tag{2.13a}$$

where

$$D_{1,2}(z) = D_0(z) \pm iq(z) , \tag{2.13b}$$

$$D_0(z) = -(2q)^{-1}\frac{dq}{dz} . \tag{2.13c}$$

With this notation Eqs. (2.4) can be written in the equivalent form which will be used in Sec. V,

$$a_{1,2}(z)f_{1,2}(z) = \mp i(2q)^{-1}[\psi'(z) - D_{2,1}(z)\psi(z)] . \tag{2.14}$$

We assume that

$$\left|\frac{D_0(z)}{q(z)}\right| \equiv \left|\frac{1}{2}\frac{d}{dz}q^{-1}\right| \ll 1 , \tag{2.15}$$

which means that the main behavior of the functions $f_{1,2}(z)$ is given by the $\exp(\pm iw)$ rather than $q^{-1/2}(z)$. Equations (2.13b) and (2.15) lead to

$$D_2(z)/D_1(z) \simeq -1 . \tag{2.16}$$

Using Eqs. (2.4), (2.10), (2.12), and (2.16) we obtain

$$\psi(z) = a_1(z_0)f_1(z)\,[1 + \tfrac{1}{2}O_{11}(\mu) + \tfrac{1}{2}O_{21}(\nu^{\rm inc})]$$
$$+ a_2(z_0)f_2(z)\{1 + \tfrac{1}{2}O_{22}(\mu) + g[\tfrac{1}{2}O_{12}(\nu^{\rm dec})$$
$$+ \tfrac{1}{4}MO_2(\mu^2)]\}\,, \qquad (2.17a)$$
$$\psi'(z) = a_1(z_0)f_1'(z)\,[1 + \tfrac{1}{2}O_{11}(\mu) + \tfrac{1}{2}\overline{O}_{21}(\nu^{\rm inc})]$$
$$+ a_2(z_0)f_2'(z)\{1 + \tfrac{1}{2}O_{22}(\mu) + g[\tfrac{1}{2}\overline{O}_{12}(\nu^{\rm dec})$$
$$+ \tfrac{1}{4}MO_2(\mu^2)]\}\,, \qquad (2.17b)$$

where

$$\overline{O}_{21} = O_{21}D_2(z)/D_1(z) \simeq -O_{21},$$
$$\overline{O}_{12} = O_{12}D_1(z)/D_2(z) \simeq -O_{12}\,, \qquad (2.17c)$$

and

$$g = \exp[2i(w - w_0)], \quad |g| \geqslant 1\,. \qquad (2.17d)$$

Equations (2.17) are valid if conditions (2.7) and (2.15) are fulfilled, and the points z_0 and z can be joined by a path along which $|\exp(iw_1)|$ increases. The O terms in Eqs. (2.17) represent the errors of the phase-integral approximation. An important point is that the O terms in Eq. (2.17a) are either the same or are simply related to those in Eq. (2.17b). Due to this fact the relatively large error terms $[O_{11}(\mu), O_{22}(\mu)$, and $O_{12}(\nu^{\rm dec})]$ can often be eliminated from the ratio $\psi'(z)/\psi(z)$; see the following Eqs. (2.19b) and (2.20a) in which the eliminated terms only influence the irrelevant details of the O symbols.

If Λ is not an anti-Stokes line ($|g| \neq 1$), and z is sufficiently far away from z_0, as measured by $|\mathrm{Im}(w - w_0)|$, we obtain $|g| \gg 1$. In that case the accuracy of the phase-integral approximation depends crucially on

$$r \equiv \frac{a_2(z_0)f_2(z)}{a_1(z_0)f_1(z)} = \frac{r_0}{g} \qquad (2.18)$$

$[r_0 = r(z = z_0)]$, the ratio of the decreasing to the increasing phase-integral term [these names refer to the exponential factors in the $f_{1,2}(z)$ along Λ]. If this ratio is not too large at z_0, so that

$$|r_0|\nu^{\rm dec} \ll 1\,, \qquad (2.19a)$$

but small at z, $|r| \ll 1$, which means that the increasing term represents the dominant contribution at z, we easily find using Eqs. (2.17a) and (2.17b),

$$\psi'(z)/\psi(z) = D_1(z)\{1 + O(\nu^{\rm inc})$$
$$+ |r_0|[\tfrac{1}{2}MO(\mu^2) + 2O(g^{-1})]\}\,. \qquad (2.19b)$$

This means that $\psi'(z)/\psi(z)$ is given, with high accuracy, by $f_1'(z)/f_1(z)$ [see Eq. (2.13a)]. A different behavior is found from Eqs. (2.17a), (2.17b) if there is no increasing term at z_0 $[a_1(z_0) = 0;\ |g| \neq 1]$,

$$\psi'(z)/\psi(z) = D_2(z)[1 + |g|O(\nu^{\rm dec})]\,, \qquad (2.20a)$$

valid if

$$|g|\nu^{\rm dec} \ll 1\,. \qquad (2.20b)$$

Condition (2.20b) is very restrictive in view of the exponential growth of $|g|$. Therefore $\psi'(z)/\psi(z)$ can be described by $f_2'(z)/f_2(z)$ at an infinitesimal distance from z_0 only; the accuracy very soon becomes intolerably high.

If Λ is an anti-Stokes line ($|g| = 1$), we obtain $\nu^{\rm inc} \simeq \nu^{\rm dec}$. In that case Eqs. (2.17a) and (2.17b) become symmetric in the subscripts 1 and 2 if we introduce

$$\widetilde{O}_{22}(\mu) = O_{22}(\mu) + \tfrac{1}{8}gMO_2(\mu^2)\,,$$
$$\widetilde{O}_{12}(\nu^{\rm dec}) = gO_{12}(\nu^{\rm dec})\,.$$

If there is only one phase-integral term at z_0, the $O(\mu)$ error term can again be eliminated from $\psi'(z)/\psi(z)$. [For example, if $a_2(z_0) = 0$, $\psi'(z)/\psi(z)$ is given by Eq. (2.19b) with $r_0 = 0$.] Another example in which the $O(\mu)$ terms are eliminated will be given in Sec. V in connection with the reflection, transmission, and absorption coefficients. In that case both phase-integral terms are present, but the anti-Stokes line in question is a segment of the real axis.

B. Application to $q(z)$ given by Eq. (2.3)

If the $q(z)$ is given by Eq. (2.3), the definition (2.13c) leads to

$$D_0(z) = -\frac{1}{2}\left\{\frac{1}{2}Q^{-2}\frac{d}{dz}Q^2 + \left[\sum_{n=0}^{N} Y_{2n}(z)\right]^{-1}\frac{d}{dz}\sum_{n=1}^{N} Y_{2n}(z)\right\}. \qquad (2.21)$$

Introducing

$$\zeta = \int Q(z)dz\,, \qquad (2.22)$$

with any convenient choice of the integration constant, and assuming that the higher-order corrections included in (2.3) are small,

$$|Y_{2n}(z_1)| \ll 1 \quad \text{along } \Lambda, \quad 1 \leqslant n \leqslant N, \qquad (2.23)$$

we can approximate dw_1 in Eqs. (2.6a) and (2.8) by $d\zeta_1$. Furthermore the leading term in the μ integral (2.6a) can be expressed in terms of the first truncated correction in Eq. (2.3) [see Eq. (28) in Ref. 4], which finally leads to

$$\mu \equiv \mu_{2N+1} \simeq \int_{\zeta_0}^{\zeta} |d\zeta_1\, 2Y_{2N+2}(\zeta_1)|\,, \qquad (2.24)$$

$$\nu^{\rm dec} \equiv \nu^{\rm dec}_{2N+1} \simeq \left|\int_{\zeta_0}^{\zeta} d\zeta_1\, 2Y_{2N+2}(\zeta_1)\exp[-2i(\zeta_1 - \zeta_0)]\right|\,, \qquad (2.25a)$$

$$\nu^{\rm inc} \equiv \nu^{\rm inc}_{2N+1} \simeq \left|\int_{\zeta_0}^{\zeta} d\zeta_1\, 2Y_{2N+2}(\zeta_1)\exp[2i(\zeta_1 - \zeta)]\right|\,. \qquad (2.25b)$$

Condition (2.15) takes the form

$$\frac{1}{2}\left|\frac{d}{dz}Q^{-1} - \left(\sum_{n=0}^{N} Y_{2n}\right)^{-1}\frac{d}{d\zeta}\sum_{n=1}^{N} Y_{2n}\right| \ll \left|\sum_{n=0}^{N} Y_{2n}\right|, \qquad (2.26)$$

which in the first order ($N = 0$) simplifies to

$$\delta(z) \equiv \left|\frac{1}{2}\frac{d}{dz}Q^{-1}\right| \ll 1\,. \qquad (2.27)$$

If $Q^2(z) = R(z)$, Eq. (2.27) reduces to the well-known validity condition for the quasiclassical approximation of quantum mechanics.

III. INTRODUCTION TO APPLICATIONS

To apply the error estimates derived in Sec. II, a proper choice of the base function $Q(x)$ and the integration path Λ is needed. In the following two sections Λ will be a segment of the real axis:

$$a<x<x_a \quad \text{or} \quad x_b<x<b. \tag{3.1}$$

Quantities needed in applications, such as the leading terms of the $Y_{2n}(x)$, the μ integrals, and the nonvanishing ν integrals, are given in Ref. 1 (Sec. VI). These results become directly applicable to the intervals (3.1) if we replace in the relevant formulas of Ref. 1

$$\begin{aligned}&(z \text{ or } x)\to x-B, && \text{for } B\neq\pm\infty,\\ &(\tilde z \text{ or } \tilde x)\to x, && \text{for } B=\pm\infty,\end{aligned} \tag{3.2}$$

where B denotes either a or b.

If $R(x)$ satisfies Eq. (1.2), the results of Ref. 1 indicate that $Q^2(z)$ can always be chosen so that (i) the μ integrals (2.24) are convergent as $x\to B$; (ii) the corrections $Y_{2n}(x)$ vanish as $x\to B$,

$$\lim_{x\to B}\left(Y_{2n},\frac{d}{d\zeta}Y_{2n},\dots\right)=0,\quad n=1,2,\dots; \tag{3.3}$$

and (iii) $\delta(x)$ is small in the intervals (3.1),

$$\lim_{x\to B}\delta(x) \text{ is either zero or a constant } (\ll 1). \tag{3.4}$$

With this choice of $Q^2(x)$ the validity conditions of Sec. II [Eqs. (2.7), (2.23), and (2.26)] and the following requirements (3.5) can always be fulfilled by choosing the shifted boundaries x_B close enough to the actual boundaries B.

The following analysis assumes that both $R(x)$ and $Q^2(x)$ are real, which implies the reality also of the $q^2(x)$ given by Eq. (2.3).[4] If Eq. (1.2) is fulfilled with a strict inequality, one can work with either the commonly behaving phase-integral approximation, or the optimized approximation. In either case the leading term of $Q^2(x)$ is the same as that for $R(x)$. If Eq. (1.2) is fulfilled with a strict equality, only the optimized approximation leads to the properties (i)–(iii); $R(x)$ and $Q^2(x)$ are given by Eqs. (5.10) and (5.8) of Ref. 1, in which $|c_R|\gg\frac{1}{4}$. In that case the leading term of $Q^2(x)$ is slightly different from that for $R(x)$ ($c=c_R-\frac{1}{4}$). Nevertheless, the main behavior of $Q^2(x)$ as $x\to B$ is always given by the behavior of $R(x)$. Therefore we can assume that $Q^2(x)$ has a fixed sign in the intervals (3.1), which furthermore coincides with the sign of $q^2(x)$,

$$R(x),Q^2(x),q^2(x)\begin{cases}<0, & \text{for bound states, Sec. IV,} \quad (3.5a)\\ >0, & \text{for wave propagation, Sec. V.} \quad (3.5b)\end{cases}$$

This means that Λ will be either a Stokes line [Eq. (3.5a)] or an anti-Stokes line [Eq. (3.5b)].

It should be pointed out that the choice of quantities such as $Q^2(x)$, the phase of $q^2(x)$, and the lower limit of integration z_l in Eq. (2.2) can be made independently for each interval (3.1). In particular, it will be convenient to take $z_l=x_B$, leading to

$$f_{1,2}(x_B)=q^{-1/2}(x_B),\quad x_B=x_a,x_b. \tag{3.6}$$

With this choice Eqs. (2.4) and (2.14) take a simple form at $z=x_B$, and $w(x)$ is pure imaginary on the Stokes lines (3.5a), and real on the anti-Stokes lines (3.5b).

IV. APPLICATION TO BOUND STATES

Limiting properties of the functions $a_{1,2}(x)$ pertaining to the wave function of a bound state are given in Ref. 3, Chap. 10. These results are applicable to the $q(x)$ given by Eq. (2.3), and $x\to B$, due to the assumed convergence of the μ integral at B, and Eq. (3.5a). Choosing the sign of $Q(x)$ so that $\exp[iw(x)]$ increases as x moves from the actual boundary B toward the shifted boundary x_B (the opposite choice as in Ref. 3), the results of Ref. 3 indicate that

$$\lim_{x\to B}a_1(x) \text{ exists and is finite,}\quad \lim_{x\to B}a_2(x)=0, \tag{4.1}$$

and

$$\lim_{x\to B}\psi(x)=0. \tag{4.2}$$

Using Eq. (4.1) and (2.19b) with $z_0=B$ we obtain ($r_0=0$),

$$[\psi'(x)/\psi(x)]_{x_B}=D_1(x_B)\{1+O_1[(\nu_B)_{2N+1}]\}, \tag{4.3}$$

where $(\nu_B)_{2N+1}\equiv\nu_{2N+1}^{\rm inc}(z_0=B,z=x_B)\ll 1$. Equation (4.3) defines the boundary condition at the shifted boundary x_B.

In the bound state calculations it is often convenient to impose the homogeneous boundary condition (4.2) but at some shifted boundary $\bar x_B$ (see for example Ref. 2 and the literature cited therein). This leads to an error in the wave function, called a *termination error*,[2] which can be estimated by using the results of Sec. II. Choosing $\bar x_B$ to be located between B and x_B, and denoting by $\bar\psi(x)$ the wave function vanishing at $\bar x_B$,

$$\bar\psi(\bar x_B)=0, \tag{4.4}$$

Eq. (2.19b) yields ($z_0=\bar x_B$, $z=x_B$, $r_0=-1$),

$$[\bar\psi'(x)/\bar\psi(x)]_{x_B}=D_1(x_B)[1+O(\bar\nu_{2N+1}^{\rm inc})+\tfrac{1}{2}MO(\bar\mu_{2N+1}^2)+2O(|g|^{-1})], \tag{4.5a}$$

where

$$|g|=\exp[2|w(\bar x_B)-w(x_B)|]\gg 1, \tag{4.5b}$$

and the integrals $\bar\nu$ nd $\bar\mu$ correspond to $z_0=\bar x_B$ and $z=x_B$. The error terms in Eqs. (4.3) and (4.5a) are related to each other. It can be shown by introducing the matrices $\mathbf{F}(\bar x_B,B)$ and $\mathbf{F}(x_B,\bar x_B)$ and using the estimates (2.10) and (2.12) that $|O_1[(\nu_B)_{2N+1}]-O(\bar\nu_{2N+1}^{\rm inc})|$ is small as compared to the last two O terms in Eq. (4.5a), which leads to [$M=\frac{1}{2}$ for the Stokes line,[3] and $\bar\mu_{2N+1}\leqslant(\mu_B)_{2N+1}\equiv\mu_{2N+1}(z_0=B, z=x_B)$]

$$\left.\frac{\bar\psi'(x)}{\bar\psi(x)}\right|_{\bar x_B} = \left.\frac{\psi'(x)}{\psi(x)}\right|_{x_B}\left\{1+\frac{1}{4}O\,[(\mu_B)^2_{2N+1}]+2O(|g|^{-1})\right\}. \tag{4.6}$$

The O terms in Eq. (4.6) represent the relative error arising in the ratio $\psi'(x)/\psi(x)$ at $x=x_B$ if the boundary condition (4.2) is shifted from B to $\bar x_B$. This error is not directly related to the error arising in the energy eigenvalue. However, if the wave function $\bar\psi(x)$ is determined numerically, some estimates for the numerical error in both $\bar\psi(x)$ and the energy eigenvalue can usually be given. They will define the overall accuracy of the bound state calculations if the O terms in Eq. (4.6) are small in comparison to ϵ_{num}, the relative error arising in $\bar\psi'(x)/\bar\psi(x)$ at $x=\bar x_B$ due to the numerical approximations:

$$\tfrac{1}{4}(\mu_B)^2_{2N+1} \ll K_B^{-1}\epsilon_{\text{num}}, \tag{4.7a}$$

$$2|g|^{-1} \ll K^{-1}\epsilon_{\text{num}}, \tag{4.7b}$$

where K_B and K are constants greater than unity. Introducing $\zeta_M(x)$ defined by Eq. (2.22) but with $Q(x)$ replaced by its leading term as $x\to B$, $Q_M(x)$, and approximating $w(x)\simeq\zeta_M(x)+\text{const}$, Eqs. (4.5b) and (4.7b) lead to

$$|\zeta_M(\bar x_B)| = |\zeta_M(x_B)| + \tfrac{1}{2}\ln(2K\epsilon_{\text{num}}^{-1}). \tag{4.8}$$

If $(\mu_B)_{2N+1}\to 0$ as $N\to\infty$, Eq. (4.7a) is fulfilled automatically in the limit $N\to\infty$. In that case x_B in Eq. (4.8) is arbitrary, subject only to the requirement that the model $Q_M^2(x)$ is not too bad at x_B. This in particular is applicable to the $R(x)$ having a second-order pole at $B\neq\pm\infty$, $R(x)\simeq c_R(x-B)^{-2}$, or to $R(x)\simeq c_R x^{-2}$, for $B=\pm\infty$ [see Eqs. (5.10) and (6.10a) in Ref. 1 ($c_R\ll-\frac{1}{4}$)]. In these cases Eq. (4.8) leads to

$$\left|\frac{x_B-B}{\bar x_B-B}\right|_{B\neq\pm\infty} = \left|\frac{\bar x_B}{x_B}\right|_{B=\pm\infty} = (2K\epsilon_{\text{num}}^{-1})^{(1/2)|c|^{-1/2}}, \tag{4.9}$$

where $c=c_R-\frac{1}{4}$. Another possibility is the $R(x)$ tending exponentially to a negative constant as $x\to\pm\infty$ (bound states in the potential tending to zero exponentially at infinity) [see Eqs. (4.18) and (6.18a) in Ref. 1]. In that case Eq. (4.8) leads to

$$|\bar x_B| = |x_B| + L\,|2\gamma|^{-1}\ln(2K\epsilon_{\text{num}}^{-1}), \tag{4.10}$$

where x_B must satisfy $|\gamma_e|\exp[-|\text{Re}(\eta x_B/L)|]\ll 1$.

In the remaining cases discussed in Ref. 1 [Secs. VI and VII] $(\mu_B)_{2N+1}$ reaches a minimum at some $N=N_0$. Choosing $K_B=2|\zeta_M(x_B)|$ and $N=N_0$, Eqs. (4.7) yield

$$|\zeta_M(\bar x_B)| = \tfrac{1}{2}\ln(KC\epsilon_{\text{num}}^{-3/2}), \quad K>1, \tag{4.11}$$

where C is given by Eq. (7.10) in Ref. 1. Equation (4.11) is valid if $\epsilon_{\text{num}}<(C/2e)^2$, which is a consequence of the required $K_B>1$; $|\zeta_M(x)|$ is either a power or an exponential function of x [see Eqs. (4.5a) and (4.28) in Ref. 1].

Equations (4.8)–(4.11) define the shifted boundary $\bar x_B$ in terms of the numerical error ϵ_{num}.

The results of this section indicate that the approach using the boundary condition (4.4) is both simple and efficient. It introduces automatically the optimum accuracy of the phase-integral approximation ($N\to\infty$ or $N=N_0$), which could never be achieved by working directly with the boundary condition (4.3) after deleting the O_1 term.

V. APPLICATION TO WAVE PROPAGATION

To give a general definition of the reflection, transmission, and absorption coefficients, applicable to any $R(x)$ satisfying Eqs. (1.2) and (3.5b), one can follow the general approach of Ref. 3, Chap. 9. An essential point is to use Eqs. (2.4) along with (2.1)–(2.3) to decompose $\psi(x)$ into the purely propagating waves as $x\to B$ ($=a,b$). In view of the convergence of the μ integrals at B and Eq. (3.5b), the four limits are well defined,[3]

$$a_k(B) = \lim_{x\to B} a_k(x)\neq\infty, \quad B=a,b, \quad k=1,2. \tag{5.1}$$

Equation (5.1) and the vanishing of the corrections $Y_{2n}(x)$ at B [Eq. (3.3)] mean that the two terms in Eq. (2.4), $a_k(x)f_k(x)$, $k=1,2$, tend to the exact solutions of Eq. (1.1) as $x\to B$. In view of Eq. (2.14) these terms are independent of the lower limit of integration z_l in Eq. (2.2). Furthermore their dependence on N and on the choice of $Q^2(x)$ disappears in the limit $x\to B$. This is a consequence of Eqs. (3.3) and (3.4), and of the fact that the leading term of the admissible $Q^2(x)$ as $x\to B$ is uniquely defined by the leading term of $R(x)$. For example, using Eq. (3.3) it can be shown that

$$\lim_{x\to B}\frac{a_k(x)f_k(x)\,|_{N>0}}{a_k(x)f_k(x)\,|_{N=0}} = 1, \quad k=1,2, \tag{5.2}$$

if $a_k(B)|_{N=0}\neq 0$; otherwise one obtains $a_k(B)=0$ for any $N\geqslant 0$. Equation (5.2) assumes a definite choice of $Q^2(x)$, and a similar behavior is obtained if $Q^2(x)$ rather than N is modified (for example, for $N=0$). This indicates that the products $a_k(x)f_k(x)$, $k=1,2$, are uniquely defined by the solution $\psi(x)$ as $x\to B$, and can be interpreted in this limit as propagating waves. In the following analysis we assume the time dependence $\exp(-i\omega t)$, $\omega>0$, and choose the phase of $q^2(x)$ so that $q(x)>0$. In that case the wave $a_1(x)f_1(x)$ propagates in the positive x direction, and $a_2(x)f_2(x)$ in the negative x direction.

The assumed reality of $R(x)$ leads to the conservation law[3]

$$S(x) = \text{Im}[\psi'(x)\psi^*(x)] = \text{const}, \tag{5.3}$$

where $S(x)$ is proportional to the quantum mechanical current, or to the energy associated with the electromagnetic wave. In the latter case the $R(x)$ may have simple poles (resonances) at $x=x_r$; see Appendix A for the details. Assuming that the wave energy is absorbed at the resonance, we obtain

$$S(x_r+0) - S(x_r-0) \leqslant 0. \tag{5.4}$$

Equation (5.4) is applicable both to the propagation in the positive x direction ($S\geqslant 0$), and in the negative x direction ($S\leqslant 0$). In any case it means a drop in $|S(x)|$ at x_r when moving in the direction of propagation. Note that the shifted boundaries must be chosen so as to enclose all resonances [see Eq. (3.5b)]. Using the fact that $w(x)$ is real (Sec. III), and $q(x)>0$, Eq. (9.5) in Ref. 3 leads to

$$S(x) = \text{const} = |a_1(x)|^2 - |a_2(x)|^2, \tag{5.5}$$

where x belongs to the intervals (3.1). Equation (5.5) indicates that $|a_k(B)|^2$, $k = 1,2$, can be interpreted as the absolute value of the energy associated with the wave $a_k(x) f_k(x)$ as $x \to B$. This leads to the following definition of the reflection (R), transmission (T), and absorption (A) coefficients, pertaining to the propagation in the positive x direction,

$$a_2(b) = 0 \quad (\text{no reflection at } b), \tag{5.6a}$$

$$R = |a_2(a)|^2 |a_1(a)|^{-2}, \tag{5.6b}$$

$$T = |a_1(b)|^2 |a_1(a)|^{-2}, \tag{5.6c}$$

$$A = [\,|S(a)| - |S(b)|\,] |a_1(a)|^{-2}. \tag{5.6d}$$

Interchanging here a with b and the subscripts 1 with 2, the definitions for the propagation in the negative x direction are obtained. In any case they lead to

$$R + T + A = 1, \tag{5.7}$$

where $A \geqslant 0$ if condition (5.4) is fulfilled. If there are no resonances, Eq. (5.6d) leads to $A = 0$.

To derive the approximate expressions for R, T, and A, and the accurate error estimates for these formulas, we introduce the F matrices connecting the actual boundaries with the shifted boundaries,

$$\mathbf{F}_B = \mathbf{F}(x_B, B) \quad B = a,b. \tag{5.8}$$

Using Eqs. (6.1) and (6.2) of Ref. 3, and our estimates (2.10b) and (2.12a), we obtain

$$(\mathbf{F}_B)_{11} = (\mathbf{F}_B)^*_{22} = 1 + O(\mu_B)/2, \tag{5.9a}$$

$$(\mathbf{F}_B)_{21} = (\mathbf{F}_B)^*_{12} = O(\nu_B)/2, \tag{5.9b}$$

$$|(\mathbf{F}_B)_{11}|^2 = |(\mathbf{F}_B)_{22}|^2 = 1 + O(\nu_B^2)/4, \tag{5.9c}$$

where $\mu_B \equiv \mu_{2N+1}$ and $\nu_B \equiv \nu^{\text{inc}}_{2N+1}$ correspond to $z_0 = B$, and $z = x_B$. We introduce the column vector $\boldsymbol{\psi}(z)$ with components $\psi(z)$ and $\psi'(z)$, and the matrix $\mathbf{M}(z,z_0)$ which propagates $\boldsymbol{\psi}(z)$ from a given point z_0,

$$\boldsymbol{\psi}(z) = \mathbf{M}(z,z_0)\boldsymbol{\psi}(z_0). \tag{5.10}$$

This M matrix can be expressed in terms of two solutions of Eq. (1.1), $\psi_1(z)$, and $\psi_2(z)$, which satisfy simple boundary conditions at z_0,

$$\mathbf{M}(z,z_0) = \begin{bmatrix} \psi_1(z) & \psi_2(z) \\ \psi_1'(z) & \psi_2'(z) \end{bmatrix}, \qquad \mathbf{M}(z_0,z_0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.11}$$

This implies

$$\det \mathbf{M}(z,z_0) = \text{Wronskian}(\psi_1,\psi_2) = 1. \tag{5.12}$$

[The matrix satisfying Eqs. (5.10)–(5.12) has been introduced recently also by Kowalski and Fry,[5] and we are using their notation.] Denoting for simplicity

$$\mathbf{M} = \mathbf{M}(x_a, x_b), \tag{5.13}$$

the matrix elements M_{11}, etc., can be determined by integrating Eq. (1.1) numerically from $z_0 = x_b$ to $z = x_a$. If there are no resonances one can integrate along the real axis, and $\mathbf{M}$ is real. Otherwise the integration path can be split into small segments bypassing the resonances in the complex plane, and the remaining real segments. The matrix $\mathbf{M}$ can then be determined as a product of the partial M matrices corresponding to this division. In any case $\mathbf{M}$ will be treated here as known from the numerical calculations.

Using Eq. (5.6a) we start from calculating $\mathbf{a}(x_b)$,

$$a_1(x_b) = (\mathbf{F}_b)_{11} a_1(b), \quad a_2(x_b) = (\mathbf{F}_b)_{21} a_1(b).$$

Then we find $\boldsymbol{\psi}(x_b)$ from Eqs. (2.4) and (3.6), and propagate it to x_a, $\boldsymbol{\psi}(x_a) = \mathbf{M}\boldsymbol{\psi}(x_b)$. And finally we find $\mathbf{a}(x_a)$ from Eqs. (2.14) and (3.6), and propagate it to a,

$$\mathbf{a}(a) = \begin{bmatrix} (\mathbf{F}_a)_{22} & -(\mathbf{F}_a)_{12} \\ -(\mathbf{F}_a)_{21} & (\mathbf{F}_a)_{11} \end{bmatrix} \mathbf{a}(x_a).$$

Using this result in the definitions (5.6b) and (5.6c), and calculating A from (5.7), the final results can be written

$$R = R_{\text{num}}[1 + O(\epsilon_R)], \tag{5.14a}$$

$$T = T_{\text{num}}[1 + O(\epsilon_T)], \tag{5.14b}$$

$$A = A_{\text{num}} + R_{\text{num}} O(\epsilon_R) + T_{\text{num}} O(\epsilon_T), \tag{5.14c}$$

where

$$R_{\text{num}} = |n_1 + n_2|^2 |d_1 + d_2|^{-2}, \tag{5.15a}$$

$$T_{\text{num}} = 4q(x_a) q(x_b) |d_1 + d_2|^{-2}, \tag{5.15b}$$

$$A_{\text{num}} = 1 - R_{\text{num}} - T_{\text{num}}, \tag{5.15c}$$

and

$$\begin{aligned} n_1 &= M_{21} - D_1(x_a) M_{11}, \\ n_2 &= D_1(x_b)[M_{22} - D_1(x_a) M_{12}], \\ d_1 &= M_{21} - D_2(x_a) M_{11}, \\ d_2 &= D_1(x_b)[M_{22} - D_2(x_a) M_{12}]. \end{aligned} \tag{5.16}$$

The O symbols are defined by Eq. (2.11); ϵ_R and ϵ_T contain the linear and the quadratic terms in ν_B, but are independent of μ_B [see the following Eqs. (5.22)]. As in Eqs. (2.19b) and (2.20a) the μ integrals only influence the irrelevant details of the O symbols. Equations (5.15) define the numerical approximations to R, T, and A. The same expressions are obtained if the actual boundaries a and b in the definitions (5.6) are replaced by the shifted boundaries x_a and x_b.

It can be seen from the above derivation that on interchanging in Eqs. (5.14)–(5.16) of D_1 with D_2 and x_a with x_b, and replacing

$$\mathbf{M} \to \mathbf{M}^{-1} = \begin{bmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{bmatrix}, \tag{5.17}$$

we obtain the results corresponding to the propagation in the negative x direction. They will be denoted by a bar, e.g.,

$$\bar{R} = \bar{R}_{\text{num}}[1 + O(\bar{\epsilon}_R)], \tag{5.18}$$

where

$$\bar{R}_{\text{num}} = |\bar{n}_1 + \bar{n}_2|^2 |\bar{d}_1 + \bar{d}_2|^{-2}, \tag{5.19}$$

and

$$\begin{aligned} \bar{n}_1 &= -M_{21} - D_2(x_b) M_{22}, \\ \bar{n}_2 &= D_2(x_a)[M_{11} + D_2(x_b) M_{12}], \\ \bar{d}_1 &= -M_{21} - D_1(x_b) M_{22}, \\ \bar{d}_2 &= D_2(x_a)[M_{11} + D_1(x_b) M_{12}]. \end{aligned} \tag{5.20}$$

Equations (5.15b) and (B4) indicate that

$$\bar{T}_{\text{num}} = T_{\text{num}}, \tag{5.21}$$

which in the limit $x_B \to B$ leads to $\bar{T} = T$. Thus the transmission coefficient is independent of the direction of propagation; $\bar{A}$ and $\bar{A}_{\text{num}}$ are given by Eqs. (5.14c) and (5.15c) specialized to the quantities with a bar ($\bar{\epsilon}_T = \epsilon_T$).

Using the relations given in Appendix B, the quantities ϵ_T, etc., representing the relative errors in T, R, and $\bar{R}$ can be written

$$\epsilon_T = \bar{R}_{\text{num}}^{1/2} \nu_b + R_{\text{num}}^{1/2} \nu_a + \tfrac{1}{2} H \nu_a \nu_b, \tag{5.22a}$$

$$\epsilon_R = \epsilon_T + R_{\text{num}}^{-1/2} (H \nu_b + \nu_a), \tag{5.22b}$$

$$\bar{\epsilon}_R = \epsilon_T + \bar{R}_{\text{num}}^{-1/2} (H \nu_a + \nu_b), \tag{5.22c}$$

where

$$H = |n_1 + n_2 D_2(x_b)/D_1(x_b)|\, |d_1 + d_2|^{-1}. \tag{5.23}$$

The shifted boundaries x_B can be matched to the required accuracy by choosing them so that

$$\nu_a = \nu_b \equiv \nu. \tag{5.24}$$

For given ν, which defines the relative error in T, Eqs. (5.24) can be solved analytically for $x_a(\nu)$ and $x_b(\nu)$ in all cases discussed in Ref. 1 (Sec. VI).

Equations (5.22b) and (5.22c) indicate that R_{num} and $\bar{R}_{\text{num}}$ are only meaningful if they are large enough in comparison to their critical values corresponding to $\epsilon_R, \bar{\epsilon}_R \sim 1$. If Eqs. (5.24) are fulfilled, we obtain

$$[R_{\text{num}}, \bar{R}_{\text{num}}]_{\text{crit}} \sim (1 + H)^2 \nu^2. \tag{5.25}$$

Note that our assumption (5.4) has not yet been used. Therefore Eqs. (5.14)–(5.25) are valid also if condition (5.4) is violated at some poles (which can lead to $A < 0$ or $\bar{A} < 0$). However, if condition (5.4) is fulfilled, all coefficients in question (R, $\bar{R}$, T, A, and $\bar{A}$), and their numerical approximations, are non-negative, and therefore bounded by zero and unity. Furthermore a simple estimate for H can be given (see Appendix B),

$$H \leqslant [1 + \min(R_{\text{num}}, \bar{R}_{\text{num}})]^{1/2} < 2^{1/2}. \tag{5.26}$$

If Eq. (5.26) holds, the H term in Eq. (5.22a) is insignificant and can be deleted.

The theory developed in this section will now be illustrated by two examples which were discussed in Ref. 6. In the present notation they correspond to

$$R(x) = R(\infty) - [R(\infty) - R(0)](x^2 + 1)^{-1}, \tag{5.27}$$

where $R(\infty) > 0$ and $R(0) < R(\infty)$, and

$$R(x) = K^2 (x^2 + 1)^2 x (x - x_r)^{-1}, \quad K > 0, \quad x_r > 0. \tag{5.28}$$

In both cases $-\infty < x < \infty$ and $R(x)$ can be approximated, as $x \to \pm\infty$, by the power model $\tilde{c} x^{\tilde{m}}$ [$\tilde{c} = R(\infty)$, $\tilde{m} = 0$ for Eq. (5.27), and $\tilde{c} = K^2$, $\tilde{m} = 4$ for Eq. (5.28); a tilde is used here in view of $B = \pm\infty$, a convention adopted in Ref. 1]. As $\tilde{m} > -2$, one can choose $Q^2(x) = R(x)$, and as $\tilde{m}$ is even, Eq. (5.24) leads to $x_a = -x_b$. To determine the ν integral we use either Eqs. (6.14) or (6.13) of Ref. 1, specialized to quantities with a tilde ($B = b$). Solving equation $(\nu_b)_{2N+1}(x_b) = \nu$ for x_b we obtain

$$x_b = \tfrac{1}{2} R^{-1/2}(\infty) \times \{4[R(\infty) - R(0)] 3 \cdot 4 \cdots (2N+3)/\nu\}^{1/(2N+4)}, \tag{5.29}$$

for Eq. (5.27) [$\tilde{p} = -2$, $\tilde{g}_p = R(0)/R(\infty) - 1$], and

$$x_b = \{3(2K)^{-1}[0.65(2N+1)!/\nu]^{1/(2N+2)}\}^{1/3}, \tag{5.30}$$

for Eq. (5.28) ($\beta_N \leqslant 1.3$, $|b| \leqslant \frac{1}{4}$). Equations (5.29) and (5.30) illustrate how the integration interval $(-x_b, x_b)$ shrinks as one increases the approximation order. On doing so one has to include the corrections $Y_{2n}(x)$ in $q(x)$, $D_1(x)$, and $D_2(x)$. However, as these quantities are only needed at the boundary points x_B, this complication has practically no influence on the computing time. A spectacular reduction in the computing time can be obtained in higher orders if ν is sufficiently small. This will be illustrated here by the drop in the computing time occurring in the third-order ($N = 1$) as compared to the first-order ($N = 0$). In a typical case of Eq. (5.27) [$R(\infty) = 100$, $R(0) = 0$] the observed drop was from 30 to 10 seconds for $\nu = 10^{-7}$, but from 130 to 20 seconds for $\nu = 10^{-10}$. The corresponding drop for Eq. (5.28) ($K = 10$, $x_r = 0.03$) was from 570 to 54 seconds for $\nu = 10^{-7}$, whereas for $\nu = 10^{-10}$ only the third-order calculation was feasible. It required 150 seconds, and the estimated increase in the computing time in the first order was by more than the factor of 100. All cases treated in Ref. 6 were recalculated by using the present (more accurate) error estimates, and the variations in R_{num}, etc., never exceeded the new error bounds. At the same time the differences between the results obtained in the first and in the third order were comparable to the error estimates. (In our earlier calculations[6] these differences were much smaller than the error estimates, and the computing times were at least twice as large as the present ones.) All calculations were performed on the CDC CYBER 73 computer, and the programs are obtainable from the author on request.

APPENDIX A: LOCAL SOLUTION AT A SIMPLE POLE OF $R(x)$

If Eq. (1.1) is continued to the complex z plane, and the $R(z)$ has a simple pole at $z = x_r$,

$$R(z) = c_r (z - x_r)^{-1} [1 + d_R(z)], \quad d_R(x_r) = 0, \tag{A1}$$

a general solution of this equation in some vicinity of x_r can be written as

$$\psi(z) = C_1 u_1(z) + C_2 u_2(z), \tag{A2}$$

where[7]

$$\begin{aligned} u_1(z) &= (z - x_r)[1 + d_1(z)], \quad d_1(x_r) = 0, \\ u_2(z) &= c_r u_1(z) \ln(z - x_r) - 1 - d_2(z), \quad d_2(x_r) = 0. \end{aligned} \tag{A3}$$

The functions $d_R(z)$ and $d_{1,2}(z)$ are analytic at x_r. If $R(x)$ is real, so are c_r, $d_R(x)$, and $d_{1,2}(x)$. Inserting in that case Eqs. (A2) and (A3) into the definition (5.3) we obtain

$$S(x_r + 0) - S(x_r - 0) = \pm c_r \pi |C_2|^2, \tag{A4}$$

where the upper sign corresponds to tracing the $\psi(z)$ in the upper half-plane. Thus to satisfy the absorption condition [Eq. (5.4)] the pole at x_r must be bypassed in the lower half-plane if $c_r > 0$, and in the upper half-plane if $c_r < 0$. This choice is unique [independent of $\psi(z)$] for given $R(z)$.

APPENDIX B: AUXILIARY RELATIONS

Equations (5.16) and (5.20) lead to the identities

$$n_1 + n_2 + \bar{d}_1 + \bar{d}_2 D_1(x_a)/D_2(x_a) = 0, \tag{B1}$$

$$\bar{n}_1 + \bar{n}_2 + d_1 + d_2 D_2(x_b)/D_1(x_b) = 0, \tag{B2}$$

$$n_1 + n_2 D_2(x_b)/D_1(x_b) + \bar{n}_1 + \bar{n}_2 D_1(x_a)/D_2(x_a) = 0, \tag{B3}$$

$$d_1 + d_2 + \bar{d}_1 + \bar{d}_2 = 0. \tag{B4}$$

Introducing $\bar{H}$ defined by Eq. (5.23) after the replacements mentioned in connection with Eq. (5.17), and using Eqs. (B3) and (B4), we obtain

$$\bar{H} = H. \tag{B5}$$

With our choice of the shifted boundaries (Sec. III) we can approximate [see Eq. (2.16)]

$$D_2(x_b)/D_1(x_b) \simeq -1, \quad D_1(x_a)/D_2(x_a) \simeq -1. \tag{B6}$$

Using Eqs. (B1), (B2), (B4), and (B6) in the definitions (5.15a) and (5.19) we obtain

$$\begin{aligned} R_{\text{num}}^{1/2} &\simeq |\bar{d}_1 - \bar{d}_2| \, |\bar{d}_1 + \bar{d}_2|^{-1}, \\ \bar{R}_{\text{num}}^{1/2} &\simeq |d_1 - d_2| \, |d_1 + d_2|^{-1}. \end{aligned} \tag{B7}$$

Equations (B7) lead to the estimates

$$\begin{aligned} [|\bar{d}_1| + |\bar{d}_2|]|\bar{d}_1 + \bar{d}_2|^{-1} &\leqslant (1 + R_{\text{num}})^{1/2}, \\ [|d_1| + |d_2|]|d_1 + d_2|^{-1} &\leqslant (1 + \bar{R}_{\text{num}})^{1/2}. \end{aligned} \tag{B8}$$

If the absorption condition (5.4) is fulfilled at all real poles x_r, we obtain

$$S(x_a) \geqslant S(x_b), \tag{B9}$$

for any solution $\psi(z)$. Specializing it to $\psi_1(z)$ and $\psi_2(z)$ [see Eq. (5.11)] with $z_0 = x_a$ or $z_0 = x_b$, we arrive at four inequalities:

$$|n_i| \leqslant |d_i|, \quad |\bar{n}_i| \leqslant |\bar{d}_i|, \quad i = 1,2. \tag{B10}$$

[$S(z_0) = 0$ in view of Eq. (5.3), whereas S at the other shifted boundary, x_b or x_a, can be determined from (5.5).] Equations (5.23), (B6), (B10), and (B8) lead to

$$H \leqslant (1 + \bar{R}_{\text{num}})^{1/2}, \quad \bar{H} \leqslant (1 + R_{\text{num}})^{1/2}, \tag{B11}$$

which is equivalent to the estimate (5.26) in view of (B5).

[1]A. A. Skorupski, J. Math. Phys. **29**, 1814 (1988).
[2]B. W. Shore, J. Chem. Phys. **59**, 6450 (1973).
[3]N. Fröman and P. O. Fröman, *JWKB-Approximation, Contributions to the Theory* (North-Holland, Amsterdam, 1965).
[4]A. A. Skorupski, Rep. Math. Phys. **17**, 161 (1980).
[5]J. M. Kowalski and J. L. Fry, J. Math. Phys. **28**, 2407 (1987).
[6]A. A. Skorupski, J. Math. Phys. **25**, 2655 (1984).
[7]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1944).

Electromagnetic angular momentum radiated by two-point charges

J. M. Aguirregabiria and J. R. Etxebarria
Fisika Teorikoa, Zientzi Fakultatea, Euskal Herriko Unibertsitatea, P. K. 644, 48080 Bilbo, Spain

(Received 29 September 1987; accepted for publication 6 April 1988)

A previous covariant approach in classical field theory to the definition of the energy and linear momentum radiated by two-point charges is extended to the case of angular momentum. First, the mixed contribution to the radiation rate of angular momentum for two proper time infinitesimal intervals is defined, shown to be independent of the hypersurface used, and computed in an exact covariant way. With this result a definition of the radiated angular momentum for a system of interacting particles is given in the framework of predictive relativistic mechanics. The lowest order of this quantity is calculated and compared with a previous result obtained by means of a different method. The agreement of the results obtained in both approaches can be interpreted as a test for the Lorentz–Dirac equation.

I. INTRODUCTION

The standard covariant approach to the definition of the energy and linear momentum radiated by a *single* point charge was proposed by Schild.[1] This approach was later extended to the definition of angular momentum and also to the case of a single spinning particle.[2] The covariant definition of the energy and linear momentum radiated by a *system* of point charges was analyzed in Ref. 3.

The purpose of this paper is to extend the later formalism to the angular momentum radiated by a system of point charges. As in Ref. 3, we shall restrict ourselves to the case of two particles, though the analysis can be immediately extended to any finite number of charges.

The definition given for the radiated quantities is a generalization of the usual covariant one for a single particle. In Sec. II we define the radiated angular momentum corresponding to two infinitesimal intervals of the world lines of the charges, as the contribution to the total electromagnetic angular momentum present at future infinity due to the fields created by the charges in the considered intervals. In a way similar to the one used by Schild[1] for the linear momentum radiated by one charge, the aforementioned definition can be seen to be independent of the hypersurface used to calculate the angular momentum at future infinity. This definition gives us a radiation rate of angular momentum that can be calculated in an exact covariant way and depends only on the positions, velocities, and accelerations of the charges.

In contrast with the case of a single charge, in order to calculate the angular momentum radiated per unit of coordinate time, one must know something else about the dynamics of the particles. To do that we shall use the framework of predictive relativistic mechanics[4] to define, in Sec. III, the angular momentum radiated by an isolated system of two-point charges, as a function only of the positions and velocities of the particles in the considered configuration.

The lowest-order contribution to this quantity and to the total linear and angular momenta radiated along the complete motion corresponding to given initial conditions is calculated in Sec. IV.

The analysis is readily extended to the case of the intrinsic angular momentum in Sec. V. This allows us to compare this result with the only previous covariant calculation of radiated quantities we know.[5] This was obtained by a completely different method without previous knowledge of the radiation rates. The particular terms given in Ref. 5 are easier to calculate by means of the way used there, but, in general—even for the other term of the same order given here for the first time—the method proposed here is more convenient because fewer orders of the dynamics need be known. The agreement of the results obtained by means of these two methods can be understood as a positive test for the coherence between the Lorentz–Dirac equation and the usual conservation laws of the classical field theory applied to the Maxwell stress tensor.

Finally, in Sec. VI, a short comment is made about the conditions for null radiation.

II. RADIATION RATE OF ANGULAR MOMENTUM

Let us consider two pointlike particles with charges e_a ($a = 1,2$). We assume that the total electromagnetic field is the sum of the retarded fields $F_a^{\alpha\beta}$ of the particles, and we take $c = 1$ and $\eta^{\alpha\beta} = \mathrm{diag}(-1,1,1,1)$. The position in the Minkowski space-time of the particle a, when τ_a is its proper time, is $x_a = \phi_a^\alpha(\tau_a)$, and its velocity and acceleration are $u_a^\alpha = d\phi_a^\alpha/d\tau_a$ and $\xi_a^\alpha = d^2\phi_a^\alpha/d\tau_a^2$.

If $T^{\alpha\beta}$ is the electromagnetic Maxwell tensor, the angular momentum density is $j^{\alpha\beta\gamma} = x^{[\alpha}T^{\beta]\gamma} \equiv x^\alpha T^{\beta\gamma} - x^\beta T^{\alpha\gamma}$, and thus the electromagnetic angular momentum corresponding to the value λ of the proper time of an inertial observer with four-velocity n^α is given by

$$J^{\alpha\beta}(n,\lambda) \equiv -\int_{\Sigma(n,\lambda)} j^{\alpha\beta\gamma}\, d^3\sigma_\gamma, \tag{1}$$

where $\Sigma(n,\lambda)$ is the spacelike hyperplane orthogonal to the world line of the particle at point λ, and $d^3\sigma^\alpha = d^3\sigma\, n^\alpha$ is the volume element of $\Sigma(n,\lambda)$.

The contribution to the angular momentum due exclusively to the electromagnetic field of the particle a is

$$J_a^{\alpha\beta}(n,\lambda) = -\int_{-\infty}^{\tau_a(n,\lambda)} d\tau_a \int_{\Sigma_a(n,\lambda,\tau_a)} j_a^{\alpha\beta\gamma}\, d^2\sigma_{a\gamma}$$
$$\equiv \int_{-\infty}^{\tau_a(n,\lambda)} d\tau_a\, \frac{\partial J_a^{\alpha\beta}}{\partial\tau_a}(n,\lambda,\tau_a). \tag{2}$$

 0022-2488/88/081832-05$02.50

In the last expression, $\Sigma_a(n,\lambda,\tau_a)$ is the intersection between $\Sigma(n,\lambda)$ and the future light cone with a vertex in x_a^α, $\tau_a(n,\lambda)$ is the value of τ_a at the intersection of the world line of particle a and $\Sigma(n,\lambda)$, and we have put $d^3\sigma^\alpha \equiv d\tau_a\, d^2\sigma_a^\alpha$.

As indicated, the integrand in the last integral depends on n^α and λ, but in the same way as proved by Schild[1] for the linear momentum, it can be seen[2] that the following limit is in fact independent of n^α and can be understood as the contribution to the rate of angular momentum radiated exclusively by the particle a:

$$\frac{dJ_{ar}^{\alpha\beta}}{d\tau_a}(\tau_a) \equiv \lim_{\lambda\to+\infty} \frac{\partial J_a^{\alpha\beta}}{\partial\tau_a}(n,\lambda,\tau_a)$$
$$= \tfrac{2}{3}e_a^2(\xi_a^\gamma \xi_{a\gamma} x_a^{[\alpha} u_a^{\beta]} + u_a^{[\alpha}\xi_a^{\beta]}). \tag{3}$$

The joint contribution due to the fields of both particles is

$$J_{12}^{\alpha\beta}(n,\lambda) = -\int_{-\infty}^{\tau_1(n,\lambda)} d\tau_1 \int_{-\infty}^{\tau_2(n,\lambda)} d\tau_2 \times \int_{\Sigma_{12}(n,\lambda,\tau_1,\tau_2)} j_{12}^{\alpha\beta\gamma}\, d\sigma_{12\gamma}$$
$$\equiv \int_{-\infty}^{\tau_1(n,\lambda)} d\tau_1 \int_{-\infty}^{\tau_2(n,\lambda)} d\tau_2 \times \frac{\partial^2 J_{12}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}(n,\lambda,\tau_1,\tau_2), \tag{4}$$

where $j_{12}^{\alpha\beta\gamma} \equiv x^{[\alpha} T_{12}^{\beta]\gamma}$ is defined by the joint contribution of both fields to the stress tensor, and $\Sigma_{12}(n,\lambda,\tau_1,\tau_2)$ is the intersection between $\Sigma(n,\lambda)$ and the light cones with vertices in x_1 and x_2. We have taken $d^3\sigma^\alpha \equiv d\tau_1\, d\tau_2\, d\sigma_{12}^\alpha$.

The integral

$$\frac{\partial^2 J_{12}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}(n,\lambda,\tau_1,\tau_2) \equiv -\int_{\Sigma_{12}(n,\lambda,\tau_1,\tau_2)} j_{12}^{\alpha\beta\gamma}\, d\sigma_{12\gamma} \tag{5}$$

is zero if the relative position vector is timelike and has a limit

$$\frac{\partial^2 J_{12r}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}(\tau_1,\tau_2) \equiv \lim_{\lambda\to+\infty} \frac{\partial^2 J_{12}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}(n,\lambda,\tau_1,\tau_2), \tag{6}$$

which is also independent of the inertial observer, as can be seen in a way very similar to the one used in Refs. 1–3 in different cases.

In consequence we shall interpret the limit (6) as the joint contribution per unit of proper time to the radiated angular momentum of the retarded fields of both particles. So the radiated angular momentum corresponding to the fields created by the two particles up to proper times τ_1 and τ_2 can be defined as follows:

$$J_r^{\alpha\beta}(\tau_1,\tau_2) \equiv \sum_{a=1}^{2} \int_{-\infty}^{\tau_a} d\tau_a \frac{dJ_{ar}^{\alpha\beta}}{d\tau_a} + \int_{-\infty}^{\tau_1} d\tau_1 \int_{-\infty}^{\tau_2} d\tau_2 \frac{\partial^2 J_{12r}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}. \tag{7}$$

To calculate the limit in Eq. (6), we shall decompose the electromagnetic field of particle a as $F_a^{\alpha\beta} = F_{a(-1)}^{\alpha\beta} + F_{a(-2)}^{\alpha\beta}$, where $F_{a(-n)}^{\alpha\beta}$ is proportional to r_a^{-n}. Here the retarded distance from the point x^α to the position x_a^α in the retarded proper time τ_a is given by $r_a \equiv -(x^\alpha - x_a^\alpha)u_{a\alpha}$. As a consequence, the mixed contribution to the Maxwell tensor will split into three parts, $T_{12(-2)}^{\alpha\beta}$, $T_{12(-3)}^{\alpha\beta}$, and $T_{12(-4)}^{\alpha\beta}$. In the case of the linear momentum only the term proportional to r_a^{-2} gives a contribution to the radiation, but for the angular momentum we have also to consider $T_{12(-3)}^{\alpha\beta}$, because of the extra factor x^α.

The actual computation of this limit is a rather long one, but the method to perform it is the one described in Sec. III of Ref. 3. Here we only quote the final result in terms of the completely symmetric tensor $I^{\alpha\beta\gamma}$ introduced there:

$$\frac{\partial^2 J_{12r}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}(\tau_1,\tau_2) = \frac{e_1e_2}{2r_{21}} I^{\mu\nu[\alpha} L_{\mu\nu}{}^{\beta]}, \tag{8}$$

with

$$L^{\alpha\beta\gamma} = \sum_{a=1}^{2} \{[(x_{aa'}\xi_{a'})+1]u_{a'}^\alpha - (x_{aa'}u_{a'})\xi_{a'}^\alpha\} u_a^{[\beta}\xi_a^{\gamma]} - [(u_1u_2)\xi_1^\alpha\xi_2^\beta - (u_1\xi_2)\xi_1^\alpha u_2^\beta - (\xi_1u_2)u_1^\alpha\xi_2^\beta + (\xi_1\xi_2)u_1^\alpha u_2^\beta](x_1^\gamma + x_2^\gamma), \tag{9}$$

where $a' \equiv 3-a$, $(A\,B) \equiv A^\alpha B_\alpha$, and

$$x_{aa'}^\alpha \equiv x_a^\alpha - x_{a'}^\alpha,\quad r_{aa'}^2 \equiv x_{aa'}^2 + (x_{aa'}u_{a'})^2. \tag{10}$$

III. RADIATED ANGULAR MOMENTUM IN COPHASE SPACE

To calculate (8) we need only to know the positions, velocities, and accelerations of the particles corresponding to the selected values of the proper times. But to find the radiated angular momentum up to times τ_1 and τ_2 or, simply, per unit of coordinate time, we need complete pieces of the world lines of both particles. As this problem is a dynamical one, we shall raise it in the framework of the predictive relativistic mechanics.[4]

Let us assume that the world lines of the particles are solutions of the following ordinary differential system:

$$\frac{dx_a^\alpha}{d\tau} = u_a^\alpha,\quad \frac{du_a^\alpha}{d\tau} = \xi_a^\alpha(x,u). \tag{11}$$

In order to be an invariant predictive system, the functions $\xi_a^\alpha(x,u)$ appearing in (11) must be Poincaré invariant and satisfy both the orthogonality condition $u_{a\alpha}\xi_a^\alpha(x,u) = 0$ and the Droz-Vincent equations[6] $\mathscr{L}(\mathbf{H}_{a'})\xi_a^\alpha(x,u) = 0$, $\mathscr{L}(\mathbf{H}_a)$ being the Lie derivative with respect to the vector field

$$\mathbf{H}_a = u_a^\alpha \frac{\partial}{\partial x_a^\alpha} + \xi_a^\alpha(x,u)\frac{\partial}{\partial u_a^\alpha}. \tag{12}$$

In classical electrodynamics the evolution equations of an isolated system of two-point charges are the delay differential equations arising from the Lorentz force or the Lorentz–Dirac equation. But it can be proved[4,5,7] that there is a unique invariant predictive system (11) with functions ξ_a^α that coincide with the accelerations of the particles moving according to the delay differential system when the relative position vector is null. This associated predictive system can be computed by means of a perturbative scheme, and all its solutions are also solutions of the corresponding delay differential system. Furthermore, because of the so-called *spontaneous predictivization*,[8] any solution of the Lorentz equations, or of the order-reduced Lorentz–Dirac equations, constructed by the methods of steps, is a solution of the associated predictive system with increasing approximation beyond the first step.

To define the radiated angular momentum in the cophase space—the space of the points (x,u)—we shall use the fact that, from Eq. (7),

$$\frac{\partial J_r^{\alpha\beta}}{\partial \tau_a}(\tau_1,\tau_2) \equiv \frac{dJ_{ar}^{\alpha\beta}}{d\tau_a}(\tau_a) + \int_{\hat\tau_{aa'}}^{\tau_{a'}} d\tau_{a'} \frac{\partial J_{12r}^{\alpha\beta}}{\partial\tau_1\,\partial\tau_2}, \quad (13)$$

where $\hat\tau_{aa'}$ is the proper time of the particle a' when it is placed on the retarded light cone of a. This result suggests to us to define the angular momentum radiated from past infinity to the configuration (x,u) as the function $J_r^{\alpha\beta}(x,u)$ defined in the cophase space by the evolution equation

$$\mathscr{L}(\mathbf{H}_a)J_r^{\alpha\beta}(x,u) = L_a^{\alpha\beta}(x,u)$$
$$\equiv M_a^{\alpha\beta}(x,u) + N_a^{\alpha\beta}(x,u),$$
$$N_a^{\alpha\beta}(x,u) \equiv \int_{\hat\tau_{aa'}}^{0} d\tau\, \phi^*_{a'\tau} M_{12}^{\alpha\beta}(x,u). \quad (14)$$

As this equation is the translation of (13) to cophase space, we have put here

$$M_a^{\alpha\beta}(x,u) \equiv \tfrac{2}{3}e_a^2(\xi_a^\gamma(x,u)\xi_{a\gamma}(x,u)x_a^{[\alpha}u_a^{\beta]} + u_a^{[\alpha}\xi_a^{\beta]}(x,u)), \quad (15)$$

by Eq. (3), and $M_{12}^{\alpha\beta}$ is Eq. (8) with the vector ξ_a^α systematically replaced by the function $\xi_a^\alpha(x,u)$ if the configuration is spacelike, and zero otherwise. The retarded proper time $\hat\tau_{aa'}(x,u)$ is the value of $\tau_{a'}$ when the point $\phi_{a'}^\alpha(x,u;\tau_{a'})$ is in the past light cone of x_a^α, $\phi_a^\alpha(x,u;\tau_a)$ being the solution of system (11) for the initial conditions (x,u). The term $\phi^*_{a'\tau}$ denotes the dual map of

$$\phi_{a'\tau}:\ (x_a^\alpha,u_a^\beta,x_{a'}^\gamma,u_{a'}^\delta) \mapsto (x_a^\alpha,u_a^\beta,\phi_{a'}^\gamma(x,u;\tau),\dot\phi_{a'}^\delta(x,u;\tau)). \quad (16)$$

Conditions (14) do not uniquely determine $J_r^{\alpha\beta}$, and we must also impose the asymptotic condition

$$\lim_{\tau\to-\infty} R_1(\tau)R_2(\tau)J_r^{\alpha\beta}(x,u) = 0, \quad (17)$$

where use has been made of the shift operator

$$R_a(\tau)f(x_a^\alpha,u_a^\beta,x_{a'}^\gamma,u_{a'}^\delta) \mapsto f(x_a^\alpha + \tau u_a^\alpha,u_a^\beta,x_{a'}^\gamma,u_{a'}^\delta). \quad (18)$$

The invariant predictive system associated with an isolated system of two-point charges is also invariant under space reflections. With the notation

$$k = -(u_1u_2), \quad \Lambda^2 = k^2 - 1,$$
$$z_a = \Lambda^{-2}[(x_{aa'}u_a) - k(x_{aa'}u_{a'})], \quad \eta_a = (-1)^{a+1}, \quad (19)$$

$$h^\alpha = x_{12}^\alpha - z_1u_1^\alpha + z_2u_2^\alpha, \quad t_a^\alpha = u_a^\alpha - ku_{a'}^\alpha, \quad (20)$$

this condition can be written as follows[4]:

$$\xi_a^\alpha = \eta_a a_a h^\alpha + l_{aa'}t_{a'}^\alpha. \quad (21)$$

The values of a_a and $l_{aa'}$ can, in principle, be calculated by a perturbative method[4,5,7] by means of expansions in powers of the charges. By means of this perturbative method one can also find successive orders of ϕ_a^α, of $L_a^{\alpha\beta}$, and, finally, of $J_r^{\alpha\beta}$ as solutions of Eqs. (14) and (17).

One nice feature of the definition given above is that to calculate the first nonzero contribution to the radiated angular momentum we need only to know the first term in the expansion of the ξ_a^α. This term is the trivial one given by the acceleration that particle a would have if it were moving in the retarded Coulomb field created by a free particle a'. In fact, if the masses of the charges are m_a, we have up to this lowest order

$$\xi_a^\alpha = (e_ae_{a'}/m_ar_{aa'}^3)(\eta_akh^\alpha - z_at_{a'}^\alpha) + O(e^4), \quad (22)$$

where $r_{aa'}^2 = h^2 + \Lambda^2 z_a^2$, $h^2 \equiv h^\alpha h_\alpha$, and $O(e^4)$ means terms vanishing at least as e^4 when $e = e_1 = e_2 \to 0$. We also need the following elementary results:

$$\phi_{a'}^\alpha(x,u;\tau) = x_{a'}^\alpha + \tau u_{a'}^\alpha + O(e^2), \quad (23)$$
$$\hat\tau_{aa'}(x,u) = \zeta_{aa'} + O(e^2), \quad \zeta_{aa'} \equiv kz_a - z_{a'} - r_{aa'}. \quad (24)$$

Substitution of these in expressions (3), (8), and (9) and in the value of $I^{\alpha\beta\gamma}$ gives, after a straightforward but lengthy calculation,

$$M_a^{\alpha\beta} = (2e_a^3e_{a'}/3m_a\Lambda^2r_{aa'}^3)(\eta_akh^{[\alpha}t_a^{\beta]} + \eta_ak^2h^{[\alpha}t_{a'}^{\beta]}$$
$$+ z_at_a^{[\alpha}t_{a'}^{\beta]}) + O(e^6), \quad (25)$$
$$M_{12}^{\alpha\beta}(x,u) = e_1^2e_2^2(H_1h^{[\alpha}t_1^{\beta]}$$
$$+ H_2h^{[\alpha}t_2^{\beta]} + Tt_1^{[\alpha}t_2^{\beta]}) + O(e^6), \quad (26)$$

where

$$H_a = \frac{1}{2}\eta_a\left[\frac{k}{\Lambda^2r_{aa'}}\left(\frac{1}{m_ar_{aa'}^3} - \frac{1}{m_{a'}r_{a'a}^3}\right) - \frac{z_{a'}}{m_{a'}h^2r_{a'a}^3}\left(\frac{z_a}{r_{aa'}} - \frac{z_{a'}}{r_{a'a}}\right) + \frac{1}{\Lambda^2r_{aa'}^3r_{a'a}}\left(\frac{k(kz_a - z_{a'})^2}{m_{a'}} - \frac{(kz_{a'} - z_a)^2}{m_a}\right)\right], \quad (27)$$

$$T = \frac{1}{2\Lambda^2}\left[\frac{z_1}{m_1r_{12}^4}\frac{z_2}{m_2r_{21}^4} + \frac{z_1(kz_1 - z_2)^2}{m_2r_{12}^3r_{21}^3} - \frac{z_2(kz_2 - z_1)^2}{m_1r_{12}^3r_{21}^3}\right]. \quad (28)$$

It should be stressed that the first nonzero components of the radiated angular momentum rates are of fourth order in the charges, while in the case of the radiated linear momentum they are of sixth order.[3]

In order to calculate the first approximation of functions $N_a^{\alpha\beta}$ on Eq. (14), we have to compute

$$N_a^{\alpha\beta} = \int_{\zeta_{aa'}}^{0} d\tau\, R_{a'}(\tau)M_{12}^{\alpha\beta} + O(e^6)$$
$$= \int_{\hat z_{a'}}^{z_{a'}} dz_{a'}\, M_{12}^{\alpha\beta} + O(e^6), \quad (29)$$

where $\hat z_{a'} \equiv R_{a'}(\zeta_{aa'})z_{a'} = kz_a - r_{aa'}$.

The fourth-order term in expression (29) is the one proportional to $e_1^2e_2^2$ and can be computed by means of Eqs. (26)–(28). The final result is a rather complex one, and we shall not write it explicitly here.

Once we know $L_a^{\alpha\beta}$, the radiated angular momentum per unit coordinate time can be computed by using[9]

$$\frac{dJ^{\alpha\beta}}{dt}(t,x_a^i,v_b^j) = \sum_{c=1}^{2} \frac{\overline{L}_c^{\alpha\beta}}{\bar{u}_c^0}, \tag{30}$$

where a bar over a function in cophase space means to take the following values of the arguments:

$$\begin{aligned} &\bar{x}_1^0 = \bar{x}_2^0 = t, \quad \bar{x}_a^i = x_a^i, \\ &\bar{u}_a^0 = (1 - v_a^i v_{ai})^{-1/2}, \quad \bar{u}_a^i = \bar{u}_a^0 v_a^i. \end{aligned} \tag{31}$$

IV. TOTAL RADIATED MOMENTA

In Ref. 3 a definition was given for the linear momentum radiated along the evolution of the isolated system from past infinity to the considered point of cophase space, but the explicit computation was not carried out. We shall indicate here how this can be performed. But for convenience we shall begin with the angular momentum.

The lowest-order contributions to the evolution equation and asymptotic conditions for $J_r^{\alpha\beta}$ on Eqs. (14) and (17) are the differential conditions

$$\frac{\partial J_r^{\alpha\beta}}{\partial z_a} = L_a^{\alpha\beta} + O(e^6); \tag{32}$$

the integrability conditions

$$\frac{\partial L_a^{\alpha\beta}}{\partial z_{a'}} = \frac{\partial L_{a'}^{\alpha\beta}}{\partial z_a} + O(e^6), \tag{33}$$

which are automatically fulfilled as a consequence of the definition and the Droz-Vincent conditions; and the limit conditions

$$\lim_{z_a \to -\infty} J_r^{\alpha\beta} = 0. \tag{34}$$

By making use of Eq. (14) and the fact that $N_a^{\alpha\beta}$ vanishes except when

$$kz_{a'} - r_{a'a} = \hat{z}_a \leqslant z_a \leqslant \check{z}_a \equiv kz_{a'} + r_{a'a},$$

it can be seen that the solution of the system (32)–(34) is

$$J_r^{\alpha\beta} = \sum_{a=1}^{2} \int_{-\infty}^{z_a} dz_a\, M_a^{\alpha\beta} + \int_{\hat{z}_1}^{z_1} dz_1\, N_1^{\alpha\beta} + \int_{-\infty}^{\hat{z}_1} dz_1\, N_1^{\alpha\beta}\Big|_{z_2 = \check{z}_2} + O(e^6), \tag{35}$$

or the equivalent result obtained by systematically interchanging the indices 1 and 2. From this expression for the angular momentum radiated up to the considered configuration, the total angular momentum radiated along the complete evolution, from past infinity to future infinity, corresponding to initial conditions (x,u), is given by

$$\begin{aligned} J_{tr}^{\alpha\beta}(x,u) &= \lim_{\tau \to +\infty} \phi_{1\tau}^{*} \phi_{2\tau}^{*} J_r^{\alpha\beta}(x,u) \\ &= \lim_{\tau \to +\infty} R_1(\tau) R_2(\tau) J_r^{\alpha\beta}(x,u) + O(e^6) \\ &= \sum_{a=1}^{2} \int_{-\infty}^{\infty} dz_a\, M_a^{\alpha\beta} + \int_{-\infty}^{+\infty} dz_1\, N_1^{\alpha\beta}\Big|_{z_2 = \check{z}_2} + O(e^6). \end{aligned} \tag{36}$$

If the integrals in Eq. (36) are performed, then after a rather long calculation the following result can be obtained:

$$\begin{aligned} J_{tr}^{\alpha\beta} = &\frac{2e_1^2 e_2^2 k\,[k\Lambda - \ln(k+\Lambda)]}{m_1 m_2 h^2 \Lambda^4} \\ &\times (m_1 h^{[\alpha} u_1^{\beta]} - m_2 h^{[\alpha} u_2^{\beta]}) \\ &- \frac{4}{3} \sum_{a=1}^{2} \frac{e_a^3 e_{a'} k}{m_a h^2 \Lambda} \eta_a h^{[\alpha} u_a^{\beta]} + O(e^6). \end{aligned} \tag{37}$$

This calculation can be performed in a very similar way for the total radiated linear momentum using the results of Ref. 3. Here we only quote the final result:

$$\begin{aligned} P_{tr}^{\alpha} = &\frac{e_1^3 e_2^3 \pi}{4 m_1 m_2 h^3 \Lambda^6} [\Lambda(3k^3 - 4k^2 + 9k - 4) \\ &- (3k^2+1)\ln(k+\Lambda)](t_1^\alpha + t_2^\alpha) \\ &- \sum_{a=1}^{2} \frac{e_a^4 e_{a'}^2 \pi}{12 m_a^2 h^3 \Lambda^3} (3k^2+1)(t_a^\alpha + k t_{a'}^\alpha) + O(e^8). \end{aligned} \tag{38}$$

V. RADIATED INTRINSIC ANGULAR MOMENTUM

Since we assume that the system of the two charges is isolated, the total linear momentum of the system, made up of the mechanical and the electromagnetic contributions, is a constant of motion, and the mass of the system $M \equiv (-P^\alpha P_\alpha)^{1/2}$ and the velocity of the center of mass $U^\alpha \equiv M^{-1} P^\alpha$ can be defined and are also constants of motion. If we consider the same inertial observer of Sec. II, the electromagnetic intrinsic angular momentum is given by

$$\begin{aligned} W^\alpha n,\lambda) &\equiv -\frac{1}{2} \int_{\Sigma(n,\lambda)} \epsilon^{\alpha\beta\gamma\delta} U_\beta\, j_{\gamma\delta\mu}\, d^3\sigma^\mu \\ &= \tfrac{1}{2} \epsilon^{\alpha\beta\gamma\delta} U_\beta J_{\gamma\delta}, \end{aligned} \tag{39}$$

where $J^{\alpha\beta}$ is given by expression (1). From here, all the work made with the linear and angular momenta can be repeated. Let us only say that the intrinsic angular momentum radiated along the full evolution corresponding to initial conditions (x,u) is given, taking into account Eq. (36), by

$$W_{tr}^\alpha \equiv \tfrac{1}{2} \epsilon^{\alpha\beta\gamma\delta} U_\beta J_{tr\gamma\delta}. \tag{40}$$

To calculate the first contribution to Eq. (40) we need the lowest-order terms of $J_{rt}^{\alpha\beta}$ and U^α. The first one is given by Eq. (37) and the second is of a purely mechanical nature:

$$\begin{aligned} &U^\alpha = \Delta \sum_{a=1}^{2} m_a u_a^\alpha + O(e^2), \\ &\Delta \equiv (m_1^2 + 2k m_1 m_2 + m_2^2)^{-1/2}. \end{aligned} \tag{41}$$

The final result is

$$W_{tr} = kW\Delta \epsilon^{\alpha\beta\gamma\delta} h_\beta u_{1\gamma} u_{2\delta} + O(e^6),$$

with

$$W \equiv 4e_1^2 e_2^2 \frac{k\Lambda - \ln(k+\Lambda)}{h^2 \Lambda^4} - \frac{4}{3} \sum_{a=1}^{2} e_a^3 e_{a'} \frac{m_{a'}}{m_a h^2 \Lambda}. \tag{42}$$

The two terms proportional to $e_1^3 e_2$ and $e_1 e_2^3$, respectively, were found by Lapiedra and Molina[5] by a different method, without previous calculation of the radiation rates. The term proportional to $e_1^2 e_2^2$ is new, as far as we know.

As was said before, to calculate Eqs. (41) and (42) we have used only the first nonzero term in the perturbative expansion of the accelerations of the particles. This term corresponds to the retarded Coulomb fields of particles moving without acceleration, and thus it is the same for the delay equations obtained from the Lorentz force or for the ones derived from the Lorentz–Dirac equation, for instance. However, in the method of Lapiedra and Molina, two terms coming from the very structure of the Lorentz–Dirac equation are used to calculate the contributions proportional to $e_1^3 e_2$ and $e_1 e_2^3$ in (42). This can be seen as a partial proof of the coherence between the Lorentz–Dirac equation and the usual interpretation of the Maxwell tensor in the conservation laws of the classical electrodynamics. This result is a point against the equations proposed by Mo and Papas[10] and Herrera[11] as an alternative to the Lorentz–Dirac equation, because both equations give—in the method of Lapiedra and Molina—no contribution proportional to $e_a^3 e_{a'}$, as can be easily seen.

VI. FINAL COMMENTS

It is obvious that if the charges are not accelerated there is no radiation. In the usual low-velocity approximation to the radiation, it is also easily seen that if a system of two-point charges radiates no energy, the charges must move without acceleration. By making use of the result of Ref. 3 corresponding to Eq. (8) (which is based only upon classical field theory) and no additional assumption, it can be seen[12] also in this covariant approach that the only possibility not to have radiated energy is that the two particles move with constant velocity.[13] However, this converse result is not necessarily true for macroscopic systems[14] that can be seen as composed of an infinite number of point charges.

[1] A. Schild, J. Math. Anal. Appl. **1**, 127 (1960). See also F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, MA, 1965).

[2] J. Cohn and H. Wiebe, J. Math. Phys. **17**, 1496 (1976). See also C. Teitelboim, D. Villarroel, and Ch. G. Van Weert, Riv. Nuovo Cimento **3**, 9 (1980).

[3] J. M. Aguirregabiria and L. Bel, Phys. Rev. D **29**, 1099 (1984).

[4] L. Bel and X. Fustero, Ann. Inst. H. Poincaré **XXV**, 411 (1976), and references therein.

[5] R. Lapiedra and A. Molina, J. Math. Phys. **20**, 1308 (1979).

[6] Ph. Droz-Vincent, Lett. Nuovo Cimento **1**, 839 (1969); Phys. Scr. **2**, 129 (1970); R. Arens, Arch. Ration. Mech. Anal. **47**, 255 (1972); L. Bel, Ann. Inst. H. Poincaré **XII**, 307 (1970).

[7] L. Bel, A. Salas, and J. M. Sanchez, Phys. Rev. D **7**, 1099 (1973).

[8] L. Bel, in *Relativistic Action at a Distance: Classical and Quantum Aspects, Lecture Notes in Physics*, Vol. 162, edited by J. Llosa (Springer, Berlin, 1982); C. R. Acad. Sci. **294**, 463 (1982).

[9] L. Bel and J. Martin, C. R. Acad. Sci. **277**, 233 (1973).

[10] T. C. Mo and C. H. Papas, Phys. Rev. D **4**, 3566 (1971).

[11] J. C. Herrera, Phys. Rev. D **15**, 453 (1977); **21**, 384 (1980).

[12] J. R. Etxebarria, Ph.D. thesis, Euskal Herriko Unibertsitatea, Bilbo, 1986.

[13] We are indebted to Dr. B. Coll for suggesting to us the analysis of this problem.

[14] See, for instance, B. B. Laud, *Electromagnetics* (Wiley, New Delhi, 1983), p. 231; T. A. Abbot and D. J. Griffiths, Am. J. Phys. **53**, 1203 (1985).

Particlelike solutions to nonlinear classical real scalar field theories

J. A. Okolowski
Department of Electrical Engineering, Widener University, Chester, Pennsylvania 19013
M. Slomiana
Department of Mechanical Engineering, Widener University, Chester, Pennsylvania 19013

(Received 18 August 1987; accepted for publication 6 April 1988)

A classical Lorentz-covariant nonlinear model scalar field theory, which admits an exact, closed-form, spatially localized, nonsingular solution, is considered. This solution corresponds to a dynamically unstable extended "particle" of finite positive "rest mass." An approximate two-particle solution is constructed, and the associated effective two-particle Lagrangian is derived by means of a Rayleigh–Ritz procedure.

I. INTRODUCTION

Recent interest has been attached to the study of spatially localized finite energy solutions to nonlinear Lorentz-covariant classical field theories.[1–3] The existence of such solutions (referred to as solitons) has been established for a family of classical field theories characterized by nonderivative interactions and unbroken SO(2) internal symmetry.[1] Absolute stability and spherical symmetry are fundamental properties of these solutions, which become unstable when coupled to the quantized neutrino field.[2] For a class of SO(n)-invariant relativistic quantum field theories with nonderivative interactions, one-particle propagators (accurate to order $\hbar$) have been derived from a soliton-based representation of the generating functional for time-ordered Green's functions.[3] A noteworthy consequence of this result is that classical instability implies one-particle quantum instability if the quantum field theory has a classical counterpart that admits a time-independent soliton solution.

Exact solutions have been obtained for the hydrodynamical version of certain nonlinear Schrödinger equations,[4,5] namely, the generalized nonlinear Schrödinger–Langevin equation[4] and the logarithmic nonlinear Schrödinger equation,[5] both with prescribed external potentials. From the hydrodynamical considerations of Refs. 4 and 5, it follows that each of these nonlinear Schrödinger equations admits a nonspreading (solitonlike) wave-packet solution when the form of the external potential is appropriately specialized.

In this paper, particlelike[6] (soliton) solutions to specific nonlinear classical field theories are investigated.

Let us consider a Lorentz-covariant classical field theory based on a Lagrangian density of the general form[7]

$$\mathscr{L} \equiv \tfrac{1}{2}(\dot{\phi})^2 - \tfrac{1}{2}|\nabla\phi|^2 + f(\phi), \tag{1}$$

where $\phi = \phi(\mathbf{x},t)$ is a real scalar field. With the Langrangian functional[8]

$$L = L[\phi, \dot{\phi}] \equiv \int \mathscr{L}\, d^3\mathbf{x}, \tag{2}$$

the action principle

$$\delta \int_{t_1}^{t_2} L\, dt = 0 \tag{3}$$

leads to the field equation

$$\ddot{\phi} - \nabla^2\phi = f'(\phi). \tag{4}$$

At a solution to (4), the total field energy

$$E = \int \left[\frac{1}{2}(\dot{\phi})^2 + \frac{1}{2}\left|\nabla\phi\right|^2 - f(\phi)\right] d^3\mathbf{x} \tag{5}$$

is an important constant of the motion ($\dot{E} = 0$).

In the present work, the self-interaction Lagrangian density, $f(\phi)$, is specialized to the form

$$f(\phi) \equiv -\lambda_1|\phi|^3 + \lambda_2\phi^4, \tag{6}$$

where λ_1 and λ_2 are arbitrary positive coupling constants. The resulting model field theory admits a rigorous, closed-form, time-independent, dynamically unstable particlelike solution of finite positive energy. Moreover, an approximate time-dependent solution is constructed which describes two interacting nonrelativistic extended "particles" separated by a distance large compared to their "radii." Dynamical consequences of the latter solution are deduced from the associated effective two-particle Lagrangian.

A generalization of (6) is discussed in the Appendix. There it is shown that (6) is the only polynomial member of a class of self-interaction Lagrangian densities for which (4) has static particlelike solutions, each with finite positive energy.

II. A SOLVABLE NONLINEAR SCALAR FIELD THEORY

For a model theory derived from (1) with the self-interaction Lagrangian density (6), the field equation (4) becomes

$$\ddot{\phi} - \nabla^2\phi = -3\lambda_1|\phi|\phi + 4\lambda_2\phi^3. \tag{7}$$

This nonlinear equation admits the static, singularity-free, spatially localized, spherically symmetric solution

$$\phi_0 = \phi_0(|\mathbf{x}|) \equiv \pm(2/3\lambda_1)\,(|\mathbf{x}|^2 + 2\lambda_2/9\lambda_1^2)^{-1}. \tag{8}$$

Observe that the particlelike solution (8) is characterized by a "size parameter"

$$Z \equiv (2\lambda_2/9\lambda_1^2)^{1/2}, \tag{9}$$

which gives the order of magnitude of the "particle radius." Evaluating (5) with (6) and (8) produces the total static field energy, or "particle rest mass,"

$$\pi^2\lambda_1(2\lambda_2^3)^{-1/2} \equiv m_0. \tag{10}$$

Now consider the dynamical stability of (8) when the perturbation about (8) is given by

 0022-2488/88/081837-03$02.50

$$\phi(\mathbf{x},t) = \phi_0(|\mathbf{x}|) + (\eta(u)/Zu)\cos kt,$$
$$u \equiv |\mathbf{x}|Z^{-1}, \qquad (11)$$

where the constant k may be either purely real or purely imaginary, and

$$|\eta(u)|/Zu \ll |\phi_0(|\mathbf{x}|)|,$$

for all $\mathbf{x}$. We substitute (11) into (7) and retain only terms linear in η to obtain the Schrödinger-like eigenvalue equation

$$-\eta''(u) + F(u)\eta(u) = (kZ)^2\eta(u), \qquad (12)$$

in which

$$F(u) = 4/(1+u^2) - 24/(1+u^2)^2. \qquad (13)$$

Equation (12) must be supplemented with the appropriate boundary conditions for a spatially localized nonsingular perturbation,

$$\eta(0) = \lim_{u\to\infty} \eta(u) = 0. \qquad (14)$$

Upper and lower bounds for k^2 follow from considerations based on the generic eigenvalue equation for κ, ξ, and some suitably chosen real-valued function $G(u)$,

$$-\xi''(u) + G(u)\xi(u) = \kappa\xi(u), \qquad (15)$$

with the real-valued eigenfunction $\xi(u)$ subject to the boundary conditions

$$\xi(0) = \lim_{u\to\infty} \xi(u) = 0. \qquad (16)$$

When $G(u)$ is defined by

$$G(u) \equiv 12/(1+u^2) - 24/(1+u^2)^2, \qquad (17)$$

we have $G(u) > F(u)$ for all $u \geqslant 0$. With (17), Eq. (15) has the solution

$$\xi(u) = u/(1+u^2)^2 \qquad (18)$$

belonging to the eigenvalue $\kappa = 0$. Since (18) is nodeless for all $u > 0$, we infer that $k_0^2 \equiv \min k^2 < 0$. Furthermore, when $G(u)$ is defined by

$$G(u) \equiv -\frac{64}{11} + \frac{20}{1+u^2} - \frac{35}{(1+u^2)^2}, \qquad (19)$$

we have $G(u) \leqslant F(u)$ for all $u \geqslant 0$. With (19), Eq. (15) has the solution

$$\xi(u) = u/(1+u^2)^{5/2} \qquad (20)$$

belonging to the eigenvalue $\kappa = -\frac{64}{11}$. Since (20) is nodeless for all $u > 0$, we infer that $(k_0Z)^2 > -\frac{64}{11}$. Consequently, the minimum value of k^2 is bounded according to $0 > (k_0Z)^2 > -\frac{64}{11} \cong -5.818$, and the perturbation term in (11) grows exponentially with time in a dynamically unstable manner.

In order to estimate the value of $|k_0|$, we employ a Rayleigh–Ritz procedure. For this purpose we express (12) as a variational principle

$$\delta\gamma = 0, \quad \gamma \equiv \int_0^\infty [\eta'(u)^2 + F(u)\eta(u)^2]du, \qquad (21)$$

with $\eta(u)$ subject to the normalization condition

$$\int_0^\infty \eta(u)^2\, du = 1. \qquad (22)$$

We seek an approximate solution of the form

$$\eta(u) = \sqrt{2}u/[B(\alpha - \tfrac{3}{2},\tfrac{3}{2})(1+u^2)^\alpha]^{1/2}, \qquad (23)$$

where α is a variational parameter ($> \frac{3}{2}$), and B is standard notation for the beta function. By substituting (23) into the definition part of (21), we find

$$\gamma = (2\alpha - 3)(3\alpha^2 - 40\alpha + 32)/4\alpha(\alpha + 1), \qquad (24)$$

from which we conclude that $\min \gamma \cong -5.447 \cong (k_0Z)^2$ for $\alpha \cong 5.3$. Thus $|k_0| \cong 2.334Z^{-1}$, and we obtain the characteristic time τ for the exponential dissolution of (8),

$$\tau \equiv |k_0|^{-1} \cong 0.428Z. \qquad (25)$$

Let us turn our attention to the interaction of two identical extended particles that correspond to the particlelike scalar fields generated from (8):

$$\phi_0(i) \equiv \pm (2/3\lambda_1)(|\mathbf{x} - \boldsymbol{\zeta}_i|^2 + Z^2)^{-1} \quad (i = 1, 2), \qquad (26)$$

where the coordinates $\boldsymbol{\zeta}_i = \boldsymbol{\zeta}_i(t)$ locate the center of the ith particle. In terms of (26), the field equation (7) admits an approximate solution of the form

$$\tilde{\phi} \equiv \phi_0(1) + \phi_0(2)$$
$$= \pm (2/3\lambda_1)[(|\mathbf{x} - \boldsymbol{\zeta}_1|^2 + Z^2)^{-1}$$
$$+ (|\mathbf{x} - \boldsymbol{\zeta}_2|^2 + Z^2)^{-1}], \qquad (27)$$

for slowly moving ($|\dot{\boldsymbol{\zeta}}_i(t)| \ll 1$), widely separated ($|\boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1| \gg Z$) particles. Here we regard $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$ as variational parameters to be determined by the action principle (3) with $L = L[\tilde{\phi}, \dot{\tilde{\phi}}]$. Hence by substituting (27) into (2) and using well-known integration techniques to evaluate the resulting integrals, we obtain the effective nonrelativistic two-particle Lagrangian

$$L[\tilde{\phi}, \dot{\tilde{\phi}}] \equiv \tilde{L} = \tilde{L}(\boldsymbol{\zeta}_i, \dot{\boldsymbol{\zeta}}_i)$$
$$= (m_0/2)|\dot{\boldsymbol{\zeta}}_1|^2 + (m_0/2)|\dot{\boldsymbol{\zeta}}_2|^2 - 2m_0$$
$$+ U(\boldsymbol{\zeta}_i, \dot{\boldsymbol{\zeta}}_i) - V(|\boldsymbol{\zeta}|), \qquad (28)$$

where

$$U(\boldsymbol{\zeta}_i, \dot{\boldsymbol{\zeta}}_i) = (16\pi^2/9\lambda_1^2 |\boldsymbol{\zeta}|^3)[\mathscr{F}(\sigma)\dot{\boldsymbol{\zeta}}_1\cdot\dot{\boldsymbol{\zeta}}_2$$
$$+ \mathscr{G}(\sigma)(\mathbf{n}\cdot\dot{\boldsymbol{\zeta}}_1)(\mathbf{n}\cdot\dot{\boldsymbol{\zeta}}_2)], \qquad (29)$$

$$\boldsymbol{\zeta} \equiv \boldsymbol{\zeta}_2 - \boldsymbol{\zeta}_1, \quad \sigma \equiv Z/|\boldsymbol{\zeta}| \ll 1, \quad \mathbf{n} \equiv \boldsymbol{\zeta}/|\boldsymbol{\zeta}|, \qquad (30)$$

$$\mathscr{F}(\sigma) = \frac{\pi}{4} - \frac{\sigma}{1+4\sigma^2} - \frac{1}{2}\tan^{-1}(2\sigma), \qquad (31)$$

$$\mathscr{G}(\sigma) = 2\sigma/(1+4\sigma^2)^2 - 3\mathscr{F}(\sigma), \qquad (32)$$

and

$$V(|\boldsymbol{\zeta}|) = -128\pi^2 Z/9\lambda_1^2 (|\boldsymbol{\zeta}|^2 + 4Z^2)^2. \qquad (33)$$

Because of (3) and (28), the particle coordinates $\boldsymbol{\zeta}_i$ satisfy the Euler–Lagrange equations

$$\frac{\partial\tilde{L}}{\partial\boldsymbol{\zeta}_i} - \frac{d}{dt}\frac{\partial\tilde{L}}{\partial\dot{\boldsymbol{\zeta}}_i} = 0, \qquad (34)$$

which ensure conservation of the total two-particle energy

$$\tilde{E} \equiv \left(\dot{\boldsymbol{\zeta}}_1\cdot\frac{\partial}{\partial\dot{\boldsymbol{\zeta}}_1} + \dot{\boldsymbol{\zeta}}_2\cdot\frac{\partial}{\partial\dot{\boldsymbol{\zeta}}_2} - 1\right)\tilde{L}$$
$$= (m_0/2)|\dot{\boldsymbol{\zeta}}_1|^2 + (m_0/2)|\dot{\boldsymbol{\zeta}}_2|^2 + 2m_0$$
$$+ U(\boldsymbol{\zeta}_i, \dot{\boldsymbol{\zeta}}_i) + V(|\boldsymbol{\zeta}|) = \text{const.} \qquad (35)$$

In addition to the usual kinetic energy and rest mass energy terms, the total two-particle energy $\tilde{E}$ includes the velocity-dependent interaction energy $U(\zeta_i,\dot{\zeta}_i)$, as well as the mutual potential energy $V(|\zeta|)$. Note that in the limiting case $\sigma\to 0$, we have

$$\lim_{\sigma\to 0} V(|\zeta|) = 0, \tag{36}$$

while the quantity $U(\zeta_i,\dot{\zeta}_i)$ reduces to the interaction energy of two "dipoles,"

$$\lim_{\sigma\to 0} U(\zeta_i,\dot{\zeta}_i) = [\mathbf{P}_1\cdot\mathbf{P}_2 - 3(\mathbf{n}\cdot\mathbf{P}_1)(\mathbf{n}\cdot\mathbf{P}_2)]|\zeta|^{-3}, \tag{37}$$

with velocity-induced "dipole moments"

$$\mathbf{P}_i \equiv (2\pi^{3/2}/3\lambda_1)\dot{\zeta}_i. \tag{38}$$

APPENDIX: PARTICLELIKE SOLUTIONS TO REAL SCALAR FIELD THEORIES FEATURING A GENERALIZATION OF EQ. (6)

With arbitrary positive coupling constants λ_1 and λ_2, a generalization of (6) is given by

$$f(\phi) \equiv -\lambda_1|\phi|^N + \lambda_2\phi^{2(N-1)}, \tag{A1}$$

where N is a positive real number (>2), not necessarily an integer. By virtue of (A1), Eq. (4) takes the time-independent form

$$-\nabla^2\hat{\phi} = -\lambda_1 N|\hat{\phi}|^{(N-2)}\hat{\phi} + 2\lambda_2(N-1)\hat{\phi}^{(2N-3)}, \tag{A2}$$

for the admissible real scalar field $\hat{\phi} = \hat{\phi}(\mathbf{x})$. We require a necessary condition for the existence of particlelike solutions to (A2); therefore we evoke the pseudovirial theorem[9] to obtain

$$6\int(-\lambda_1|\hat{\phi}|^N + \lambda_2\hat{\phi}^{2(N-1)})d^3\mathbf{x} = \int|\nabla\hat{\phi}|^2\, d^3\mathbf{x}. \tag{A3}$$

Next we multiply (A2) by $\hat{\phi}$, and then integrate over all $\mathbf{x}$ to find

$$\int(-\lambda_1 N|\hat{\phi}|^N + 2\lambda_2(N-1)\hat{\phi}^{2(N-1)})d^3\mathbf{x} = \int|\nabla\hat{\phi}|^2\, d^3\mathbf{x}. \tag{A4}$$

The validity of both (A3) and (A4) is contingent upon a $\hat{\phi}(\mathbf{x})$ of function class C^2 for all $\mathbf{x}$, and such that

$$\lim_{|\mathbf{x}|\to\infty}[|\mathbf{x}|^{3/s}\hat{\phi}(\mathbf{x})] = 0, \quad s\equiv\min\{N, 6\}. \tag{A5}$$

From (A3) and (A4) we derive the global relation

$$3\lambda_1\left(\frac{N-2}{4-N}\right)\int|\hat{\phi}|^N d^3\mathbf{x} = \int|\nabla\hat{\phi}|^2\, d^3\mathbf{x} > 0, \tag{A6}$$

which implies that particlelike solutions to (A2) can exist only if $2<N<4$. Under this restriction on values of N, Eq. (A2) admits the spherically symmetric particlelike solution

$$\hat{\phi}_0 = \hat{\phi}_0(|\mathbf{x}|) \equiv \pm\left[\frac{\lambda_1 N(N-2)^2}{2(4-N)}(|\mathbf{x}|^2 + \hat{Z}^2)\right]^{1/(2-N)},$$
$$\hat{Z} \equiv (4-N)(2\lambda_2)^{1/2}/N(N-2)\lambda_1, \tag{A7}$$

with the associated total field energy (particle rest mass)

$$\hat{m}_0 = \frac{8\sqrt{2}\pi}{3(N-2)^3}B\left(\frac{5}{2}, \frac{6-N}{2N-4}\right)\times\left[\left(\frac{\lambda_1 N}{4-N}\right)^{(4-N)}\lambda_2^{(N-6)/2}\right]^{1/(N-2)}, \tag{A8}$$

computed by means of (A7), (A1), and (5). For the special case $N=3$, treated in Sec. II, we have the only integer value of N for which (A2) has a particlelike solution of finite positive energy.

By taking the limits $\lambda_1\to 0$ and $N\to 4$ in such a way that $(4-N)/\lambda_1\to 8\beta^2$, where β is an arbitrary dimensionless positive constant, Eqs. (A7) and (A8) are brought to their respective limiting forms,

$$\hat{\phi}_0\to\theta_0 \equiv \pm\beta(|\mathbf{x}|^2 + Z_0^2)^{-1/2},$$
$$\hat{Z}\to Z_0 \equiv \beta^2(2\lambda_2)^{1/2}, \tag{A9}$$

and

$$\hat{m}_0\to M_0 \equiv (\pi^2/4)(2\lambda_2)^{-1/2}. \tag{A10}$$

The time-independent particlelike scalar field θ_0, with particle rest mass M_0, satisfies the nonlinear field equation

$$\ddot{\phi} - \nabla^2\phi = 6\lambda_2\phi^5, \tag{A11}$$

derived from the Lagrangian density

$$\mathscr{L} = \tfrac{1}{2}(\dot{\phi})^2 - \tfrac{1}{2}|\nabla\phi|^2 + \lambda_2\phi^6. \tag{A12}$$

A model scalar field theory based on (A12) has been previously investigated.[10–13]

[1]S. Coleman, Nucl. Phys. B **262**, 263 (1985).
[2]A. Cohen, S. Coleman, H. Georgi, and A. Manohar, Nucl. Phys. B **272**, 301 (1986).
[3]R. Raczka and L. Roszkowski, Nuovo Cimento A **90**, 435 (1985).
[4]A. B. Nassar, Phys. Rev. A **33**, 2134 (1986).
[5]A. B. Nassar, Phys. Rev. A **33**, 3502 (1986).

[6]In the present paper, the adjective "particlelike" is equivalent to the phrase "spatially localized and singularity-free."
[7]Physical units are chosen such that $\hbar = c = 1$.
[8]The formulation of classical field theory in terms of functionals is discussed in G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).

[9]G. Rosen, J. Math. Phys. **7**, 2066 (1966).
[10]G. Rosen, J. Math. Phys. **6**, 1269 (1965).
[11]G. Rosen, J. Math. Phys. **8**, 573 (1967).
[12]G. H. Derrick and W. Kay-Kong, J. Math. Phys. **9**, 232 (1968).
[13]G. Pinski, J. Math. Phys. **9**, 1323 (1968).

Relativistic Schrödinger equations

L. Bel
Institut Henri Poincaré, Laboratoire de Physique Théorique, 11 rue Pierre et Marie Curie, 75231 Paris Cédex 05, France

E. Ruiz
Departamento de Física Teórica, Universidad de Salamanca, 37008 Salamanca, Spain

(Received 30 March 1988; accepted for publication 20 April 1988)

A representation of the Poincaré group by means of operators acting on the function space of the nonrelativistic quantum mechanics is introduced. The conditions that the Hamiltonian of an interacting particle system must verify in order to define a relativistic invariant Schrödinger equation are found out. The relation with usual quantum constraint theories is also studied.

I. INTRODUCTION

Our aim in this paper is to introduce a relativistic quantum mechanics that describes systems of N interacting particles by means of a noncovariant formalism. As far as we know, the best known theories on this topic available in the literature are based on manifestly covariant formalisms.[1] Nevertheless, *noncovariant formalisms* are closer to the standard form of nonrelativistic quantum mechanics and first-order relativistic corrections to it. As many authors have pointed out, covariance is not an essential requirement for a relativistic theory, even though it may be desirable in many cases. Covariant formalisms accomplish Poincaré's invariance in a clear and easy to handle fashion. On the contrary, noncovariant formalisms make the relativistic invariance of the theory rather involved. In fact, much of the work in this paper is devoted to elucidating that question.

What does a noncovariant formalism mean? In brief, it is a formalism that retains all the basic framework of nonrelativistic quantum mechanics, except for Galileo's invariance. Let us be more precise in those points which confer a noncovariant character to the approach.

First of all, we have to say what we call a N-particle wave function. Let s_a $(a,b,... = 1,2,...,N)$ be the spin of the particle labeled a, and let $\{\chi_a{}^{P_a}\}$ $(P_a,Q_a,... = -s_a, -s_a + 1,...,s_a)$ be a basis of the spin space of this particle [$(2s_a+1)$-dimensional complex vector space]. We denote by

$$\chi^{(P)} \equiv \chi^{P_1,P_2,...,P_N} = \chi_1{}^{P_1} \otimes \chi_2{}^{P_2} \otimes \cdots \otimes \chi_N{}^{P_N} \tag{1.1}$$

the basis vectors of the tensor product of all particle spin spaces. Wave functions read in this notation as in nonrelativistic quantum mechanics,

$$\psi = \psi_{P_1,...,P_N}(t,\mathbf{x}_a)\chi^{P_1,...,P_N} \equiv \psi_{(P)}\chi^{(P)} \tag{1.2}$$

(here $\mathbf{x}_a$ is the "position vector" of the particle a referred to an inertial observer at the time t measured by this observer). The wave function space of the theory is a subset of the function space defined by (1.2).

Even in nonrelativistic quantum mechanics, frames related by Galileo transformations use different wave functions to describe the same physical state of a system. Therefore, the theory supplies a wave function transformation mechanism that tells us how wave functions relative to different inertial observers are related. It amounts to giving a representation of the invariance group of the theory (Galileo's group) by means of linear operators acting on the wave function space of the theory. To generalize this structure to cover the relativistic case (i.e., to substitute Galileo's group by Poincaré's group) may be rather cumbersome if we maintain the wave function space of nonrelativistic quantum mechanics. The standard action of the Poincaré group in Minkowski's space-time leads naturally to a representation of the Poincaré group on a space of functions depending on the space-time coordinates of all the particles (i.e., N-time wave functions). The images by a boost of points of the space-time that are simultaneous in an inertial frame have different time coordinates in a boosted frame. We shall introduce in Sec. II a representation of the Poincaré group that preserves the one-time dependence of wave functions.

On the other hand, wave functions (1.2) are well suited to a time evolution driven by a Schrödinger-like equation

$$(\hbar = c = 1), \quad i\frac{\partial\psi}{\partial t} = H\psi. \tag{1.3}$$

However, the relativistic invariance of Schrödinger's equation is rather obscure. It has to be understood in connection with the representation of the Poincaré group we have mentioned above. We shall examine that question throughout Sec. III. Let us say in advance that the Hamiltonian H must verify some commuting relations with the other infinitesimal generators of the representation in order to define a Poincaré invariant Schrödinger equation.

The last sections will be devoted to exploring the relation of our formalism with the covariant formalism that is based on a system of N Klein–Gordon equations.[2]

II. REPRESENTATION OF THE POINCARÉ GROUP ON THE ONE-TIME FUNCTION SPACE

We proceed to build up a representation of the Poincaré group on the function space (1.2); that is, we shall define a one-to-one correspondence $(L,A)\leftrightarrow \mathbf{U}(L,A)$ that associates a linear operator $\mathbf{U}(L,A)$ to each Poincaré transformation (L is a Lorentz matrix and A a four-vector). These operators must satisfy the group composition law,

$$\mathbf{U}(L_2L_1,A_1 + L_1{}^{-1}A_2) = \mathbf{U}(L_2,A_2)\mathbf{U}(L_1,A_1). \tag{2.1}$$

An operator of the representation acting on a one-time function ψ gives as output another one-time function

$$\psi' = D^{(Q)}{}_{(P)}(L)\,U(L,A)\psi_{(Q)}(t,\mathbf{x}_a)\chi^{(P)}, \tag{2.2}$$

where

 0022-2488/88/081840-07$02.50

$$D^{(Q)}{}_{(P)}(L) \equiv \prod_{b=1}^{N} (D_b)^{Q_b}{}_{P_b}(L). \tag{2.3}$$

The term $(D_b)^{Q_b}{}_{P_b}$ is an irreducible (or not) representation matrix of the Lorentz group on the spin space of the particle b, and $U(L,A)$ is a linear operator belonging to a representation of the Poincaré group on the one-time scalar function space (wave functions used for describing systems of spinless particles). Our task in this section will consist of obtaining the explicit form of this unusual representation of the Poincaré group.

Let us consider the momentum space of an N-particle system (we can assume spin-0 particles without loss of generality); $k_a{}^\alpha$ will denote the linear momentum of the particle labeled a ($\alpha,\beta,... = 0,1,2,3$). We choose a surface

$$\Sigma\colon\ F_{a'}(k_b{}^\mu k_{c\mu}) = 0 \quad (a',b',... = 2,3,...,N) \tag{2.4}$$

invariant under Lorentz transformations

$$k'_a{}^\alpha = L^\alpha{}_\beta k_a{}^\beta, \tag{2.5}$$

We focus our attention on those surfaces that admit the parametric equations

$$k_{a'}{}^0 = f_{a'}(k_1{}^0, \mathbf{k}_b\cdot\mathbf{k}_c). \tag{2.6}$$

That is, the set $\{k_1{}^0,\mathbf{k}_a\}$ is a system of admissible coordinates for the surface Σ. We now introduce a new coordinate E defined implicitly by the equation

$$k_1{}^0 + \epsilon^{a'} f_{a'}(k_1{}^0,\mathbf{k}_b\cdot\mathbf{k}_c) = E, \tag{2.7}$$

This leads to the parametric form we use henceforth,

$$\begin{aligned} k_1{}^0 &= h_1(E,\mathbf{k}_b\cdot\mathbf{k}_c), \\ k_{a'}{}^0 &= h_{a'}(E,\mathbf{k}_b\cdot\mathbf{k}_c) \equiv f_{a'}[h_1(E,\mathbf{k}_d\cdot\mathbf{k}_e),\mathbf{k}_f\cdot\mathbf{k}_g]. \end{aligned} \tag{2.8}$$

The natural action of the Lorentz group (2.5) restricted to the invariant surface Σ defines a family of mappings of the surface onto itself which constitutes a nonlinear action of the Lorentz group ($i,j,... = 1,2,3$),

$$\begin{aligned} E' &= L^0{}_\alpha E + L^0{}_j P^j \quad (P^i = \epsilon^a k_a{}^i,\ \epsilon^a = 1), \\ k'_a{}^i &= L^i{}_0 h_a(E,\mathbf{k}_b\cdot\mathbf{k}_c) + L^i{}_j k_a{}^j. \end{aligned} \tag{2.9}$$

By considering the pullback under the mappings (2.9) of functions defined on the surface Σ, we get a representation of the Lorentz group on this function space. It reads

$$\begin{aligned} U_\kappa(L,0)\phi(E,\mathbf{k}_a) &= \phi[(L^{-1})^0{}_0 E + (L^{-1})^0{}_j P^j, (L^{-1})^i{}_0 h_d(E,\mathbf{k}_b\cdot\mathbf{k}_c) \\ &\quad + (L^{-1})^i{}_m k_d{}^m]. \end{aligned} \tag{2.10}$$

The next step will lead us to the desired result. We may regard functions $\phi(E,\mathbf{k}_a)$ as Fourier transforms of one-time functions[3]

$$\psi(t,\mathbf{x}_a) = \mathrm{Fourier}(\phi) \equiv (2\pi)^{-(3N+1)/2} \int_\Sigma \phi(E,\mathbf{k}_a) \times \exp[i(-Et + \mathbf{k}_b\cdot\mathbf{x}^b)]\, dE \wedge d\mathbf{k}_1 \wedge \cdots \wedge d\mathbf{k}_N. \tag{2.11}$$

Therefore, we may regard (2.10) as the representation operators in momentum space. We complete the representation by adding the space-time translation operators

$$U_\kappa(1,A)\phi(E,\mathbf{k}_a) = \exp[i(-EA^0 + \mathbf{P}\cdot\mathbf{A})]\phi(E,\mathbf{k}_a). \tag{2.12}$$

It can be easily proved that defining the generic representation operator by

$$U_\kappa(L,A) = U_\kappa(L,0)\,U_\kappa(1,A), \tag{2.13}$$

the group composition law (2.1) is satisfied. [There is an alternative expression to the group law we have chosen here; it is performed by changing the order of the operators on the right-hand side of (2.13).]

Finally, we define the representation operators in position space in the usual way; that is, if $\psi = \mathrm{Fourier}(\phi)$, then

$$U(L,A)\psi = \mathrm{Fourier}[U_\kappa(L,A)\phi]. \tag{2.14}$$

Let us remark that restricting (2.14) to space rotations and space-time translations one gets the well known operators used in nonrelativistic quantum mechanics to describe these Poincaré transformations.

We come back to the general case (nonzero spin particles) and end this section by writing the infinitesimal generators of the representation in momentum space. They read

$$\begin{aligned} T &= E, \quad P_i = \epsilon^a k_{ai}, \\ J_i &= S_i + i\epsilon_i{}^j{}_m k_a{}^m \frac{\partial}{\partial k_a{}^j}, \\ K_i &= Q_i + iP_i \frac{\partial}{\partial E} + ih_a(E,\mathbf{k}_b)\frac{\partial}{\partial k_a{}^i}. \end{aligned} \tag{2.15}$$

The operators S_i and Q_j are the infinitesimal generators of the representation of the Lorentz group (2.3). Their action on functions (1.2) can be expressed in terms of the infinitesimal generators of the Lorentz group representations on individual spin spaces; for instance,

$$\begin{aligned} S_i\psi &\equiv (S_i)^{(Q)}{}_{(P)}\psi_{(Q)}\chi^{(P)} \\ &= \sum_{a=1}^{N} \delta^{Q_1}{}_{P_1}\cdots(S_{ai})^{Q_a}{}_{P_a}\cdots\delta^{Q_N}{}_{P_N}\psi_{Q_1,...,Q_N}\chi^{P_1,...,P_N}. \end{aligned} \tag{2.16}$$

Since operators carrying different particle indices act on different spin spaces, they commute. Moreover, each pair S_{ai},Q_{aj} closes the Lie algebra of the Lorentz group,

$$\begin{aligned} [S_{ai},S_{aj}] &= i\epsilon_{ij}{}^k S_{ak}, \\ [S_{ai},Q_{aj}] &= i\epsilon_{ij}{}^k Q_{ak}, \\ [Q_{ai},Q_{aj}] &= -i\epsilon_{ij}{}^k S_{ak}. \end{aligned} \tag{2.17}$$

Obviously, S_i and Q_j have the same commuting relations. It is necessary to use some simple properties of the parametrization of the surface Σ to check up that (2.15) close to the Lie algebra of the Poincaré group

$$\begin{aligned} &[T,P_i] = [T,J_i] = [P_i,P_j] = 0, \\ &[P_i,J_j] = i\epsilon_{ij}{}^k P_k, \quad [J_i,J_j] = i\epsilon_{ij}{}^k J_k, \\ &[K_i,T] = iP_i, \quad [K_i,P_j] = i\delta_{ij}T, \\ &[K_i,J_j] = i\epsilon_{ij}{}^k K_k, \quad [K_i,K_j] = -i\epsilon_{ij}{}^k J_k. \end{aligned} \tag{2.18}$$

III. POINCARÉ INVARIANT SCHRÖDINGER EQUATIONS

We shall say that the Schrödinger equation (1.3) is Poincaré invariant if and only if the Poincaré transform

$\psi' = \mathbf{U}(L,A)$ of every solution ψ is also a solution for every operator $\mathbf{U}(L,A)$ of the representation (2.2). It ensures that inertial observes will all use the same Schrödinger equation to describe a definite N-particle system, even though they use different solutions of it to perform individual physical states of the system.

A straightforward consequence of the definition is the following set of necessary conditions:

$$\frac{\partial H}{\partial t} = 0, \tag{3.1a}$$

$$[P_i,H]\psi = [J_i,H]\psi = ([K_i,H] - iP_i)\psi = 0. \tag{3.1b}$$

The last equations only must hold if the functions ψ are solutions of the Schrödinger equation. Therefore, the Hamiltonian H does not coincide with the time translation operator T over the whole wave function space, but it does all over the Schrödinger equation solution subspace.

It is due to this annoying presence of the general solution of (1.3) in Eqs. (3.1b) that they are not very useful. Hence it would be desirable to come to a new set of equations where the solution dependence has been dropped out.

Schrödinger's equation determines without ambiguity the wave function at a time t provided we know it at a given time t_0. We can set $t_0 = 0$, since H does not depend explicitly on t. Thus we have

$$\psi(t) = \exp(-iHt)\psi(0). \tag{3.2}$$

Equation (3.2) defines a one-to-one correspondence between solutions and initial data. Then, if we assume (1.3) to be Poincaré invariant (i.e., Poincaré transform of solutions are also solutions), we can regard the action of the representation operators on the solutions of (1.3) as changes in the Cauchy data. This point of view amounts to considering a new representation of the Poincaré group on the initial data function space. We call it the induced representation. There are at least two ways to build up that representation [see Eq. (A12) of the Appendix]. They, of course, are equivalent. We have preferred here to proceed in the nonrelativistic quantum mechanics way: we regard (3.2) as the equation that makes the induced representation into the previous one (i.e., Schrödinger's representation into Heisenberg's representation). Thus the induced representation operators $\overline{\mathbf{U}}(L,A)$ read

$$\overline{\mathbf{U}}(L,A) = \exp(iHt)\mathbf{U}(L,A)\exp(-iHt). \tag{3.3}$$

A straightforward calculation gives the following expressions for the infinitesimal generators:

$$\begin{aligned}
&\overline{P}_i = P_i = -i\epsilon^a \frac{\partial}{\partial x_a{}^i} \equiv \epsilon^a p_{ai},\\
&\overline{J}_i = J_i = \epsilon_{ij}{}^k x^{aj} p_{ak} + S_i, \quad \overline{T} = H,\\
&\overline{K}_i = \sum_{n=0}^{\infty}\left(\frac{1}{n!}\right)\frac{\partial^n}{\partial E^n} h_a(0,\mathbf{p}_b\cdot\mathbf{p}_c) x^a{}_i H^n + Q_i.
\end{aligned} \tag{3.4}$$

The boost generator formula needs a more careful treatment. In fact, the above expression is rather formal and it cannot be applied blindly. We shall give a detailed discussion of it in the Appendix.

Since under a change of representation the commutator algebra is preserved, the necessary conditions on the Hamiltonian remain formally unchanged. However, we can suppress the function on which the commutators act. Because it is now a Cauchy data, there is not in principle any relevant constraint on this sort of function. Therefore, we get

$$[H,P_i] = [H,J_i] = 0, \quad [\overline{K}_i,H] = iP_i. \tag{3.5}$$

These equations are completely equivalent to (3.1) by means of purely algebraic calculations. Moreover, they ensure that H, P_i, J_j, and $\overline{K}_m$ close the Lie algebra of the Poincaré group. Therefore, if we assume that the conditions (3.5) are also sufficient, every solution H of (3.5) defines a Poincaré invariant Schrödinger equation.

We have no doubt that sufficiency may be proved by means of standard group techniques; that is, current procedures go from infinitesimal to finite transformations. Clearly, only transformations connected with the identity can be recovered by these procedures (i.e., neither space inversion nor time reversal).

IV. A NOTE ON EQUIVALENT REPRESENTATIONS

We have used an arbitrary Lorentz invariant surface (2.4) to build up the representation of the Lorentz group (2.10). It does not mean that we may obtain as many distinct representations as surfaces (2.4) are in the momentum space. Even though a more complete study of all these representations will deserve an entire paper, we proceed here to demonstrate the (local) equivalence of some of them. We shall make use of this result in the next section.

Let us consider two surfaces[4]

$$\begin{aligned}
&\Sigma_1\colon\ F_{1a'}(k^2{}_b) = 0 \quad (k^2{}_b = -k_b{}^\mu k_{b\mu}),\\
&\Sigma_2\colon\ F_{2a'}(q^2{}_b) = 0 \quad (q^2{}_b = -q_b{}^\mu q_{b\mu}).
\end{aligned} \tag{4.1}$$

Since the functions defining both surfaces do not depend on the scalar products of different momenta, they also define two curves in R^N. Let us call them C_1 and C_2. We assume that $C_1 \simeq C_2$; that is,

$$\{k^2{}_a\}\in C_1 \leftrightarrow \{q^2{}_b = \sigma_b(k^2{}_a)\}\in C_2 \tag{4.2}$$

is a diffeomorphism. We further assume that (4.2) does not change the number of $k^2{}_a$ that are positive, negative, or null. Roughly speaking, it conserves separately (and locally) the three possible kinds of particles: massive particles, zero mass particles, and tachyons. It is clear that (4.2) does not ever exist. However, we think that it may be defined piecewise in many cases. That will be enough for practical purposes, in which global properties of the representation can be ignored.

We define under the above assumption a one-to-one correspondence between the two surfaces that is compatible with the action of the Lorentz group (2.10) on each one; that is, if $\{k_a{}^\mu\}\leftrightarrow\{q_b{}^\rho\}$, then

$$\{L^\alpha{}_\mu k_a{}^\mu\}\leftrightarrow\{L^\phi{}_\rho q_b{}^\rho\}. \tag{4.3}$$

First, if $k^2{}_a > 0$, we move the point $k_a{}^\mu$ to the vertex of the hyperboloid $k^2{}_a = \text{const}$ by means of a pure Lorentz transformation; if $k^2{}_a < 0$, we perform a rotation to bring the point to the $\{k^0,k^1\}$ plane ($k^1 > 0$), and after that, we set $k_a{}^0 = 0$ by means of a boost rotation in that plane; if $k^2{}_a = 0$ ($k_a{}^\mu \neq 0$), we make a rotation as before, and after that we perform a boost in order to get $k_a{}^0 = \pm 1$, $k_a{}^1 = 1$. Let us call $L_a{}^\alpha{}_\beta$ to the Lorentz transformation we need in any of the three cases. Then, we have

$$k_a{}^\mu = \pm (L_a{}^{-1})^\mu{}_0 k_a \quad (k^2{}_a > 0),$$
$$k_a{}^\mu = + (L_a{}^{-1})^\mu{}_1 k_a \quad (k^2{}_a < 0), \qquad (4.4)$$
$$k_a{}^\mu = \pm (L_a{}^{-1})^\mu{}_0 + (L_a{}^{-1})^\mu{}_1 \quad (k^2{}_a = 0),$$

where $k_a = +|k^2{}_a|^{1/2}$ and the signs $+$ and $-$ in the first and third equation refer, respectively, to $k_a{}^0 > 0$ and $k_a{}^0 < 0$. We now use (4.2) to identify hyperboloids contained in Σ_1 with those in Σ_2. Finally, we apply the Lorentz transformations (4.4) to reach the point $\{q_b{}^\rho\}$ of Σ_2

$$q_b{}^\rho = k_b{}^{-1} |\sigma_b(k^2{}_a)|^{1/2} k_b{}^\rho \quad (k_a{}^2 \neq 0),$$
$$q_b{}^\rho = k_b{}^\rho \quad (k^2{}_a = 0). \qquad (4.5)$$

We recall that $k_a{}^\mu$ and $q_a{}^\rho$ belong to invariant surfaces. Therefore, we should rewrite (4.5) by making explicit the parametrization; it leads to

$$E_2 = \epsilon^a k_a{}^{-1} |\sigma_a(k^2{}_b)|^{1/2} h_{1a}(E_1, \mathbf{k}^2{}_c),$$
$$q_a{}^i = k_b{}^{-1} |\sigma_b(k^2{}_a)|^{1/2} k_a{}^i, \qquad (4.6)$$

where $k^2{}_a = h^2{}_{1a}(E_1, \mathbf{k}^2{}_c) - \mathbf{k}^2{}_a$, and we have taken $k_b{}^{-1} |\sigma_b(k^2{}_a)|^{1/2} = 1$ when $k_b{}^2 = 0$.

The way in which (4.6) has been constructed ensures that this correspondence verifies the condition (4.3). It implies that the representations we associate with Σ_1 and Σ_2 are equivalent in the usual sense.

V. CONSTRAINT THEORY

The theory we have developed throughout the preceding sections may be considered to some extent as the quantum counterpart of the classical predictive relativistic mechanics in its manifestly predictive form (i.e., Newtonian-like equations of motion). On the other hand, constraint theory can be regarded as the result upon quantization of the same classical theory, but in its manifestly covariant form.[5] One proceeds in this later case to make the N constants of motion that express the invariance in time of the particle masses into linear operators by means of the covariant correspondence principle. The result is a set of N Klein–Gordon equations (constraints) to determine the N-time wave functions of the physical system.[6] This suggests that there must be some relationship between both quantum theories since they seem to come from the same classical theory.

We consider henceforth surfaces of the kind (4.1). It is convenient for the following to introduce some constant in the definition of the surface:

$$\Sigma: \quad F_{a'}(k^2{}_b) - c_{a'} = 0. \qquad (5.1)$$

We define the lift of a function $\phi(E, \mathbf{k}_a)$ as a distribution on the whole momentum space (we can omit the spin wave function without loss of generality),

$$\phi_\uparrow(k_a{}^\mu) = \prod_{a'=2}^{N} \delta(F_{a'} - c_{a'}) \phi(\epsilon^b k_b{}^0, \mathbf{k}_c). \qquad (5.2)$$

We recover the function $\phi(E, \mathbf{k}_a)$ by using the simple inversion formula,

$$\int \phi_\uparrow(E, \mathbf{k}_c, C_{a'}) dC_2 \wedge \cdots \wedge dC_N = \phi(E, \mathbf{k}_b), \qquad (5.3)$$

where $\{E = \epsilon^a k_a{}^0, C_{a'} = F_{a'}(k^2{}_b), \mathbf{k}_d\}$ is an alternative system of coordinates for the momentum space. It turns out that the lift (5.2) is compatible with the natural action of the Poincaré group on functions defined in momentum space

$$U_{ST}(L,A)\phi_\uparrow(k_b{}^\mu) = \exp[i\epsilon^a (L^{-1})^\alpha{}_\beta k_a{}^\beta A_\alpha] \phi_\uparrow[(L^{-1})^\rho{}_\mu k_a{}^\mu], \qquad (5.4)$$

that is, the following condition is verified:

$$U_{ST}(L,A)\phi_\uparrow(k_b{}^\mu) = \{U_\kappa(L,A)\phi(E, \mathbf{k}_a)\}_\uparrow. \qquad (5.5)$$

Therefore, there is a strong correlation between the action of the group in the two levels. It allows us to operate with the group whether up (the lifted function space) or down (Σ-function space) and at the end to translate the result into the language of one or another function space by projecting or lifting.

The lifted function space can be obviously defined in terms of the solutions of the system of equations,

$$\{F_{a'}(k^2{}_a) - c_{a'}\}\phi_\uparrow(k_b{}^\alpha) = 0. \qquad (5.6)$$

This starts to seem like the usual constraint theory. However, we expect that the constraints have the form of Klein–Gordon equations and (5.6) do not actually have it. In fact, it will be seen later that (5.6) are consequences of the fundamental system of constraints.

Let us assume that the functions $h_a(E, \mathbf{k}^2, c_{d'})$ are analytic in E (this requirement can be relaxed to include functions having a pole at $E = 0$ by means of the techniques shown in the Appendix). If we only consider solutions of the Poincaré invariant Schrödinger equation,

$$H\phi(E, \mathbf{k}_a) = E\phi(E, \mathbf{k}_a), \qquad (5.7)$$

we can write the identity

$$H_a(\mathbf{k}^2{}_b, c_{d'}, H)\phi \equiv \sum_{n=0}^{\infty} \left(\frac{1}{n!}\right) \frac{\partial^n h^2{}_a}{\partial E^n}(0, \mathbf{k}^2{}_c, c_{g'}) H^n \phi = h^2{}_a(E, \mathbf{k}^2{}_d, c_{f'})\phi. \qquad (5.8)$$

Thus we obtain

$$\Omega_a \phi \equiv (H_a - \mathbf{k}^2{}_a)\phi = (h^2{}_a - \mathbf{k}^2{}_a)\phi. \qquad (5.9)$$

We apply the lift (5.2) to these identities in order to get another set that is verified by the lifted Schrödinger equation solutions. We define the lifted operators by

$$\Omega_{a\uparrow}\phi_\uparrow \equiv \prod_{b'=2}^{N} \delta(F_{b'} - c_{b'})\Omega_a \phi. \qquad (5.10)$$

This does not yield a unique operator. For instance, we may add to $\Omega_{a'}$ a sum of functions (5.1) multiplied by arbitrary nonconstant coefficients. However, its action on lifted functions is well defined. Moreover, (5.10) ensures that $\Omega_{a\uparrow}$ commutes with the $F_{a'}$'s all over the lifted function space.

The lift of (5.9) reads

$$\{k^2{}_a - \Omega_{a\uparrow}\}\phi_\uparrow = 0. \qquad (5.11)$$

We remark that all the operators Ω_a (resp. $\Omega_{a\uparrow}$) are not independent. In fact, we can use the functions defining the surface to express some of them as functions of the others. It will be in general rather formal, since the $F_{a'}$'s may be very complicated functions of the $k^2{}_a$'s. Nevertheless, we can use the equivalence among representations we have shown in the preceding section to come to a surface defined by very simple functions. It will allow us to give a definite and easy form to the above-mentioned relationships. Let us choose the following surface:

$$k^2_{a'} - k^2_1 = c_{a'} \equiv m^2_{a'} - m^2_1, \tag{5.12}$$

where $m_a > 0$ is the mass of the particle labeled a. Independently of the values of the masses, it always exists a domain in R^N where all the k^2_a's are positive and verify (5.12); that is, (5.12) can be used to describe a system of N massive particles.

Subtracting the identity (5.9) referred to the particle labeled 1 to those referred to any other particle and taking into account (5.12), we get

$$\{\Omega_{a'} - m^2_{a'}\}\phi = \{\Omega_1 - m^2_1\}\phi \equiv V\phi. \tag{5.13}$$

Substituting this result in (5.11), we obtain

$$\{k^2_a - m^2_a - V_\uparrow\}\phi_\uparrow = 0. \tag{5.14}$$

It is obvious that the equations defining the lifted function space (5.6) are contained in (5.14).

We may summarize the situation as follows: the lift of every solution of the Poincaré invariant Schrödinger equation (5.7) verifies the system of N Klein–Gordon equations (5.14). A simple argument based on the relativistic invariance of (5.7) and on the definition of lifted functions (5.2) and operators (5.10) allows us to partially prove some properties of the Klein–Gordon system. First, the system (5.14) is Poincaré invariant. Second, Klein–Gordon operators having different particle indices commute. Unfortunately, Poincaré's invariance is only true when the operators act on functions such that their projections are solutions of the Schrödinger equation. Since the space of solutions of (5.14) is larger than the set of functions generated by solutions of (5.7), this property does not hold to the extent we judge necessary to our purpose. We conjecture as a possible solution to this problem that several Poincaré invariant Hamiltonians define the same operator V in such a way that the space of solution of (5.14) is filled with lifted solutions of the Schrödinger equations associated to these Hamiltonians. Roughly speaking, given an operator V we search the operators H which verify (5.13). The second property is valid for a larger function space: the lifted function space. However, we think that the function space on which the commutators is zero might be extended to cover all the momentum function space by using the freedom we have to define $V_\uparrow$ outside of the lifted function space. This would be in full agreement with the strong commuting condition that is assumed in the constraint approach.

Let us finish this section by making some comments on the lift (5.2) and the projection (5.3) from the point of view of the position space. The Fourier transform of a lifted function gives an N-time function,

$$\psi_\uparrow(x_a{}^\mu) \equiv (2\pi)^{-2N} \int \phi_\uparrow(k_b{}^\rho)\exp(ik_{c\beta}x^{c\beta})dk_1 \wedge \cdots \wedge dk_N$$
$$= (2\pi)^{-2N}\int_\epsilon \phi(E,\mathbf{k}_b)\exp[-ih_c(E,\mathbf{k}_d)x^{c0} + i\mathbf{k}_c\cdot\mathbf{x}^c]J^{-1}(E,\mathbf{k}^2_e)dE \wedge d\,\mathbf{k}_1 \wedge \cdots \wedge d\,\mathbf{k}_N, \tag{5.15}$$

where

$$J^{-1}(E,\mathbf{k}^2_e) = \left|\frac{\partial(k_a{}^0)}{\partial(E,C_{b'})}\right|_{C_{d'}=c_{d'}}. \tag{5.16}$$

The natural projecting process consists of taking $x_a{}^0 = t$,

$$\psi(t,\mathbf{x}_a) = (2\pi)^{-2N}\int_\Sigma J^{-1}(E,\mathbf{k}^2_e)\phi(E,\mathbf{k}_b) \times\exp[i(-Et + \mathbf{k}_a\cdot\mathbf{x}^a)] \times dE \wedge \sim d\,\mathbf{k}_1 \wedge \cdots \wedge d\,\mathbf{k}_N. \tag{5.17}$$

This is consistent with the projection defined in momentum space (5.3). Now, functions in position and momentum space are related by the invariant Fourier transform. [$J^{-1}(E,\mathbf{k}_a)dE \wedge d\,\mathbf{k}_1 \wedge \cdots \wedge d\mathbf{k}_N$ is a volume element invariant under the projected action of the Lorentz group (2.10).]

The use of the invariant Fourier transform instead of the usual one changes the expression of the operators of the representation in the position space. However, it leads to a representation equivalent to that we defined in (2.14). In fact, we can maintain (2.14) to define the representation in the position space and to make the following change of representation in the momentum space:

$$\phi(E,\mathbf{k}_a) \to J^{-1}(E,\mathbf{k}^2_e)\phi(E,\mathbf{k}_b). \tag{5.18}$$

The new infinitesimal generators that result from applying (5.18) to (2.15) are equal to the old ones except for the boost generator,

$$K_i = Q_i + iP_i\frac{\partial}{\partial E} + i\frac{\partial}{\partial k_a{}^1}h_a(E,\mathbf{k}^2_b), \tag{5.19}$$

where the last term on the right-hand side must be reordered. It amounts to moving the $x_a{}^i$'s to the first place on the left-hand side of (3.4). Obviously, the Hamiltonian is also changed, $H \to J^{-1}HJ$. Nevertheless, the new H must verify the system of equations (3.5), provided we substitute the old generators by the new ones on it.

VI. FINAL REMARKS

It is well-known that probabilities are the natural language of quantum mechanics. All the information that a quantum theory supplies about the results of experiments involves probabilities. On mathematical grounds, it amounts to introducing a scalar product in the wave function space (i.e., to give it a structure of Hilbert space). Moreover, relativistic invariance demands that probabilities must be inertial observer independent. In other words, the physical scalar product must be invariant under the representation of the Poincaré group that we use to describe changes of system of reference.[7] In our case, it implies

$$(\psi_1,\psi_2) = (\mathbf{U}(L,A)\psi_1,\mathbf{U}(L,A)\psi_2) \tag{6.1}$$

for every pair of solutions of the Schrödinger equation (1.3) and every operator of the representation (2.2). It implies that $\mathbf{U}(L,A)$ are unitary operators with respect to the relativistic scalar product. Let us remark that (6.1) holds just for solutions of the Schrödinger equation, not for any pair of one-time functions, as it occurs in nonrelativistic quantum mechanics (e.g., invariance under time translations).

Let us emphasize one aspect of the scalar product question which may be useful in order to determine it. We may essay to define it either in the one-time function space (strictly speaking, in the subspace of Schrödinger equation solutions), as we suggest above, or in the initial data space. It does not matter, since both have the same functional form. Let us assume that we know the invariant scalar product in the initial data space. It can be extended to include functions depending on time if we regard t as a parameter. This prescription defines a scalar product which verifies (6.1) as a straightforward consequence of its invariance under the induced representation.

It might be surprising that the relativistic scalar product shares this property with the scalar product of the nonrelativistic theory. However, a deeper insight shows that it is a consequence of the time translation invariance of the Schrödinger equation, which is a property that both theories have in common.

Unfortunately, we have not yet succeeded in finding such a scalar product in the general case.[8] Nevertheless, it can be determined exactly for free particle systems, and approximately, up to the post-Newtonian order, for the N interacting particle system (both cases will be extensively treated in a subsequent paper).

ACKNOWLEDGMENTS

One of the authors (E. R.) wishes to thank the Institute Henri Poincaré for its hospitality.

His work was supported by a grant from the Ministerio de Educación y Ciencia under the Plan de Formación de Personal Investigador scheme.

APPENDIX: THE INFINITESIMAL GENERATORS OF THE INDUCED REPRESENTATION

One of the difficulties arising in the calculation of the infinitesimal generators of the induced representation comes from the lack of knowledge of the commutators (3.1b) out of the space of solutions of the Schrödinger equation. It compels us to be careful in the use of (3.1b), checking at each time if they are applied or not to a Schrödinger equation solution. The space-time translation and the spatial rotation generators are easy to evaluate. The boost generator adds to the above mentioned difficulty some other problems related to the definition of the operator in the position space. We only deal with this generator henceforth.

According to (3.3),

$$\overline{K}_i\psi(0) = \exp(iHt)K_i\psi(t) = \sum_{n=0}^{\infty}\left(\frac{i^n}{n!}\right)t^nH^nK_i\psi(t), \tag{A1}$$

where we have expanded the exponential and used that $[t,H]=0$. Taking into account that if ψ is a solution of the Schrödinger equation, and $H\psi$ is a solution as well, we get by means of applying (3.3) repeatedly,

$$H^nK_i\psi(t) = \{K_iH^n - inP_iH^{n-1}\}\psi(t). \tag{A2}$$

Substituting (A2) in (A1), we obtain

$$\overline{K}_i\psi(0) = \sum_{n=0}^{\infty}\left(\frac{i^n}{n!}\right)t^nK_iH^n\psi(t) + tP_i\psi(0). \tag{A3}$$

We now use the following remarkable identity:

$$t^nK_i = \sum_{r=0}^{n}\binom{n}{r}[t,K_i]^{(r)}t^{n-r}, \tag{A4}$$

where we have introduced the definition,

$$[t,K_i]^{(r)} \equiv [t\,[t\cdots(n\ \text{times})\cdots[t,K_i]\cdots]] \quad (r>0),$$
$$[t,K_i]^{(0)} \equiv K_i. \tag{A5}$$

A straightforward calculation leads to the formula

$$\overline{K}_i\psi(0) = \sum_{r=0}^{\infty}\left(\frac{i^r}{r!}\right)[t,K_i]^{(r)}H^r\psi(0) + tP_i\psi(0). \tag{A6}$$

The commutators (A5) can be easily calculated in the momentum space, where the operator t becomes $-i\,\partial/\partial E$. Then, we have

$$[t,K_i]^{(r)} = (-i)^r\frac{\partial^r h_a}{\partial E^r}(E,\mathbf{k}_b\cdot\mathbf{k}_c)i\frac{\partial}{\partial k_a{}^i}. \tag{A7}$$

The derivatives of $h_a(E,\mathbf{k}_b\cdot\mathbf{k}_c)$ are well defined operators in momentum space. We denote them as operators in the position space by substituting the momentum space variables E, $\mathbf{k}_a$ by, respectively, the operators T, $\mathbf{p}_a$. This formal notation hides sometimes a lack of definition of these operators in the position space. We shall briefly discuss it in some cases.[9]

First of all, if the functions $h_a(E,\mathbf{k}_b\cdot\mathbf{k}_c)$ are analytic in a neighborhood of $E=0$, we can use its Taylor expansion at $E=0$ to define its derivatives with respect to E as operators in the position space. This amounts to substituting E by $i\,\partial/\partial t$ in its Taylor series. Since $\overline{K}_i$ acts on functions independent of t, all the terms of the expansion except the first one, which does not include $i\,\partial/\partial t$, becomes zero. That is we implicitly assumed when we wrote (3.4).

A more realistic case allows $h_a(E,\mathbf{k}_b\cdot\mathbf{k}_c)$ to have a pole at $E=0$.[10] Then, (3.4) is meaningless. Let us consider any negative power of E in the Laurent expansion of $h_a(E,\mathbf{k}_b\cdot\mathbf{k}_c)$ at $E=0$, and evaluate its contribution to $\overline{K}_i$ (we now work in momentum space),

$$\overline{K}^{(s)i} = \theta_a{}^{(s)}(\mathbf{k}_b\cdot\mathbf{k}_c)i\frac{\partial}{\partial k_a{}^i}\sum_{n=0}^{\infty}\left(\frac{1}{n!}\right)\frac{\partial n}{\partial E^n}(E^{-s})H^n. \tag{A8}$$

The term $\theta_a{}^{(s)}(\mathbf{k}_b\cdot\mathbf{k}_c)$ is the Laurent coefficient of the term E^{-s} $(s>0)$. We sum up the series that appears in the right-hand side of (A8) by means of the following formula valid for numbers:

$$\sum_{n=0}^{\infty}\left(\frac{1}{n!}\right)\frac{\partial^n}{\partial E^n}\left(\frac{1}{E^s}\right)H^n$$
$$=(E+H)^{-s}$$
$$=\int_{-\infty}^{0}du_s\int_{-\infty}^{u_s}du_{s-1}\cdots\int_{-\infty}^{u_2}du_1\exp[u_1(E+H)] \tag{A9}$$

$[E+H>0$; if negative we must change the intervals of integration into $(\ ,+\infty)]$. Since the final result is meaningful when E and H are operators ($[E,H]=0$), we accept that (A9) also holds in this case. We now go to the position space

$$\overline{K}^{(s)i} = \theta_a{}^{(s)}(\mathbf{p}_b\cdot\mathbf{p}_c)x^s{}_i\int_{-\infty}^{0}du_s\int_{-\infty}^{u_s}du_{s-1}\cdots\int_{-\infty}^{u_2}du_1$$
$$\times\exp(u_1H). \tag{A10}$$

We have set $E=0$ for the same reason as we did before.

Even though it might be that (A9) is rather formal, the expression for $\overline{K}_i$ that is given by an alternative but equivalent definition of this operator

$$\overline{K}_i\psi(0)\equiv\{K_i\psi(t)\}_{t=0} \tag{A11}$$

is exactly the same.

Finally, let us point out that the same trick can be used to define (5.8) when $h_a(E,\mathbf{k}_b\cdot\mathbf{k}_c)$ has a pole at $E=0$. For example, if ϕ is a solution of (5.7), we have the following identity:

$$\int_{-\infty}^{0} du_s \int_{-\infty}^{u_s} du_{s-1}\cdots\int_{-\infty}^{u_2} du_1 \exp(u_1 H)\phi = E^{-s}\phi. \tag{A12}$$

This permits us to define $H_a(\mathbf{k}^2{}_b,c_{a'},H)$ by using the Laurent expansion of $h_a^2(E,\mathbf{k}^2{}_d,c_{f'})$.

[1]See, for instance, A. Komar, Phys. Rev. D **18**, 1887 (1978); P. Droz-Vincent, *ibid.* **19**, 702 (1979); L. P. Horwitz and F. Rohrlich, *ibid.* **24**, 1528 (1981); V. A. Rizov, M. Sazdjian, and I. T. Todorov, Ann. Phys. (NY) **165**, 59 (1985); G. Longhi and L. Lusanna, Phys. Rev. D **18**, 3707 (1986). A noncovariant approach can be found in F. Coester and W. N. Polyzou, Phys. Rev. D **26**, 1348 (1982). Systems of two spin-$\frac{1}{2}$ particles have been studied by H. Sazdjian, Phys. Rev. D **33**, 3401 (1986); H. W. Crater and P. Van Alstine, J. Math. Phys. **23**, 1697 (1982); Phys. Rev. Lett. **53**, 1527 (1984). See also W. N. Polyzou, Phys. Rev. D **32**, 995 (1985).

[2]The possible connection with two spin-$\frac{1}{2}$ particle theories based on a couple of Dirac equations will not be considered here.

[3]$d\mathbf{k}_a \equiv dk_a{}^1\wedge dk_a{}^2\wedge dk_a{}^3$.

[4]The signature of the Minkowski metric is $+2$.

[5]L. Bel and J. Martín, Ann. Inst. H. Poincaré A **33**, 409 (1980); A **34**, 231 (1981).

[6]L. Bel, Phys. Rev. D **28**, 1308 (1983).

[7]E. P. Wigner, Ann. Math. **40**, 141 (1939); V. Bargmann, *ibid.* **59**, 1 (1954).

[8]An attempt of defining such scalar products provided that the Hamiltonian verifies some conditions has been made by L. Bel and E. Ruiz, in *Constraint's Theory and Relativistic Dynamics*, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987).

[9]An alternative way to deal with this problem is the deformation algebra theory. This has been used in the paper cited in Ref. 8. For additional information see L. Bel, in *Contribution to Differential Geometry and Relativity*, edited by M. Cahen and M. Flato (Reidel, Dordrecht, Holland, 1976).

[10]That is the situation which appears in the two spinless particle problem (see Ref. 8).

Generalized exponential, circular, and hyperbolic functions for nonlinear wave equations

J. Díaz Bejarano and A. Martín Sánchez
Departamento de Fisica, Universidad de Extremadura, 06071 Badajoz, Spain

(Received 2 July 1987; accepted for publication 6 April 1988)

Wave functions are presented in the form of generalized exponentials that are solutions of some of the most usual linear and nonlinear wave equations. The solutions are given in terms of the elliptic functions of Jacobi and presented in a form as similar as possible to the usual circular functions. Some simple theorems are demonstrated to present the solutions as the simplest possible extension of the usual exponentials.

I. INTRODUCTION

The construction of a simple model of nonlinear quantum mechanics (QM) should probably clarify the relationships between QM and general relativity. It is also well-known that the simplest equations for the harmonic oscillator in special relativity[1] are equivalent to nonlinear differential equations of the type

$$\dot{x}^2 + Ax^2 + Bx^4 = E, \quad \dot{x} = \frac{dx}{dt}, \tag{1}$$

and in consequence the solutions can be given in terms of the Jacobian elliptic functions. Due to this nonlinear character, these oscillators have played no role in the development of the usual linear QM. The equations for the orbits of particles in general relativity can be transformed to a one-dimensional problem of the type

$$\dot{x}^2 + P(x) = E, \tag{2}$$

where $P(x)$ is a third-order polynomial. The solutions can be given in a very simple form in terms of Weierstrassian or, equivalently, squares of Jacobian elliptic functions.[2] We know of only one previous attempt[3] to construct a simple model of nonlinear QM using, for the fundamental oscillators, the simplest possible nonlinear extension of the usual linear fundamental oscillators, i.e., oscillators with equations of motion of types (1) or (2).

In this paper we deal with one of the many problems that have to be solved. We present wave functions in the form of simple generalized exponentials that are solutions to some of the most usual linear and nonlinear wave equations. The solutions are given in terms of the Jacobi elliptic functions [solutions of Eq. (1)], so that the results may be presented in a form very similar to the standard presentation in terms of the usual circular, hyperbolic, and exponential functions that are the solutions of Eq. (1) when $B = 0$.

The problem we have solved is how to find, in a systematic way, solutions of the type $x = \exp\phi(t)$ to equations of types (1) and (2) that are a simple extension of the usual exponentials and can be used to construct wave functions that are solutions of certain linear and nonlinear wave equations. The exponentials can also be used to find the corresponding generalized circular and hyperbolic functions. This opens the way to generalized Fourier series[4] and transforms, useful also for finding approximate solutions to problems in other branches of physics, e.g., the classical mechanics of mass points. Certain plane-wave solutions using Jacobian elliptic functions that correspond to some of our generalized exponential, circular, or hyperbolic functions are given in the paper of Petiau,[3] but in a nonsystematic way.

Our exponentials can be used, for example, in solutions for certain partial differential equations (PDE's), namely, (a) the usual linear wave equation,

$$\Box\psi = 0,$$

where our functions are particular cases of the arbitrary solutions $f(x \pm ct)$; (b) the nonlinear generalization of the Klein–Gordon equation,

$$\Box\psi + A_2\psi + A_4\psi^3 = 0; \tag{3}$$

(c) the Korteweg–de Vries equation,

$$\psi_t + (c_0 + c_1\psi)\psi_x + \nu\psi_{xxx} = 0;$$

and (d) the sine–Gordon equation,

$$\Box\psi = \sin\psi,$$

$$\psi_x = \frac{\partial\psi}{\partial x}, \quad \psi_t = \frac{\partial\psi}{\partial t},$$

$$\Box\psi = (1/c^2)\psi_{tt} - (\psi_{xx} + \psi_{yy} + \psi_{zz}).$$

We shall consider only the generalized exponential solutions for the type (a) of Petiau, i.e., for the plane-wave solutions

$$\psi(\mathbf{x},t) = \psi(\theta),$$

where the phase is

$$\theta = \omega t \pm (\mathbf{k}\cdot\mathbf{x}).$$

We are mostly interested in stationary solutions, i.e., waves that propagate with constant velocity and without deformation of the profile. In this case the phase velocity is

$$u = \omega/k = \text{const},$$

and, given a profile at $t = 0$, the perturbation $\psi_0(\mathbf{x},0)$ propagates with constant velocity u without change of form.

When, as in the cases quoted and many others, the evolution equation admits a stationary solution, the calculation is simple because we pass from the PDE to the ODE (ordinary differential equation) as follows.

In $l + 1$ dimensions,

$$\theta = x - ut = \text{const}.$$

 0022-2488/88/081847-07$02.50

Then

$$\frac{\partial^n\psi}{\partial t^n}=(-u)^n\frac{d^n\psi}{d\theta^n},\quad \frac{\partial^n\psi}{\partial x^n}=\frac{d^n\psi}{d\theta^n},$$
$$\frac{\partial^n\psi}{\partial t^s\,dx^{n-s}}=(-u)^s\frac{d^n\psi}{d\theta^n}. \tag{4}$$

We are interested in evolution equations that transform to ODE's with solutions in terms of Jacobian elliptic functions and, in order to simplify the discussion, to equations such as Eq. (3) that, with Eqs. (4), give

$$(u^2-1)\psi_{\theta\theta}+A_2\psi+2A_4\psi^3=0$$

or

$$\tfrac{1}{2}(u^2-1)\psi_\theta^2+A_2\psi^2+A_4\psi^4=E',$$

i.e.,

$$\psi_\theta^2+A\psi^2+B\psi^4=E,$$

where $\psi_\theta=d\psi/d\theta$ and A,B,E are constants, i.e., equations of type (1).

In spite of many efforts, simple formulas for the energy levels or transmission coefficients of the quantum anharmonic symmetrical oscillators (ASO's) and anharmonic asymmetrical oscillators (AAO's) are not known. The formulas that we and many other authors have found[5,6] are approximate and awkward. In order to help find simple nonlinear momentum and energy operators for these oscillators, we shall give relations between the nonlinear equations of type (1) and the formulas for the relativistic harmonic oscillator.

One interesting result we shall show for our nonlinear solutions, in sharp contrast with the usual circular functions, is that they can be used not only as generalized circular or hyperbolic functions of a given angle, but also as generalized exponentials of a different angle.

There are many obstacles to be overcome in achieving a simple model for nonlinear QM: one must find exact energies and wave functions for the fundamental levels of the oscillator, simple and useful lowering and raising operators, simple relationships with the corresponding covariant space-time derivative, agreement between theoretical formulas and physical experiments, etc. Nevertheless there are two principal advantages to the use of elliptic functions for the fundamental oscillators: (a) many of the formulas reduce to the usual ones when the parameter of the elliptic functions m is set to zero; and (b) using our generalized exponential, circular, and hyperbolic functions, the theory can probably be constructed in a way similar to the usual linear QM.

Our formulas can be useful in the field of classical mechanics, helping to solve in exact or approximate ways the very old, unresolved problems of the damped, forced, etc., ASO's and AAO's.[7] We shall not insist in this paper on the problems of nonlinear QM and PDE's. We shall give first some simple theorems on scaling, reciprocals, and logarithmic derivatives, necessary for a systematic construction of the generalized exponential solutions of Eq. (1). Using these exponentials, we shall define the corresponding generalized circular and hyperbolic functions that are also solutions of Eq. (1) and the generalized exponentials that are solutions to equations of type (2). The final section gives the relationships with special relativity and the space-time curvature of world lines.

II. THEOREMS

A. Scaling

If R and ω are nonzero constants, then x is a solution of the ODE

$$x'^2+Ax^2+Bx^4=E(s),\quad x'=\frac{dx}{ds},\quad s=\omega t \tag{5}$$

if and only if $y(t)=Rx(\omega t)$ is a solution of the ODE

$$\dot y^2+A\omega^2y^2+B(\omega^2/R^2)y^4=R^2\omega^2E(\omega t),\quad \dot y=\frac{dy}{dt}.$$

The proof of this theorem is trivial. Some conclusions can be drawn: if $+Rx$ is a solution of the ODE (5), then $-Rx$ is a solution of the same equation; in the accompanying tables a $\pm$ sign could be written before any solution.

Three particular cases of the theorem are $R=i$; $\omega=i$; $R=i$ and $\omega=i$. If $A=0$, then $R=i^{1/2}$, $\omega=+i^{1/2}$ gives $\dot y^2+By^4=-E$, and $R=-i^{1/2}$, $\omega=+i^{1/2}$ gives $\dot y^2+By^4=+E$. The latter two cases are related to scaling of the parameter k that is only simple for $A=0$ ($m=k^2=m_1=k'^2=\frac{1}{2}$), with $m(k)$ being the parameter of the elliptic functions[8] that are solutions of Eq. (5) if $E=$ const, i.e., Eq. (1).

B. Reciprocals

One has that x is a solution of the ODE

$$\dot x^2+Ax^2+Bx^4=E(t)$$

if $y=1/x$ is a solution of the ODE

$$\dot y^2+Ay^2-E(t)y^4=-B.$$

The proof is also obvious. Neither the scaling nor the reciprocal theorem requires the assumption $E(t)=$ const, which would be the special case of the elliptic functions, solutions of Eq. (1).

If $y=i/x$, then using the scaling theorem,

$$\dot y^2+Ay^2+Ey^4=+B.$$

Then if, for an ODE type (5), $E=+B$, then x and i/x are solutions of the same ODE.

If $E=-B$, then x and $1/x$ are solutions of the same ODE.

The scaling and reciprocal theorems are very useful in constructing and checking the tables of solutions for the different signs of A,B,E in the different regions of interest. In Table I we give as an example the solutions for the different regions of x. The two types of solutions given are useful for different initial conditions. From the four possible sign combinations we have selected only one. More details and physical applications are given in Refs. 4–7. Given A,B,E, the region of interest, etc., it is easy to find the values of R, ω, and $m=k^2$ from the formulas of Table I. This is shown for some of the pairs of solutions in Table II. More details and physical applications can be found in the references given above. Observe that the pairs (sn,cd), (k'sd,cn) reduce to (sin,cos) for $m=0$, $m_1=1-m=1$.

TABLE I. Solutions of the equation $\dot{x}^2 + Ax^2 + Bx^4 = E$. Type $A \geqslant 0$, $B \geqslant 0$. In this table, $m = k^2$, $m_1 = 1 - m$; $k'^2 = 1 - k^2$; $V_m = -A^2/4B$; $E_R = E/V_m$; and $x_1 \leqslant x_2$ are the real roots of the equation $V(x) = E$.

Function $x(t)$	Conditions		$A \geqslant 0$	$B \geqslant 0$	E	E_R
$R\,\mathrm{cn}(\omega t;k)$ $Rk'\,\mathrm{sd}(\omega t;k)$ $(iR/k)\mathrm{ds}(\omega t;k)$ $(iRk'/k)\mathrm{nc}(\omega t;k)$	$x \leqslant x_1$, $\dot{x}^2 \geqslant 0$ } $E > V$ $x \geqslant x_2$, $\dot{x}^2 < 0$ } $E < V$	$m < \frac{1}{2}$; $E > 0$; $E_R \leqslant 0$	$(1-2m)\omega^2 = -B(x_1^2 + x_2^2)$	$m\omega^2/R^2$	$R^2\omega^2 m_1 = -Bx_1^2 x_2^2$	$-4mm_1(1-2m)^{-2}$
$iR\,\mathrm{sn}(\omega t;k)$ $iR\,\mathrm{cd}(\omega t;k)$ $(iR/k)\mathrm{ns}(\omega t;k)$ $(iR/k)\mathrm{dc}(\omega t;k)$	$x \leqslant x_1$ $\dot{x}^2 \leqslant 0$ $x \geqslant x_2$ $\dot{x}^2 < 0$	$E \leqslant V$; $V_m < E < 0$ $A^2 + 4BE > 0$	$(1+m)\omega^2$	$m\omega^2/R^2$	$-R^2\omega^2$	$4m(1+m)^{-2}$

TABLE II. Formulas for the oscillators $\dot{x}^2 + Ax^2 + Bx^4 = E$. The symbols are the same as in Table I. The value of R^2 as a function of A, B, and E can be obtained from the expression for ω^2.

Signs of constants	Solution x	ω^2	$m(A,B,R)$	$m(E)$
$A>0, B>0, E>0$	$R\,\mathrm{cn}(\omega t,k)$ $Rk'\,\mathrm{sd}(\omega t,k)$	$A + 2BR^2 = (A^2+4BE)^{1/2}$; $\omega^2 > A$	$BR^2/(A+2BR^2)$; $0<m<\frac{1}{2}$	$\frac{1}{2}\{1 - [(1-E_R)^{1/2}/(1-E_R)]\}$
$A>0, B>0, 0<E<V_m$	$R\,\mathrm{cd}(\omega t,k)$ $R\,\mathrm{sn}(\omega t,k)$	$A + BR^2 = \frac{1}{2}\{A + [A^2 + (BE)^{1/2}]\}$; $A/2 < \omega^2 < A$	$-BR^2/(A+BR^2)$; $0<m<1$	$[2 - E_R - 2(1-E_R)^{1/2}]E_R^{-1}$
$A<0, B>0, E>0$	$R\,\mathrm{cn}(\omega t,k)$ $Rk'\,\mathrm{sd}(\omega t,k)$	$A + 2BR^2 = (A^2+4BE)^{1/2}$; $\omega^2 > A$	$BR^2/(A+2BR^2)$; $\frac{1}{2}<m<1$	$\frac{1}{2}\{1 + [(1-E_R)^{1/2}/E_R]\}$
$A<0, B>0, V_m<E<0$	$R\,\mathrm{dn}(\omega t,k)$ $Rk'\,\mathrm{nd}(\omega t,k)$	$BR^2 = \frac{1}{2}\{-A + [A^2 + (BE)^{1/2}]\}$; $-A/2 < \omega^2 < -A$	$2 + (A/BR^2)$; $0<m<1$	$[2E_R - 2 + 2(1-E_R)^{1/2}]E_R^{-1}$

TABLE III. Logarithmic derivative solutions of the equation $\dot{x}^2 + Ax^2 + Bx^4 = E$. Type $A>0$, $B<0$. The symbols are the same as in Table I. Now $B = -R^2$ everywhere.

Function	A	E	E_R	$m(A,B,E)$	Conditions
$(m\omega\,\mathrm{sn}\,\omega t\,\mathrm{cn}\,\omega t)/(R\,\mathrm{dn}\,\omega t)$ $(m\omega\,\mathrm{sd}\,\omega t\,\mathrm{cd}\,\omega t)/(R\,\mathrm{nd}\,\omega t)$	$2\omega^2(2-m)$	$m^2\omega^4/R^2$	$[m/(2-m)]^2$	$2E_R^{1/2}(1+E_R^{1/2})^{-1}$	$0 \leqslant E \leqslant V_m$; $E_R \leqslant 1$; $A^2 + 4BE \geqslant 0$
$(\omega\,\mathrm{sn}\,\omega t\,\mathrm{dn}\,\omega t)/(R\,\mathrm{cn}\,\omega t)$ $(\omega\,\mathrm{cn}\,\omega t\,\mathrm{nd}\,\omega t)/(R\,\mathrm{sn}\,\omega t)$	$2\omega^2(2m-1)$	ω^4/R^2	$(2m-1)^{-2}$	$(2E_R^{1/2})^{-1}(1+E_R^{1/2})$	$m \geqslant \frac{1}{2}$; $V_m \leqslant E$; $E_R \geqslant 1$; $A^2 + 4BE \leqslant 0$

When $E < V_m = A^2/4B$ (Table I) there is no simple solution in terms of one of the 12 elliptic Jacobian functions with R, ω, and m real. Simple solutions for this and other cases exist in the form of logarithmic derivatives or generalized exponentials, as will be discussed below.

C. Logarithmic derivatives

Logarithmic derivatives (bipolar functions) are discussed in Refs. 9 and 10.

If x is a solution of the ODE (1), the logarithmic derivative $z \equiv \dot{x}/x$ is a solution of the ODE,

$$\dot{z}^2 - 2Az^2 - z^4 = A^2 + 4BE,$$

and using the scaling theorem, $Ry = \dot{x}/x$ is a solution of

$$\dot{y}^2 - 2Ay^2 - R^2y^4 = (A^2 + 4BE)/R^2. \tag{6}$$

The proof is not given because it is not very difficult.

Comparing Eqs. (1) and (6) gives

$$A_y = -2A_x, \quad B_y = -R^2, \\ E_y = (A_x^2 + 4B_xE_x)/R^2, \tag{7}$$

where subscripts x and y refer to Eqs. (1) and (6), respectively. To solve the inverse problem, define $x_+ \equiv e^{+\phi}$ and $x_- \equiv e^{-\phi} \equiv 1/e^{+\phi}$.

We are interested in functions of x that are solutions of Eq. (1). Using the reciprocal theorem one must have $E_x = -B_x$. Then from Eqs. (7),

$$R^2E_y = (A_y^2/4) - 4B_x^2,$$

i.e.,

$$(4B_x)^2 = (4E_x)^2 = A_y^2 - 4R^2E_y.$$

If $\bar{z} \equiv \dot{z}/z$, then $\bar{z}$ also satisfies an equation of type (1). If $\bar{\bar{z}} = \dot{\bar{z}}/\bar{z}$, then an infinite series of logarithmic derivatives satisfy equations of type (1).

Using the solutions in terms of the 12 Jacobian elliptic functions as given in Table I and the theorems and formulas previously discussed, it is easy to construct solutions in terms of logarithmic derivatives. One case is shown in Table III. We give only solutions with R,ω,m real, which are those of most interest. Formulas similar to the ones given in Table II are easily obtained for these solutions. More details and physical applications can be found in Ref. 6.

III. GENERALIZED EXPONENTIALS

If $B \neq 0$, Eq. (1) has no solutions of the type

$$x = Re^{\pm i\omega t} \quad \text{or} \quad x = Re^{\pm \omega t}. \tag{8}$$

These are solutions only if $B = 0$. However, we can find solutions of the type

$$x = Re^{\pm i\phi(t)} \quad \text{or} \quad x = Re^{\pm \phi(t)}. \tag{9}$$

These solutions are simple generalizations of Eq. (8), and we call them generalized exponentials. We shall give below a systematic derivation using the logarithmic derivatives. But it is more intuitive to present two of these functions that reduce to Eq. (8) if $B = 0$, i.e., the case in which the elliptic functions reduce to the usual circular or hyperbolic functions ($m = 0$ or $m = 1$).

Using systematic methods and formulas to be discussed below, we have found all the solutions necessary to solve Eq. (1) for different values of A,B,E, different regions, etc. An example of some of these solutions is given in Table IV. We again give the more interesting solutions, i.e., solutions with R, ω, and m real and with $0 \leqslant m \leqslant 1$. Formulas similar to the ones given in Table II can be easily obtained for these solutions. Observe that only two of the generalized exponential solutions given in Table IV are bounded on the real axis of the argument. These are the reciprocals of 9 and 12. Types 3 and 6 and the reciprocals of 9 and 12 can be written in the form that appears in the top line of Table III but with arguments $(\omega t + \alpha_l)/2$, where $l = 3,6,9,12$; by virtue of the Landen transformation[8] they are

$$\mathrm{sn}[(1+k')(\omega t + \alpha_l)/2;(1-k')/(1+k')],$$

$\alpha_3 = -iK', \alpha_6 = K - iK', \alpha_9 = K, \alpha_{12} = 0$. It is also easy to find the correspondence with our exponential forms, e.g.,

$$[k/(1+k')]\mathrm{sn}[\tfrac{1}{2}(1+k')\omega t;\ (1-k')/(1+k')] \\ = k\,\mathrm{sn}(\omega t/2)\mathrm{cn}(\omega t/2)/\mathrm{dn}(\omega t/2)$$

TABLE IV. Generalized exponential solutions of the equation $\dot{x}^2 + Ax^2 + Bx^4 = E$. Type $A<0$, $B>0$, $E<0$. The numbers in the first column are the types according to the definitions given in Table V.

Type	Function $x_i(t) = \mathrm{Re}^{\lambda\phi i}$	A	B	E	E_R	Conditions
7	$iR[\mathrm{sc}(\omega t;m_1) + \mathrm{nc}(\omega t;m_1)]$					
3	$R[\mathrm{sn}(i\omega t;m) + i\,\mathrm{cn}(i\omega t;m)]$					
4	$R[\mathrm{nd}(\omega t;m_1) + k'\,\mathrm{sd}(\omega t;m_1)]$					
6	$R[\mathrm{cd}(i\omega t;m) + ik'\,\mathrm{sd}(i\omega t,m)]$					
10	$(iR/k)[\mathrm{cs}(\omega t;m_1) + \mathrm{ds}(\omega t;m_1)]$	$(m-2)\omega^2/2$	$m\omega^2/4R^2$	$-mR^2\omega^2/4$	$[m/(2-m)]^2$	$E<V_m$; $0<E_R<1$
12	$(R/k)[\mathrm{ns}(i\omega t;m) + \mathrm{ds}(i\omega t;m)]$					
1	$(R/k)[\mathrm{dn}(\omega t;m_1) \pm k'\,\mathrm{cn}(\omega t;m_1)]$					
9	$(R/k)[\mathrm{dc}(i\omega t;m) + k'\,\mathrm{nc}(i\omega t;m)]$					
8	$iR[k\,\mathrm{sc}(\omega t;m_1) \pm \mathrm{dc}(\omega t;m_1)]$					
2	$R[k\,\mathrm{sn}(i\omega t;m) + i\,\mathrm{dn}(i\omega t;m)]$					
5	$R[k\,\mathrm{nd}(\omega t;m_1) \pm ik'\,\mathrm{cd}(\omega t,m_1)]$					
5	$R[k\,\mathrm{cd}(i\omega t;m) + ik'\,\mathrm{nd}(i\omega t;m)]$					
11	$iR[\mathrm{cs}(\omega t;m_1) \pm \mathrm{ns}(\omega t;m_1)]$	$(1-2m)\omega^2/2$	$\omega^2/4R^2$	$-R^2\omega^2/4$	$(2m-1)^{-2}$	$\tfrac{1}{2}<m<1$; $E>V_m$; $E_R>1$
11	$R[\mathrm{ns}(i\omega t;m) + \mathrm{cs}(i\omega t;m)]$					
2	$R[\mathrm{dn}(\omega t;m_1) \pm ik'\,\mathrm{sn}(\omega t;m_1)]$					
8	$R[\mathrm{dc}(i\omega t;m) + k'\,\mathrm{sc}(i\omega t;m)]$					

$$= (1/k)[\text{ns}\,\omega t - \text{ds}\,\omega t] = k/[\text{ns}\,\omega t + \text{ds}\,\omega t].$$

All the solutions given in Tables IV and V, and more to be discussed later, are solutions of type (9) of Eq. (1), but not many of them are valid when $B = 0$, and only the y_3 and y_6 solutions of Table IV tend to type (8):

$$\phi_3(u) \equiv \int x_3(u)du = \int \text{dn}\,u\,du,$$

$$y_3^{\pm} = e^{\pm i\phi_3(u)} = \text{cn}\,u \pm i\,\text{sn}\,u;$$

$$\phi_6(u) \equiv \int x_6(u)du = \int \text{nd}\,u\,du,$$

$$y_6^{\pm} = e^{\pm i\phi_6(u)} = \text{cd}\,u \pm ik'\,\text{sd}\,u.$$

If $B = 0$, then $k = 0$, $k' = 1$, $\phi_3(u) = \phi_6(u) = u$, and

$$\text{cn} = \cos, \quad \text{sn} = \sin, \quad \text{dn} = 1,$$

$$\text{cd} = \cos, \quad \text{sd} = \sin, \quad \text{nd} = 1.$$

Using, as in Tables I–III, $u = \omega t$,

$$\phi_3(\omega t) = \phi_6(\omega t) = \omega t,$$

and we have Eq. (8).

From well-known properties of the elliptic functions,[8]

$$\text{cn}(iu,k) = \text{nc}(u,k'), \quad \text{cn}(iu,0) = \text{nc}(u,1) = \cosh u,$$

$$\text{cd}(iu,k) = \text{nd}(u,k'), \quad \text{cd}(iu,0) = \text{nd}(u,1) = \cosh u,$$

$$\text{sn}(iu,k) = i\,\text{sc}(u,k'), \quad \text{sn}(iu,0) = i\,\text{sc}(u,1) = i\sinh u,$$

$$\text{sd}(iu,k) = i\,\text{sd}(u,k'), \quad \text{sd}(iu,0) = i\,\text{sd}(u,1) = i\sinh u,$$

and consequently the generalized exponentials y_3 and y_6 give, in the limit $B = 0$,

$$\exp(\pm\omega t) = \cosh\omega t + \sinh\omega t = e^{\pm\omega t}.$$

Some properties of the usual exponentials are *not* satisfied by the generalized exponentials, especially the very important property of linear addition, i.e., linear combinations of solutions of a given equation are *not* solutions of the same equation. Also, the parameter m for solutions of Table IV is not the same parameter m for the same E in Table I.

For a systematic search for generalized exponentials we use

$$y = \exp\left(R\int x(u)du\right) \quad \text{and} \quad x = \exp\left(R\int z(u)du\right),$$

where $x_1, x_2, \ldots, x_{12}$, $z_1, \ldots, z_{12}$ are given in Tables I–III. We have

$$(\ln y)' = (y'/y) = Rx, \quad (\ln x)' = (x'/x) = Rz.$$

Table IV is obtained from the equations given in Sec. II C. The integrals in Table V have been chosen to satisfy $e^{+\phi}\cdot e^{-\phi} = 1$, $e^{+i\phi}\cdot e^{-i\phi} = 1$. For that we have chosen the arbitrary constants of integration and/or the limits of the integrals using the properties of sums and differences of squares of the functions as given in Table V. For example,

$$\pm\phi_1 = \pm\int_z^K kx_1(u)du = \pm k\int_z^K \text{sn}\,u\,du$$
$$= \ln y_1^{\pm} = (x_3 \mp kx_2)/k',$$

so that

$$\exp(+\phi_1) = (\text{dn}\,u - k\,\text{cn}\,u)/k'$$
$$= k'/(\text{dn}\,u + k\,\text{cn}\,u)$$
$$= \cosh\phi_1 + \sinh\phi_1,$$
$$\exp(-\phi_1) = (\text{dn}\,u + k\,\text{cn}\,u)/k'$$
$$= k'/(\text{dn}\,u - k\,\text{cn}\,u)$$
$$= \cosh\phi_1 - \sinh\phi_1.$$

Then

$$2\cosh\phi_1 = e^{-\phi_1} + e^{-\phi_1} = 2x_3/k',$$
$$\cosh\phi_1 = (1/k')\text{dn}\,u,$$
$$2\sinh\phi_1 = e^{+\phi_1} - e^{-\phi_1} = 2kx_2/k',$$
$$\sinh\phi_1 = (k/k')\text{cn}\,u.$$

From these equations it is easy to obtain all the hyperbolic functions corresponding to ϕ_1 and, using Table I and/or the method to be described in the next section, to find the equations of type (1) that they satisfy, or vice versa.

TABLE V. Properties of elliptic functions useful for the systematic search for generalized exponentials.

Functions	Logarithmic derivatives	$\phi_i = \int x_i\,du$	Squares
$x_1 \equiv \text{sn}\,u$	$z_1 = x_2x_3/x_1$	$\phi_1 = (1/k)\ln[(x_3 - kx_2)/k']$	$x_3^2 - k^2x_2^2 = k'^2$
$x_2 \equiv \text{cn}\,u$	$z_2 = -x_1x_3/x_2$	$i\phi_2 = (1/k)\ln(x_3 + ikx_1)$	$x_3^2 + k^2x_1^2 = 1$
$x_3 \equiv \text{dn}\,u$	$z_3 = -k^2x_1x_2/x_3$	$i\phi_3 = \ln(x_2 + ix_1)$	$x_2^2 + x_1^2 = 1$
$x_4 \equiv \text{cd}\,u$	$z_4 = -k'^2x_5x_6/x_4$	$\phi_4 = (1/k)\ln(x_6 + kx_5)$	$x_6^2 - k^2x_5^2 = 1$
$x_5 \equiv \text{sd}\,u$	$z_5 = +x_4x_6/x_5$	$i\phi_5 = (1/kk')\ln(k'x_6 - ikx_4)$	$k'^2x_6^2 + k^2x_4^2 = 1$
$x_6 \equiv \text{nd}\,u$	$z_6 = +k^2x_4x_5/x_6$	$i\phi_6 = (1/k')\ln(x_4 + ik'x_5)$	$x_4^2 + k'^2x_5^2 = 1$
$x_7 \equiv \text{dc}\,u$	$z_7 = +k'^2x_8x_9/x_7$	$\phi_7 = \ln(x_8 + x_9)$	$x_8^2 - x_9^2 = 1$
$x_8 \equiv \text{nc}\,u$	$z_8 = +x_7x_9/x_8$	$\phi_8 = (1/k')\ln(x_7 + k'x_9)$	$x_7^2 - k'^2x_9^2 = 1$
$x_9 \equiv \text{sc}\,u$	$z_9 = +x_7x_8/x_9$	$\phi_9 = (1/k')\ln[(x_7 + k'x_8)/k]$	$x_7^2 - k'^2x_8^2 = k^2$
$x_{10} \equiv \text{ns}\,u$	$z_{10} = -x_{11}x_{12}/x_{10}$	$\phi_{10} = \ln[(x_{11} + x_{12})/k']$	$x_{11}^2 - x_{12}^2 = k'^2$
$x_{11} \equiv \text{ds}\,u$	$z_{11} = -x_{10}x_{12}/x_{11}$	$\phi_{11} = \ln(x_{10} + x_{12})$	$x_{10}^2 - x_{12}^2 = 1$
$x_{13} \equiv \text{cs}\,u$	$z_{12} = -x_{10}x_{11}/x_{12}$	$\phi_{12} = \ln[(x_{10} + x_{11})/k]$	$x_{10}^2 - x_{11}^2 = k^2$

IV. GENERALIZED CIRCULAR AND HYPERBOLIC FUNCTIONS

From the generalized exponentials $e^{+i\phi}$, $e^{-i\phi}$, we define in the usual way the corresponding circular functions

$$\cos\phi = (e^{+i\phi} + e^{-i\phi})/2, \quad \sin\phi = (e^{+i\phi} - e^{-i\phi})/2i .$$

The other circular functions are defined in the usual way. We shall study below the problem of finding the ODE that is satisfied by $\cos\phi$, knowing the ODE for which $e^{+i\phi}$ is a solution, by finding the ODE that is satisfied by $2y = x + (1/x)$, where x is assumed to be a solution of Eq. (1).

The generalized hyperbolic functions are defined in a similar way,

$$\cosh\phi = (e^{+\phi} + e^{-\phi})/2, \quad \sinh\phi = (e^{+\phi} - e^{-\phi})/2,$$

as we have explained in detail for ϕ_1.

The problem of finding the ODE that is satisfied by $y = \cosh\phi_1$, knowing the ODE that is satisfied by $x = \exp(+\phi)$, is the problem of determining the ODE satisfied by $2y = x + (1/x)$ knowing the ODE satisfied by x.

If x is a solution of Eq. (1), $2y_1 = x - (1/x)$ is a solution of

$$\dot{y}_1^2 + (A - 6B)y_1^2 + 4By_1^4 = A + 2E , \tag{10}$$

and $2iy_2 = x - (1/x)$ is solution of

$$\dot{y}_2^2 + (A + 6B)y_2^2 - 4By_2^4 = A - 2E = A + 2B . \tag{11}$$

From the definition of y_1 and y_2,

$$2\dot{y}_1 = \dot{x} - (\dot{x}/x^2) = (\dot{x}/x)[x - (1/x)] = (\dot{x}/x)2iy_2 ,$$
$$2i\dot{y}_2 = \dot{x} + (\dot{x}/x^2) = (\dot{x}/x)2y_1 .$$

Thus

$$\dot{y}_1 = iy_2(\dot{x}/x), \quad \dot{y}_2 = -iy_1(\dot{x}/x) . \tag{12}$$

Also,

$$\begin{aligned} 4y_1^2 &= -4y_2^2 + 4, \\ y_1 &= (1 - y_2^2)^{1/2}, \quad y_2 = (1 - y_1^2)^{1/2} . \end{aligned} \tag{13}$$

From (12) and (13),

$$\dot{y}_1/(1 - y_2^2)^{1/2} = -\dot{y}_2/(1 - y_2^2)^{1/2} = i\dot{x}/x \equiv -\dot{\phi} .$$

Hence, using Eq. (1),

$$\begin{aligned} -\dot{x}^2/x^2 &= A + 2Bx^2 - (E/x^2) \\ &= (i\dot{x}/x)^2 = \dot{y}_1^2/(1 - y_1^2) = \dot{y}_2^2/(1 - y_2^2). \end{aligned}$$

If $E = -B$,

$$\begin{aligned} \dot{y}_1^2/(1 - y_1^2) &= \dot{y}_2^2/(1 - y_2^2) \\ &= A + B(4y_1^2 - 2) = A - B(4y_2^2 - 2), \end{aligned}$$

and hence Eqs. (10) and (11) are as stated.

All the definitions used in our tables have been made to maintain coherence with these elementary definitions. One has

$$\begin{aligned} &\int (1 - y_1^2)^{-1/2}\, dy_1 \\ &\quad = \arccos y_1 \\ &\quad = -\int (1 - y_2^2)^{-1/2}\, dy_2 = -\arcsin y_2 \\ &\quad = i\int x^{-1}\, dx = i\ln x = -\phi, \quad \ln x = i\phi. \end{aligned}$$

Then, as expected,

$$x = e^{+i\phi} = \cos\phi + i\sin\phi = y_1 + iy_2,$$
$$1/x = e^{-i\phi} = \cos\phi - i\sin\phi = y_1 - iy_2.$$

The most trivial example is to select as angles for the generalized exponentials the integrals (with appropriate factors) of the logarithmic derivatives given in Table III, and thus calculate the corresponding circular and hyperbolic generalized functions. One obtains the functions of Tables I and II, except for trivial factors.

It is less trivial to use as generalized exponentials the functions from Table I with the necessary factors, and to calculate the corresponding circular and hyperbolic functions. It is first necessary to find a pair of reciprocal functions that are solutions of the same Eq. (1). Some of these pairs are

$$k^{1/2}\,\mathrm{sn}\,\omega t, \quad 1/(k^{1/2}\,\mathrm{sn}\,\omega t) = (\mathrm{ns}\,\omega t)/k^{1/2};$$
$$A = (1 + k^2)\omega^2, \quad B = -E = -k\omega^2;$$
$$(\mathrm{dn}\,\omega t)/k'^{(1/2)}, \quad k'^{(1/2)}\,\mathrm{nd}\,\omega t;$$
$$A = (1 + k'^2)\omega^2, \quad B = -E = +k'\omega^2.$$

In consequence, as noted above, the Jacobian elliptic functions can be used not only as generalized circular functions, but also as generalized exponentials. Nevertheless, probably the most useful generalized exponentials are not the functions of Tables I, II (the x's), or III (the z's) with appropriate factors, but the functions of Table IV with the factors given there.

V. GENERALIZED EXPONENTIAL SOLUTIONS FOR EQUATIONS OF TYPE (2)

The following theorem allows these solutions to be found. The function x is a solution of Eq. (1) if $y = x^2$ is a solution of the ODE

$$\dot{y}^2 + A_3 y^3 + A_2 y^2 + A_1 y = 0,$$

where $A_1 = -4E(t)$, $A_2 = +4A$, $A_3 = 4B$, i.e., the general cubic without an independent term. A simple translation yields the general cubic with an independent term. Observe that A_1 is more general than needed for the case with elliptic functions, where $E = \text{const}$.

From $y = x^2$ one has $\dot{y}^2 = 4x^2\dot{x}^2$ and, using Eq. (1),

$$\dot{y}^2 + 4Ay^2 + 4By^3 = 4yE(t),$$

which proves the theorem.

Using this theorem, it is a trivial matter to find from the generalized exponentials, etc. [solutions of Eq. (1)], the corresponding generalized exponential solution of Eq. (2).

VI. RELATION WITH SPECIAL RELATIVITY AND SPACE-TIME CURVATURE

From Eq. (1),

$$Bx^4 + Ax^2 + (\dot{x}^2 - E) = 0,$$

and hence

$$2Bx^2 + A = \pm (A^2 + 4BE - 4B\dot{x}^2)^{1/2}.$$

Using $\ddot{x} = -x(A + 2Bx^2)$, one has $\ddot{x}(A^2 + 4BE - 4B\dot{x}^2)^{-1/2} = -x$.

Taking $4B/(A^2 + 4BE) = \pm (1/c^2)$ and $A^2 + 4BE = \omega^4$ we have

$$\ddot{x}[1 \pm (\dot{x}/c)^2]^{-1/2} = -\omega^2 x. \tag{14}$$

This equation is formally equivalent to the equation we have called the "transverse relativistic oscillator."[1] The solutions of Eq. (14) are readily obtained using our tables.

For the curvature of a space-time trajectory $x = x(ict)$,

$$-\kappa/c^2 = x''[1 - (x'/c)^2]^{-3/2},$$

where

$$x' = (1/ic)\frac{dx}{dt}.$$

One sees that the equation

$$\ddot{x}[1 - (x/c)^2]^{-3/2} = -\omega^2 x \tag{15}$$

that corresponds to the usual "relativistic oscillator"[1] implies that the motion of this oscillator is a space-time trajectory such that the curvature is proportional to the position. For the integration of Eq. (15), define, as in the special theory of relativity,

$$d\tau = [1 - (\dot{x}/c)^2]^{1/2}\,dt, \quad dt = \left[1 + \frac{1}{c^2}\left(\frac{dx}{d\tau}\right)^2\right]^{1/2} d\tau.$$

Then

$$\frac{d}{d\tau}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}\left[1 - \frac{1}{c^2}\left(\frac{dx}{dt}\right)^2\right]^{-1/2} = \frac{d^2x}{d\tau^2}\left[1 + \frac{1}{c^2}\left(\frac{dx}{d\tau}\right)^2\right]^{-3/2},$$

$$\frac{d}{dt}\left(\frac{dx}{d\tau}\right) = \frac{d^2x}{d\tau^2}\left[1 + \frac{1}{c^2}\left(\frac{dx}{d\tau}\right)^2\right]^{1/2} = \frac{d^2x}{dt^2}\left[1 - \frac{1}{c^2}\left(\frac{dx}{dt}\right)^2\right]^{-3/2}.$$

In other words, after simple changes of variables, equations of type (15) are transformed into equations of type (14).

VII. CONCLUSIONS

We have developed a systematic method for finding the generalized exponential, circular, and hyperbolic functions that are solutions of the most simple extensions of the simple harmonic oscillator (SHO), i.e., for the ASO with potential $V(x) = Ax^2 + Bx^4$ and the AAO with $V(x) = Ax^2 + Bx^3$. This opens the way for the development of the corresponding Fourier series, Fourier and Laplace transforms, etc., that is, for a systematic linearization of the simplest possible nonlinear problems. Following the ideas of Petiau[3] and the relations we give with the special relativity SHO and the space-time curvature, our formulas can be used for nonlinear PDE's and for the construction of simple models of nonlinear quantum mechanics.

ACKNOWLEDGMENTS

Thanks are due to the mathematicians Professor Benítez and Professor M. Fernández for having convinced us that a more formal presentation of the subject adds nothing substantial. Discussions with R. Chatwin have contributed to improvements of this and previous papers[4–7] on nonlinear mechanics. Thanks are also due to the anonymous referee for his helpful comments.

The authors acknowledge the Comisión Asesora de Investigación Científica y Técnica, Spain (Project No. 1179-84) for financial support.

[1]J. Díaz Bejarano and A. Martín Sánchez, An. Fis. Ser. A **79**, 8 (1983).

[2]J. Díaz Bejarano, A. Martín Sánchez, and C. Miró Rodríguez, J. Chem. Phys. **85**, 5128 (1986).

[3]G. Petiau, Suppl. Nuovo Cimento X, 542 (1958).

[4]S. Bravo Yuste and J. Díaz Bejarano, J. Sound Vib. **110**, 347 (1986).

[5]Simple formulas for the quantum ASO or AAO similar to the well-known formulas of the simple harmonic oscillator (SHO) are not known. Some references to the enormous literature on approximations for the type (a) ($A > 0$, $B > 0$) ASO energy levels are given in A. Martín Sánchez and J. Díaz Bejarano, J. Phys. A: Math. Gen. **19**, 887 (1986). References for the other ASO types are less numerous and given in the same paper. The scarce references for the AAO energy levels may be found in Ref. 2.

[6]J. A. Caballero Carretero and A. Martín Sánchez, J. Math. Phys. **28**, 636 (1987).

[7]There are no known simple exact formulas for the damped, forced, etc., classical ASO or AAO similar to those for the SHO. To our knowledge no reference exists for exponentials, Fourier series, Fourier transforms, etc., for these oscillators (the only exception being Ref. 4). Some references on the very abundant literature on approximations for the damped ASO, type (a) are given in S. Bravo Yuste and J. Díaz Bejarano, J. Sound Vib. **114**, 33 (1987).

[8]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).

[9]E. H. Neville, *Jacobian Elliptic Functions* (Clarendon, Oxford, 1951).

[10]F. Tölke, *Praktische Funktionenlehre* (Springer, Heidelberg, 1968), Vols. II–V.

The Heisenberg–Weyl group in the coherent state basis and the Bargmann transform

Debabrata Basu and T. K. Kar
Department of Physics, Indian Institute of Technology, Kharagpur 721302, West Bengal, India

(Received 4 June 1987; accepted for publication 24 February 1988)

It is shown that the representation matrices of the quantum mechanical group in the coherent state basis can be used as bases for the expansion of arbitrary square integrable functions on the group. Although this basis is nonorthogonal the expansion can be inverted in a manner similar to an orthogonal basis. Next a coherent state "wave function" is defined in analogy with the Schrödinger wave function and the above matrix elements are used to determine the action of the group in the space of the coherent state wave functions. The function space in this realization is Bargmann's Hilbert space of analytic functions. It is shown that the mixed basis matrix element between the coordinate and coherent state bases is essentially the integral kernel of Bargmann. The transpose of this matrix element yields the kernel for the inversion of the transform. In this construction the Bargmann transform appears as the unitary transformation connecting the Schrödinger and coherent state wave functions. This considerably simplifies Bargmann's original arguments and yields the integral transform as well as its inversion formula in a simple way.

I. INTRODUCTION

The group theoretic content of Heisenberg's commutation relation between the position and momentum operators was discovered by Weyl[1] about 60 years ago. The quantum mechanical or the Heisenberg–Weyl (HW) group is a noncompact non-Abelian group with a nontrivial center and in one dimension consists of three generators. It is the simplest nilpotent group and plays the same fundamental role in the general theory of nilpotent groups[2] as SU(2) does in the representation theory of semisimple groups. However, detailed descriptions of the group are not abundant in mathematical physics literature. We may cite the works of Barut and Raczka,[3] Wolf and co-workers,[4] Hermann,[5] and Talman[6] who have studied some aspects of the representations of this group.

In this paper we investigate some consequences of the reduction of the unitary irreducible representations (UIR's) of the HW group in the coherent state basis. The coherent states were originally introduced by Glauber,[7] Klauder,[8] and Sudarshan[9] as the appropriate quantum mechanical states for the description of intense beam of photons. In the present paper we intend to use the Glauber coherent states as a possible basis for the HW group. A similar coherent state basis for the SU(1,1) group has been developed by Barut and Girardello.[10] We hope to apply the method of our paper in future for a parallel investigation of the Barut–Girardello coherent states.

In this paper we consider two related problems: (i) expansion of an arbitrary function on the group in terms of the matrix elements of the group in the coherent state basis and (ii) a group theoretic derivation of Bargmann's integral transform.[11] Although Glauber pointed out the connection of the coherent states with Bargmann's Hilbert space of analytic functions,[11] he noted only the coincidence of the scalar product of two arbitrary oscillator states with that in the Bargmann space. The connection becomes clearer by taking coherent states as a basis for the UIR's of the HW group.

We start from the realization of the UIR's of the group in the space of $L^2(R)$ functions as given by Wolf[4] and Barut and Raczka[3] and calculate the matrix elements in this basis. We show that these matrix elements can be used as a basis for the expansion of arbitrary square integrable functions on the group. Although the matrix elements constitute a nonorthogonal basis the expansion can be inverted and the coefficients calculated in the traditional way through an "integral equation." The process of expansion and its inversion that yields the coefficients of the expansion constitutes an integral transform and its inversion formula. The integral equation in conjunction with the condition of finiteness of the norm yields a theorem (Sec. III) that states that the expansion coefficient is expressible in terms of an entire analytic function of the appropriate parameters of the expansion. Following a previous paper[12] we next introduce a coherent state wave function analogous to the Schrödinger wave function and determine the action of the group operator in the space of this wave function. The above theorem ensures that the representation space consists of entire analytic functions of an appropriate complex variable and coincides with Bargmann's Hilbert space. We next show that the mixed basis matrix element of the group between the coordinate and coherent state bases is essentially the integral kernel of Bargmann. The transpose of this matrix element yields the kernel for the inversion of the transform. In this construction Bargmann's integral transform appears as the unitary transformation connecting the Schrödinger and coherent state wave functions. The role of the Bargmann transform connecting these two wave functions is similar to that of the Fourier transform connecting the coordinate and momentum space wave functions. This observation considerably simplifies Bargmann's involved analytical arguments and yields in a simple and unitary way Bargmann's integral transform as well as its inversion formula.

II. FUNDAMENTAL EQUATIONS AND THE EXPANSION PROBLEM

In one dimension the HW group has three infinitesimal operators Q, P, and H, which satisfy the commutation relations

$$[Q,P] = iH,\ [Q,H] = [P,H] = 0. \tag{2.1}$$

Following Wolf[4] we first introduce a three-dimensional non-Hermitian representation

$$Q=\begin{pmatrix} 0 & -i & -i \\ i & 0 & 0 \\ -i & 0 & 0 \end{pmatrix},\quad P=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix},$$
$$H=\begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & -2 & -2 \end{pmatrix}. \tag{2.2}$$

A finite group element can be obtained by exponentiating the generators (2.2) and is given by

$$\begin{aligned} g(x,y,z) &= e^{i(xQ+yP+zH)} \\ &= e^{ixQ}e^{iyP}e^{i(z+xy/2)H} \\ &= \begin{pmatrix} 1 & x+iy & x+iy \\ -x+iy & 1+2iz-(x^2+y^2)/2 & 2iz-(x^2+y^2)/2 \\ x-iy & -2iz+(x^2+y^2)/2 & 1-2iz+(x^2+y^2)/2 \end{pmatrix}, \end{aligned} \tag{2.3}$$

where x,y,z are the three group parameters. The group multiplication law follows by multiplying two matrices of the form (2.3) and is given by

$$g(x_1,y_1,z_1)g(x_2,y_2,z_2) = g(x_1+x_2,\ y_1+y_2,\ z_1+z_2+\tfrac{1}{2}[y_1x_2-y_2x_1]). \tag{2.4}$$

The group identity is $e=g(0,0,0)$ and the inverse $g^{-1}(x,y,z) = g(-x,-y,-z)$. The right and left invariant Haar measure is simply the three-dimensional volume element

$$dg = dx\,dy\,dz. \tag{2.5}$$

We now introduce a square integrable function $f(g)$ on the group transforming as

$$T_{g'}f(g) = f(g'^{-1}g), \tag{2.6}$$

where

$$g' = g(x',y',z').$$

This evidently yields a representation in the space of these functions.

We next introduce the decomposition

$$g(x,y,z) = g(0,y,z-xy/2)\ g(x,0,0) \tag{2.7}$$

and consider functions

$$f_c = f_c(y,u) = f_c(y,z-xy/2) \tag{2.8}$$

on the space of left cosets. Since in the space of these functions

$$H = i\left(\frac{\partial}{\partial u}\right)$$

the decomposition of the representation space into the eigenspaces of irreducible representations is achieved by the Fourier transformation

$$f_c(y,u) = \int d\lambda\ \phi_\lambda(y)e^{-i\lambda u}\,d\lambda. \tag{2.9}$$

This immediately yields the multiplier representation of Wolf[4] and Barut and Raczka[3]

$$T_{g'}\phi_\lambda(y) = e^{i\lambda(z'+x'y/2-x'y)}\phi_\lambda(y-y'), \tag{2.10}$$

where $\phi_\lambda(y)$ is a square integrable function of the parameter y. The representation (2.10) is unitary with respect to the scalar product

$$(\phi_\lambda,\Psi_\lambda) = \int \bar{\phi}_\lambda(y)\Psi_\lambda(y)dy. \tag{2.11}$$

The Hermitian generators of the group are now given by

$$Q = -\lambda y,\quad P = i\left(\frac{\partial}{\partial y}\right),\quad H=\lambda. \tag{2.12}$$

The orthonormal coordinate and momentum bases are given by

$$\begin{aligned} f_q &= |\lambda|^{1/2}\delta(q+\lambda y), \\ \chi_p &= (2\pi)^{-1/2}\exp(-ipy). \end{aligned} \tag{2.13}$$

The coherent state basis is the eigenfunction of the annihilation operator

$$a = (2|\lambda|)^{-1/2}(Q + i\,\mathrm{sgn}\,\lambda P) \tag{2.14}$$

and is given by

$$\begin{aligned} \Psi_\alpha &= (|\lambda|/\pi)^{1/4}\exp(-\alpha_1^2 - i\alpha_1\alpha_2) \\ &\quad\times\exp(-(|\lambda|/2)[y^2 + 2\sqrt{2/|\lambda|}\ \mathrm{sgn}\,\lambda\alpha y]), \end{aligned} \tag{2.15}$$

where the complex number α stands for the eigenvalue of a and α_1 and α_2 are its real and imaginary parts. The choice of phase ensures that the overlap between two eigenstates is given by Glauber's formula,

$$(\Psi_\beta,\Psi_\alpha) = \exp(\bar{\beta}\alpha - |\alpha|^2/2 - |\beta|^2/2). \tag{2.16}$$

The "completeness" condition is given by

$$\frac{1}{\pi}\int \bar{\Psi}_\alpha(y')\Psi_\alpha(y)d^2\alpha = \delta(y'-y), \tag{2.17}$$

where $d^2\alpha = d\alpha_1\,d\alpha_2$.

It is now easy to obtain the matrix element of the group in any basis,

$$D^\lambda_{mn}(g') = (u^\lambda_m, T_{g'}u^\lambda_n). \tag{2.18}$$

It should be pointed out, however, that the matrix element in the coherent state basis can be obtained quite easily without any integration using only the commutation relation between a and its Hermitian conjugate $a^\dagger$. We list below the matrix elements in the various orthogonal and nonorthogonal bases:

$$D^{\lambda}_{\beta\alpha}(g) = \exp\Big(\bar{\beta}\alpha - \frac{|\alpha|^2}{2} - \frac{|\beta|^2}{2} + i\lambda z - \frac{|\lambda|}{2}\Big[\frac{x^2+y^2}{2} + \sqrt{2/|\lambda|}\,\{\bar{\beta}(\operatorname{sgn}\lambda y - ix) - \alpha(\operatorname{sgn}\lambda y + ix)\}\Big]\Big), \tag{2.19a}$$

$$D^{\lambda}_{qq'}(g) = \delta(\lambda y + q - q')\exp i(\lambda z + \tfrac{1}{2}x[q+q']), \tag{2.19b}$$

$$D^{\lambda}_{\beta q}(g) = (|\lambda|\pi)^{-1/4}\exp\Big(-\frac{|\beta|^2}{2} + i\lambda\Big[z - \frac{xy}{2}\Big] - \frac{|\lambda|y^2}{2} - \frac{1}{2}\Big[\bar{\beta}^2 + \frac{q^2}{|\lambda|}\Big] + \sqrt{2/|\lambda|}\,\bar{\beta}\,[q - \lambda y] + q[ix + \operatorname{sgn}\lambda y]\Big). \tag{2.19c}$$

We now consider the problem of expansion of an arbitrary square integrable function $f(g)$ on the group in terms of the matrix elements (2.19a):

$$f(g) = \int d\lambda \int\int d^2\beta\, d^2\alpha\, a(\lambda;\beta,\alpha)\, D^{\lambda}_{\beta\alpha}(g). \tag{2.20}$$

Such an expansion is unique only if

$$\int\int d^2\beta\, d^2\alpha\, a(\lambda;\beta,\alpha) D^{\lambda}_{\beta\alpha}(g) = \int\int d^2\gamma\, d^2\delta\, D^{\lambda}_{\gamma\delta}(g)\,\frac{1}{\pi^2}\int\int d^2\beta\, d^2\alpha (\Psi_\beta,\Psi_\gamma) \times(\Psi_\delta,\Psi_\alpha) a(\lambda;\beta,\alpha). \tag{2.21}$$

This therefore requires

$$a(\lambda;\gamma,\delta) = \frac{1}{\pi^2}\int\int d^2\beta\, d^2\alpha (\Psi_\beta,\Psi_\gamma)\ (\Psi_\delta,\Psi_\alpha) a(\lambda;\beta,\alpha). \tag{2.22}$$

Using Eq. (2.16) we finally obtain

$$a(\lambda;\gamma,\delta) = \exp\Big[\frac{-(|\delta|^2+|\gamma|^2)}{2}\Big]\frac{1}{\pi^2}\int\int d^2\beta\, d^2\alpha \times\exp\Big[\bar{\beta}\gamma + \bar{\delta}\alpha - \frac{|\alpha|^2+|\beta|^2}{2}\Big] a(\lambda;\beta,\alpha). \tag{2.23}$$

We shall now show that the expansion (2.20) can be inverted and the coefficients calculated in the traditional way using Eq. (2.23). Multiplying both sides of the expansion (2.20) by $\bar{D}^{\lambda'}_{\beta'\alpha'}(g)$ and integrating over g we obtain

$$\int \bar{D}^{\lambda'}_{\beta'\alpha'}(g) f(g)\, dg = \int d\lambda \int d^2\beta\, d^2\alpha\, a(\lambda;\beta,\alpha) \times\int \bar{D}^{\lambda'}_{\beta'\alpha'}(g) D^{\lambda}_{\beta\alpha}(g)\, dg. \tag{2.24}$$

We now note that

$$\int \bar{D}^{\lambda'}_{\beta'\alpha'}(g) D^{\lambda}_{\beta\alpha}(g)\, dg = \frac{4\pi^2}{|\lambda'|}\delta(\lambda - \lambda')\exp\Big[\frac{-(|\alpha|^2+|\beta|^2)}{2} - \frac{|\alpha'|^2+|\beta'|^2}{2} + \bar{\beta}\beta' + \bar{\alpha}'\alpha\Big]. \tag{2.25}$$

This yields

$$\frac{4\pi^2}{|\lambda'|}\exp\Big[\frac{-(|\alpha'|^2+|\beta'|^2)}{2}\Big]\int d^2\alpha \int d^2\beta \times\exp\Big(\beta'\bar{\beta} + \bar{\alpha}'\alpha - \frac{|\alpha|^2}{2} - \frac{|\beta|^2}{2}\Big) a(\lambda;\beta,\alpha) = \int \bar{D}^{\lambda'}_{\beta'\alpha'}(g) f(g)\, dg. \tag{2.26}$$

By Eq. (2.23) the lhs turns out to be $a(\lambda';\beta',\alpha')$ multiplied by $4\pi^4/|\lambda'|$ and we have

$$a(\lambda';\beta',\alpha') = \frac{|\lambda'|}{4\pi^4}\int \bar{D}^{\lambda'}_{\beta'\alpha'}(g) f(g)\, dg. \tag{2.27}$$

In the next section we show that the expansion coefficient $a(\lambda;\beta,\alpha)$ is of the form

$$a(\lambda;\beta,\alpha) = \exp[-(|\alpha|^2+|\beta|^2)/2] E(\lambda;\beta,\bar{\alpha}),$$

where $E(\lambda;\beta,\bar{\alpha})$ is an entire function analytic in β and $\bar{\alpha}$. The completeness relation for the matrix elements now reads

$$\int d\lambda \frac{|\lambda|}{4\pi^2}\int \bar{D}^{\lambda}_{\beta\alpha}(g') D^{\lambda}_{\beta\alpha}(g)\, d^2\beta\, d^2\alpha = \pi^2\delta(x'-x)\delta(y'-y)\delta(z'-z), \tag{2.28}$$

which can be verified by explicit calculation. This may be compared with the completeness condition for the matrix elements in the orthonormal p or q basis:

$$\int d\lambda \frac{|\lambda|}{4\pi^2}\int \bar{D}^{\lambda}_{rr'}(g') D^{\lambda}_{rr'}(g)\, dr\, dr' = \delta(x'-x)\delta(y'-y)\delta(z'-z). \tag{2.29}$$

Although the additional factor of π^2 in Eq. (2.28) suggests overcompleteness we see that the expansion (2.20) can be inverted in a manner similar to an orthonormal basis. Equations (2.20) and (2.27), therefore, constitute an integral transform and its inverse.

III. THE COHERENT STATE WAVE FUNCTION AND THE BARGMANN TRANSFORM

It should be noted that the realization (2.12) for the generators of the group on the y manifold is not the usual one in quantum mechanics. In the Schrödinger realization the eigenvalue q of the operator Q is regarded as the configuration space coordinate. The Schrödinger wave function can be built as

$$\Psi(q) = (f_q,\phi). \tag{3.1}$$

The action of the group operator in the space of the wave functions $\Psi(q)$ is then defined by

$$\hat{T}_g\Psi(q) = (f_q, T_g\phi) \tag{3.2a}$$

$$= \int D_{qq'}(x,y,z)\Psi(q')\, dq'. \tag{3.2b}$$

Using Eq. (2.19b) we obtain

$$\hat{T}_g\Psi(q) = \exp i(\lambda z + qx + \lambda xy/2)\Psi(q+\lambda y), \quad (3.3)$$

which yields

$$Q = q, \quad P = -i\lambda \frac{d}{dq}, \quad H = \lambda, \quad (3.4)$$

the familiar quantum mechanical operators.

We now define analogously the coherent state wave function

$$\phi(\beta) = (\Psi_\beta, \phi). \quad (3.5)$$

Then the coherent state realization of the group operator is given by

$$\hat{T}_g\phi(\beta) = (\Psi_\beta, T_g\phi) \quad (3.6)$$

$$= \frac{1}{\pi}\int d^2\alpha\, D^{\lambda}_{\beta\alpha}(x,y,z)\phi(\alpha). \quad (3.7)$$

Setting $x = y = z = 0$ we obtain

$$\phi(\beta) = \frac{1}{\pi}\exp\Big(-\frac{|\beta|^2}{2}\Big)\int d^2\alpha \exp\Big(\bar{\beta}\alpha - \frac{|\alpha|^2}{2}\Big)\phi(\alpha). \quad (3.8)$$

We now define

$$\phi(\alpha) = \exp(-|\alpha|^2/2)f(\bar{\alpha}), \quad (3.9)$$

where a possible α dependence of the function f has been suppressed. The function $f(\bar{\alpha})$ satisfies

$$f(\bar{\beta}) = \int \exp(\bar{\beta}\alpha)f(\bar{\alpha})d\mu(\alpha), \quad (3.10)$$

where

$$d\mu(\alpha) = (1/\pi)\exp(-|\alpha|^2)d^2\alpha.$$

The scalar product in the Hilbert space of the functions f is given by

$$(f_1,f_2) = \int \overline{f_1(\bar{\beta})}f_2(\bar{\beta})d\mu(\beta). \quad (3.11)$$

The above form of the scalar product is obtained by noting that

$$(\phi_1,\phi_2) = \frac{1}{\pi}\int (\phi_1,\Psi_\beta)\,(\Psi_\beta,\phi_2)d^2\beta$$

$$= \frac{1}{\pi}\int \overline{\phi_1(\beta)}\phi_2(\beta)d^2\beta$$

and using the definition (3.9).

We now prove the following theorem.

Theorem: The function $f(\bar{\beta})$ satisfying Eq. (3.10) and having a finite norm according to the scalar product (3.11) is an entire function analytic in $\bar{\beta}$.

Proof: To prove the analyticity we recall that a complex valued function $f(z,\bar{z})$ is analytic in z if

$$\frac{\partial f}{\partial \bar{z}} = 0. \quad (3.12)$$

Equation (3.12) is equivalent to the Cauchy–Reimann equations and defines an analytic function of z.

Differentiating both sides of Eq. (3.10) with respect to β we obtain

$$\frac{\partial f}{\partial \beta} = \int d\mu(\alpha)f(\bar{\alpha})\frac{\partial}{\partial \beta}[\exp(\bar{\beta}\alpha)] = 0,$$

because the kernel $\exp(\bar{\beta}\alpha)$ is independent of β.

Thus the function f satisfying Eq. (3.10) is analytic in $\bar{\beta}$, i.e.,

$$f = f(\bar{\beta}). \quad (3.13)$$

We next expand $\exp(\bar{\beta}\alpha)$ in Eq. (3.10) in a power series and define

$$C_n = \int \alpha^n f(\bar{\alpha})d\mu(\alpha), \quad n = 0,1,2,\ldots . \quad (3.14)$$

Thus

$$f(\bar{\beta}) = \sum_{n=0}^{\infty}\frac{C_n\bar{\beta}^n}{n!}. \quad (3.15)$$

We shall show that the radius of convergence of the power series (3.15) is infinite by requiring the squared norm (f,f) to be finite. This yields

$$\sum_{n=0}^{\infty}\frac{|C_n|^2}{n!} = \text{finite}. \quad (3.16)$$

The series (3.16) is therefore absolutely convergent and we have

$$\lim_{n\to\infty}|C_{n+1}/C_n|^2(1/n) \leqslant 1.$$

Thus, for arbitrarily large n,

$$|C_{n+1}/C_n| \leqslant \sqrt{n}.$$

If we now denote the nth term of the power series (3.15) by u_n we have

$$\lim_{n\to\infty}\left|\frac{u_{n+1}}{u_n}\right| = \lim_{n\to\infty}\left|\frac{C_{n+1}}{C_n}\right|\frac{|\bar{\beta}|}{n} \leqslant \frac{|\bar{\beta}|}{\sqrt{n}} \to 0$$

no matter however large $|\bar{\beta}|$ is.

Hence the radius of convergence of the power series (3.15) is infinite and the analytic function represented by it is an entire function.

A straightforward extension of this theorem for two complex variables states that the coefficient $a(\lambda;\beta,\alpha)$ of the expansion (2.20) satisfying Eq. (2.23) is of the form

$$a(\lambda;\beta,a) = \exp(-\tfrac{1}{2}[|\alpha|^2 + |\beta|^2])E(\lambda;\beta,\bar{\alpha}), \quad (3.17)$$

where $E(\lambda;\beta,\bar{\alpha})$ is an entire function analytic in β and $\bar{\alpha}$. This extension can be done in a manner similar to the one-variable case by requiring the square integrability of the function $f(g)$, i.e.,

$$\int |f(g)|^2\,dg = \text{finite},$$

which yields

$$\int \overline{E(\lambda;\beta,\bar{\alpha})}E(\lambda;\beta,\bar{\alpha})d\mu(\beta)d\mu(\alpha) = \text{finite}. \quad (3.18)$$

The two-variable extension of the above theorem follows easily by requiring the finiteness of the norm (3.18).

We now proceed to realize the representation of the HW group in the space of the functions $f(\bar{\beta})$. Substituting the explicit form of the matrix element $D^{\lambda}_{\beta\alpha}$ as given in Eq. (2.19a) and noting that

$$\frac{1}{\pi}\int \exp\left[\left\{\bar\beta+\left(\frac{|\lambda|}{2}\right)^{1/2}(\operatorname{sgn}\lambda y+ix)\right\}\alpha-|\alpha|^2\right]f(\bar\alpha)d^2\alpha$$
$$=f(\bar\beta+\sqrt{|\lambda|/2}(\operatorname{sgn}\lambda y+ix)) \qquad (3.19)$$

we obtain

$$\hat T_g\phi(\beta)=\exp\left(-\frac{|\beta|^2}{2}+i\lambda z-\frac{|\lambda|}{2}\left[\frac{x^2+y^2}{2}+\sqrt{2/|\lambda|}\,\bar\beta(\operatorname{sgn}\lambda y-ix)\right]\right)\times f\left(\bar\beta+\left(\frac{|\lambda|}{2}\right)^{1/2}(\operatorname{sgn}\lambda y+ix)\right). \qquad (3.20)$$

If we now define

$$\hat V_g=\exp(|\beta|^2/2)\hat T_g\exp(-|\beta|^2/2), \qquad (3.21)$$

we obtain the action of the group operator in the Hilbert space of entire analytic functions:

$$\hat V_gf(\bar\beta)=\exp\left(i\lambda z-\frac{|\lambda|}{2}\left[\frac{x^2+y^2}{2}+\left(\frac{2}{|\lambda|}\right)^{1/2}\bar\beta(\operatorname{sgn}\lambda y-ix)\right]\right)\times f\left(\bar\beta+\left(\frac{|\lambda|}{2}\right)^{1/2}(\operatorname{sgn}\lambda y+ix)\right). \qquad (3.22)$$

The generators of the group in this realization are given by

$$Q=\left(\frac{|\lambda|}{2}\right)^{1/2}\left(\bar\beta+\frac{\partial}{\partial\bar\beta}\right),\quad P=\operatorname{sgn}\lambda i\left(\frac{|\lambda|}{2}\right)^{1/2}\left(\bar\beta-\frac{\partial}{\partial\bar\beta}\right), \qquad (3.23)$$

which are the Fock–Bargmann operators. The representation (3.22) is unitary with respect to the scalar product (3.11).

We are now in a position to derive Bargmann's integral transform and its inversion. We start from Eq. (3.6) and use the completeness of the orthonormal coordinate basis f_q. Thus

$$\hat T_g\phi(\beta)=\int(\Psi_\beta,T_gf_q)\,(f_q,\phi)dq=\int D^\lambda_{\beta q}(x,y,z)\Psi(q)dq, \qquad (3.24)$$

where we have used the definition (3.1) of the Schrödinger wave function.

We now substitute the explicit form of the mixed basis matrix element as given by Eq. (2.19c). Transferring $\exp(-|\beta|^2/2)$ to the left and using Eq. (3.22) we obtain

$$\exp\left(i\lambda z-\frac{|\lambda|}{2}\left[\frac{x^2+y^2}{2}+\sqrt{2/|\lambda|}\,\bar\beta(\operatorname{sgn}\lambda y-ix)\right]\right)\times f(\bar\beta+\sqrt{|\lambda|/2}\,(\operatorname{sgn}\lambda y+ix))$$
$$=(|\lambda|\pi)^{-1/4}\exp\left[i\lambda\left(z-\frac{xy}{2}\right)-\frac{|\lambda|y^2}{2}\right]\times\int\exp\left[-\frac{\bar\beta^2+q^2/|\lambda|}{2}+\sqrt{2/|\lambda|}\bar\beta(q-\lambda y)+q(ix+\operatorname{sgn}\lambda y)\right]\Psi(q)dq. \qquad (3.25)$$

We now introduce a new complex variable

$$\xi=\bar\beta+\sqrt{|\lambda|/2}(\operatorname{sgn}\lambda y+ix)$$

and note that all the x-, y-, and z-dependent factors cancel from the two sides. We then obtain after simplification

$$f(\xi)=(|\lambda|\pi)^{-1/4}\int\exp\left[-\frac{\xi^2+q^2/|\lambda|}{2}+\left(\frac{2}{|\lambda|}\right)^{1/2}\xi q\right]\Psi(q)dq, \qquad (3.26)$$

which is the integral transform of Bargmann. To obtain the inversion formula for this transform we start from Eq. (3.2a) and use the completeness condition (2.17) for the coherent states. Thus

$$\hat T_g\Psi(q)=\frac{1}{\pi}\int(f_q,T_g\Psi_\beta)\,(\Psi_\beta,\phi)d^2\beta=\frac{1}{\pi}\int D^\lambda_{q\beta}(x,y,z)\phi(\beta)d^2\beta=\frac{1}{\pi}\int\bar D^\lambda_{\beta q}(-x,-y,-z)\phi(\beta)d^2\beta.$$

Using Eqs. (2.19c), (3.3), and (3.9), we obtain

$$\exp[i\lambda(z+xy/2)+iqx]\Psi(q+\lambda y)$$
$$=\exp\left[i\lambda\left(z+\frac{xy}{2}\right)-\frac{|\lambda|y^2}{2}\right]\int\exp\left[-\frac{\beta^2+q^2/|\lambda|}{2}+\sqrt{2/|\lambda|}\beta(q+\lambda y)+q(ix-\operatorname{sgn}\lambda y)\right]f(\bar\beta)d\mu(\beta).$$

Replacing q by $(q-\lambda y)$ and $\bar\beta$ by ξ we obtain, as before,

$$\Psi(q)=(|\lambda|\pi)^{-1/4}\int\exp\left[-\frac{\bar\xi^2+q^2/|\lambda|}{2}+\sqrt{2/|\lambda|}\,\bar\xi q\right]f(\xi)d\mu(\xi), \qquad (3.27)$$

which is Bargmann's inversion formula.

ACKNOWLEDGMENTS

The authors wish to thank the referee for suggesting a crucial modification in the proof of a theorem in Sec. III. The authors would also like to thank Professor G. P. Sastry for many helpful discussions.

[1]H. Weyl, Z. Phys. **46**, 1 (1928).
[2]A. A. Kirillov, Usp. Math. Nauk **106**, 57 (1962); Russ. Math. Surv. **17**, 53 (1962).
[3]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN-Polish Scientific, Warsaw, 1980), pp. 323 and 588.
[4]K. B. Wolf, in *Group Theory and Its Applications*, edited by E. M. Loebl (Academic, New York, 1975), Vol. 3; M. Garcia-Bulle, W. Lassner, and K. B. Wolf, J. Math. Phys. **27**, 29 (1986).
[5]R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).
[6]J. D. Talman, *Special Functions, a Group Theoretic Approach* (Benjamin, New York, 1968).

[7]R. J. Glauber, Phys. Rev. Lett. **10**, 84 (1963); Phys. Rev. **131**, 2766 (1963).
[8]J. R. Klauder, Ann. Phys. (NY) **11**, 123 (1960); J. R. Klauder and E. C. G. Sudarshan, *Fundamentals of Quantum Optics* (Benjamin, New York, 1968).
[9]E. C. G. Sudarshan, Phys. Rev. Lett. **10**, 277 (1963).
[10]A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
[11]V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).
[12]D. Basu and T. K. Kar, "The quantum mechanical group, coherent states and the Bargmann transform," submitted to the Workshop on Relativistic Invariance and Quantum Phenomena (Maryland).

A theorem on the rate of regeneration in decay processes

Michio Nishioka
Department of Physics, Faculty of Liberal Arts, Yamaguchi University, Yamaguchi 753, Japan

(Received 21 December 1987; accepted for publication 13 April 1988)

There is given a generalization of the theorem of Misra and Sinha [Helv. Phys. Acta **50**, 99 (1977)], which shows that in the decay process the regeneration of the undecayed unstable states from the decay products cannot proceed too slowly at arbitrarily short time and, correspondingly, the deviation from exponential decay at arbitrarily short time cannot be too small either. The generalization presented here is shown to be valid in the region where the theorem of Misra and Sinha cannot be applied.

I. INTRODUCTION

In the studies[1–5] of the decay processes of unstable particles, absence of regeneration of undecayed unstable states from the decay products is mathematically expressed by the semigroup property of the reduced evolution operator $Z(t) = EU_tE$:

$$Z(t+s) - Z(t)Z(s) = EU_tE^{\perp}U_sE = 0, \qquad (1.1)$$

for $t,s \geqslant 0$, where E denotes the projection into the subspace spanned by the undecayed (unstable) states of the system and $E^{\perp} = I - E$ the projection onto the subspace of decayed states, and where U_t is a one-parameter unitary group describing time evolution of the total system with the Hamiltonian H. Several studies[1–4] on this assumption have shown that it is in conflict with the physical requirement that the spectrum of the Hamiltonian H be bounded below. So Misra and Sinha[5] studied the consequence of having a nonvanishing but slow rate of regeneration and found that too slow a rate of regeneration at arbitrarily short time was again in conflict with the semiboundedness of the Hamiltonian.

The discussion of Misra and Sinha is based on the following principal theorem.

Theorem of Misra and Sinha: Let $F(t)$, $t \geqslant 0$, be a strongly continuous contractive family of bounded operators in K (the Hilbert space of state vectors of the decaying system together with its decay products), and let, for all ψ in some dense set D,

$$\|[F(t+s) - F(t)F(s)]\psi\| \leqslant C_\psi t^\alpha s^\alpha, \qquad (1.2)$$

for $t,s \geqslant 0$, with $\alpha > 1$ and C_ψ a constant independent of t,s but depending on ψ. Then $F(t)$ forms a strongly continuous semigroup, for $t \geqslant 0$:

$$F(t+s) = F(t)F(s), \quad t,s \geqslant 0. \qquad (1.3)$$

Using this theorem, Misra and Sinha derived the conclusion mentioned above.

Following some techniques Misra and Sinha used in proving the above theorem, we shall derive a new theorem that is a generalization of the Misra–Sinha theorem.

II. A GENERALIZED THEOREM

We shall prove a theorem which includes the Misra–Sinha theorem as a special case.

Generalized Theorem: Let $F(t)$, $t \geqslant 0$ be a strongly continuous contractive family of bounded operators in K, and let, for all ψ in some dense set D,

$$\|[F(t+s) - F(t)F(s)]\psi\| \leqslant C_\psi t^\alpha s^\alpha (t+s)^\beta, \qquad (2.1)$$

for $t,s \geqslant 0$, with $\alpha > 1$, $2\alpha + \beta > 0$, and C_ψ a constant independent of t,s but depending on ψ. Then $F(t)$ forms a strongly continuous semigroup, for $t \geqslant 0$:

$$F(t+s) = F(t)F(s), \quad t,s \geqslant 0. \qquad (2.2)$$

Proof: Let $\{t_i\}_{i=1}^n$ be a set of positive numbers. Then

$$\begin{aligned}
&\left[F\left(\sum_{i=1}^n t_i\right) - \prod_{i=1}^n F(t_i)\right]\psi \\
&= \left[F\left(t_1 + \sum_{i=2}^n t_i\right) - F(t_1)\prod_{i=2}^n F(t_i)\right]\psi \\
&= \left[F\left(t_1 + \sum_{i=2}^n t_i\right) - F(t_1)F\left(\sum_{i=2}^n t_i\right)\right. \\
&\quad \left. + F(t_1)\left\{F\left(\sum_{i=2}^n t_i\right) - \prod_{i=2}^n F(t_i)\right\}\right]\psi.
\end{aligned}$$

Using (2.1) and triangle inequality we obtain

$$\begin{aligned}
&\left\| \left[F\left(\sum_{i=1}^n t_i\right) - \prod_{i=1}^n F(t_i)\right]\psi \right\| \\
&\leqslant C_\psi t_1^\alpha \left(\sum_{i=2}^n t_i\right)^\alpha \left(\sum_{i=1}^n t_i\right)^\beta \\
&\quad + \left\| \left[F\left(\sum_{i=2}^n t_i\right) - \prod_{i=2}^n F(t_i)\right]\psi \right\|.
\end{aligned}$$

We iterate this process and in the end substitute $t_i = t/n$ for all i. Then

$$\begin{aligned}
&\|[F(t) - F(t/n)^n]\psi\| \\
&\leqslant C_\psi \left(\frac{t}{n}\right)^{2\alpha+\beta} \sum_{i=1}^{n-1} (i+1)^\beta i^\alpha \\
&\leqslant C_\psi t^{2\alpha+\beta} n^{-2\alpha-\beta} \int_1^{n-1} x^\alpha (x+1)^\beta \, dx \\
&\leqslant C_\psi \frac{t^{2\alpha+\beta}}{\alpha+\beta+1} (n^{1-\alpha} - 2^{\alpha+\beta+1} n^{-2\alpha-\beta}), \qquad (2.3)
\end{aligned}$$

where we have assumed $\alpha + \beta + 1 \neq 0$.

The inequality (2.3) shows that the sequence of operators $[F(t/n)]^n$ converges strongly to $F(t)$ on D as $n \to \infty$.

 0022-2488/88/081860-02$02.50

Since we have assumed that $\|[F(t/n)]^n\| \leqslant 1$ and D is dense in K, we have

$$\operatorname*{s-lim}_{n\to\infty} F(t/n)^n = F(t), \quad t \geqslant 0, \tag{2.4}$$

on the whole of K.

From the relation (2.4) we can prove the semigroup property of $F(t)$ by making use of the paper of Misra and Sinha[5] or the book by Chernoff[6]:

$$F(t+s) = F(t)F(s), \quad \text{for all } t,s \geqslant 0. \qquad \text{Q.E.D.}$$

From the relation (2.2) it is found that for $\beta = 0$ our theorem is reduced to the theorem of Misra and Sinha. So following Misra and Sinha, we also make a corollary from (1.1) and our theorem.

Corollary: If for a dense set of vectors ψ in K

$$\|EU_s E^{\perp} U_t E\psi\| \leqslant C_\psi t^\alpha s^\alpha (t+s)^\beta, \quad t,s \geqslant 0, \tag{2.5}$$

with $\alpha > 1$, $2\alpha + \beta > 0$, $\alpha + \beta + 1 \neq 0$, then

$$EU_s E^{\perp} U_t = 0, \quad \text{for all } t,s \geqslant 0;$$

that is to say, EU_tE is a semigroup for $t \geqslant 0$.

If we define the function

$$R(t,s) \equiv \|EU_s E^{\perp} U_t E\psi\|,$$

it gives us an estimate of the rate of regeneration of the undecayed unstable states from the decay products. Combining $R(t,s)$ with (2.5), we obtain the bound on regeneration,

$$R(t,s) \leqslant C_\psi t^\alpha s^\alpha (t+s)^\beta, \quad t,s \geqslant 0, \tag{2.6}$$

with

$$\alpha > 1, \quad 2\alpha + \beta > 0, \quad \alpha + \beta + 1 \neq 0. \tag{2.7}$$

III. CONCLUSION

In this concluding discussion, $t,s > 0$ is assumed. We can easily understand that in the case $\beta > 0$ the condition (2.1) with (2.7) is sufficient, and in the case $\beta = 0$ it coincides with (1.2). For the case $\beta < 0$ the rhs of (2.1) or, equivalently, (2.6) satisfies the following inequality:

$$C_\psi t^\alpha s^\alpha / (t+s)^{-\beta} \leqslant 2^\beta C_\psi t^{\alpha+\beta/2} s^{\alpha+\beta/2}. \tag{3.1}$$

If β satisfies the inequality

$$\alpha + \tfrac{1}{2}\beta > 1, \tag{3.2}$$

with $\alpha > 1$, $2\alpha + \beta > 0$, $\alpha + \beta + 1 \neq 0$, then our theorem is subsumed by the Misra–Sinha theorem. However, if the inequality

$$\alpha + \tfrac{1}{2}\beta \leqslant 1, \tag{3.3}$$

with $\alpha > 1$, $2\alpha + \beta > 0$, $\alpha + \beta + 1 \neq 0$, holds for β, then the Misra–Sinha theorem does not hold but ours does. Moreover, we have to mention that if $\alpha + \frac{1}{2}\beta \leqslant 0$, then of course neither theorem holds.

[1] D. W. Williams, Commun. Math. Phys. **21**, 314 (1971).
[2] L. P. Horwitz, J. P. Marchand, and J. LaVita, J. Math. Phys. **12**, 2537 (1971).
[3] L. Fonda and G. C. Gihirardi, Nuovo Cimento A **7**, 180 (1972).
[4] K. Sinha, Helv. Phys. Acta **45**, 619 (1972).
[5] B. Misra and K. Sinha, Helv. Phys. Acta **50**, 99 (1977).
[6] P. R. Chernoff, *Product Formulas, Nonlinear Semigroups and Addition of Unbounded Operators, Memoirs of the American Mathematical Society* (American Mathematical Society, Providence, RI, 1974).

Field quantization for pure and mixed states and its classical limit

Pimon Ajanapon
Division of Mathematics and Computer Science, Northeast Missouri State University, Kirksville, Missouri 63501

(Received 8 December 1986; accepted for publication 13 April 1988)

The path-integral formulation of quantum mechanics in terms of the density matrix proposed in a previous paper is here extended to quantum field theory. The corresponding classical field equation is derived as the deterministic classical limit by imposing an appropriate mixed-state initial condition. Relativistic invariance is insured by construction of ten generators of the inhomogeneous Lorentz group.

I. INTRODUCTION

In a previous paper,[1] the author proposed an alternative approach to the classical-limit problem of quantum mechanics. The present paper extends the idea to quantum field theory, specifically quantum scalar field theory with self-interaction.

Although we assume that the reader has read the abovementioned paper, let us recapitulate to bring out the main ideas here. First, quantum mechanics is formulated in terms of the density matrix which includes both pure and mixed states. The law of motion is given in the form of an integral equation whose evolution kernel is the product of a Feynman probability transition amplitude and its Hermitian conjugate. The Feynman path integral is defined by means of nonstandard analysis,[2] giving a rigorous rendition for the concept in a way close to Feynman's original intuitive idea. In fact, as is shown in the previous paper, all one needs to do is to paraphrase Feynman in the language of nonstandard analysis. The equation of motion is then transformed to the mean and relative coordinates and, upon taking Fourier transforms with respect to the relative coordinates, we obtain a time-evolution equation for the phase-space representation of the density matrix (also known in the literature as the Wigner distribution function).

The approach to the classical limit is motivated by the physical interpretation of the elements of the density matrix. According to this interpretation, the diagonal elements give the probability of finding the system in a certain state and the off-diagonal elements are associated with the interference effects which are the manifestations of the wave nature of matter and hence purely quantum mechanical in nature. Thus, if quantum mechanics is to contain classical mechanics, all the classical information must lie along the diagonal of the density matrix. Furthermore, since the diagonal elements of a density matrix are all non-negative, their phase-space representation should give us a bona fide classical distribution. To extract out such a distribution from our quantum mechanical equation, we expand the potential with respect to the relative coordinates and retain only the linear part in the relative coordinates; this is done in the nonstandard universe. Again, this procedure is motivated by the fact that the classical information is contained only along the diagonal of the density matrix which becomes infinitesimally fuzzy in the nonstandard extension and first-order infinitesimals can be significant in the nonstandard universe as can small real numbers in the standard universe. When carried out in connection with the phase-space representation of the diagonal elements, it can also be interpreted as ignoring the quantum effects, or eliminating the "quantum forces." In any case, the procedure does provide us with the desired classical distribution satisfying the (classical) Liouville equation. We call this the *statistical classical limit* of quantum mechanics. To obtain the *deterministic classical limit*, i.e., Hamilton's or Newton's mechanics, we need only consider an initial distribution corresponding to the case of the complete knowledge of the motion. It is simply the way that one passes from classical statistical mechanics to deterministic classical mechanics.

Now, systems dealt with in quantum field theory are of infinite numbers of degrees of freedom. Such a system is characterized by a Lagrangian density which is a functional of a set of fields and the fields are functions of the space coordinates and of time satisfying the smoothness conditions that may be required in the particular problem. The field quantities now play the role of the generalized coordinates of the system with the space variables labeling the field quantities and hence enumerating the degrees of freedom of the system. Obviously, this infinity of degrees of freedom is uncountable. From the point of view of standard mathematical analysis, the transition from quantum mechanics, which deals only with systems of finite numbers of degrees of freedom, to quantum field theory, which deals with systems of uncountably infinite numbers of degrees of freedom, is therefore an enormous and delicate jump. Not so in nonstandard analysis! Recall that the cardinality of an infinite *finite set is uncountably infinite. This therefore suggests that we simply consider our field-theoretic systems to be of infinite *finite numbers of degrees of freedom. One way to accomplish this is to employ spatial Fourier decomposition of the fields (as usually done), nonstandardly extend the sequences of Fourier coefficients (which are functions of time), and then consider infinite *finite subsequences of these nonstandard extensions. Since such infinite *finite sequences of Fourier coefficients approximate the fields to within an infinitesimal,[3] they can be used to represent the fields for all practical purposes. Representing the fields by infinite *finite (instead of infinite) sequences of Fourier coefficients is particularly essential in our rigorous definition of the path integral, which avoids the limit problem entirely. So, the use of nonstandard analysis completely justifies the formal way we go

 0022-2488/88/081862-04$02.50

from quantum mechanics to quantum field theory. It is simply a matter of going from the finite to the infinite *finite.

II. QUANTIZATION

To illustrate the above ideas concretely and in a simple way, we consider quantum scalar field theory with self-interaction in this paper and leave the more complicated situation involving both bosonic and fermionic fields to a future paper. Suppose then that we have a system characterized by the Lagrangian density

$$\mathscr{L} = \tfrac{1}{2}(\partial_\mu \phi)^2 - \mathscr{U}[\phi], \tag{1}$$

where $\phi(\mathbf{r},t)$ is a real scalar field assumed to be sufficiently smooth and expandable at each instant of time up to an infinitesimal with respect to the family

$$\left\{\left(\frac{1}{\Omega}\right)^{1/2}, \left(\frac{2}{\Omega}\right)^{1/2} \cos\frac{2\pi}{l}(n_1x+n_2y+n_3z), \left(\frac{2}{\Omega}\right)^{1/2} \sin\frac{2\pi}{l}(n_1x+n_2y+n_3z)\right\}^N_{n_1,n_2,n_3=1}, \tag{2}$$

where Ω is the volume of a cube with edge length l and center at the origin and N an infinite *finite integer. Specifically, we write

$$\phi(\mathbf{r},t) \approx \left(\frac{1}{\Omega}\right)^{1/2} a_0(t) + \left(\frac{2}{\Omega}\right)^{1/2} \times \sum_{n_1,n_2,n_3=1}^{N} \left[a_{n_1n_2n_3}(t)\cos\frac{2\pi}{l}(n_1x+n_2y+n_3z) + b_{n_1n_2n_3}(t)\sin\frac{2\pi}{l}(n_1x+n_2y+n_3z)\right], \tag{3}$$

where $\approx$ is the "infinitely close" relation. Setting

$$q = (q_1,\dots,q_n) = (1/c)(a_0,a_{111},b_{111},\dots,a_{NNN},b_{NNN}), \tag{4}$$

the Lagrangian

$$L = \int_\Omega \mathscr{L}\, d^3r \tag{5}$$

takes the form

$$L = \tfrac{1}{2}\dot{q}^2 - V(q), \tag{6}$$

where

$$V(q) = \int_\Omega \mathscr{V}[\phi] d^3r, \tag{7}$$

and

$$\mathscr{V}[\phi] = \tfrac{1}{2}(\nabla\phi)^2 + \mathscr{U}[\phi]. \tag{8}$$

The Lagrangian L in (6) is in exactly the same form as that considered in the previous paper. The difference is that the q here is an infinite *finite n-tuple, while it is just a finite n-tuple in the previous paper. We can now proceed to quantize our field-theoretic system in the same way as we do our quantum mechanical system in the previous paper. Let ω be a positive infinite *finite integer. Divide the time interval $t^b - t^a$ into ω equal parts of infinitesimal duration $\epsilon = (t^b - t^a)/\omega$ and denote the successive times by $t^0 = t^a, t^1, \dots, t^\omega = t^b$ with $t^k - t^{k-1} = \epsilon$. Specify each path $q(t)$ by the *finite sequence of points $q^0 = q(t^0)$, $q^1 = q(t^1),\dots,q^\omega = q(t^\omega)$. Then define the infinitesimal probability transition amplitude by

$$K(q^k,q^{k-1};\epsilon) = (1/A)\exp\{(i/\hbar)S_\epsilon(q^k,q^{k-1})\}, \tag{9}$$

where $S_\epsilon(q^k,q^{k-1})$ is the infinitesimal action from time $t^{k-1} = t^a + (k-1)\epsilon$ to time $t^k = t^a + k\epsilon$ defined by

$$S_\epsilon(q^k,q^{k-1}) = \frac{\epsilon}{2}\left[L\left(q^k, \frac{q^k - q^{k-1}}{\epsilon}\right) + L\left(q^{k-1}, \frac{q^k - q^{k-1}}{\epsilon}\right)\right]. \tag{10}$$

Now, define the finite probability transition amplitude in terms of the infinitesimal amplitude as the following improper *multiple integral:

$$K(q^b,q^a;t^b - t^a) = \int \prod_{k=1}^{\omega} K(q^k,q^{k-1};\epsilon) \prod_{k=1}^{\omega-1} dq^k. \tag{11}$$

Specify the general (pure or mixed) state by a density matrix $\rho(q',q'',t)$ and postulate its time evolution by

$$\rho(q^{b'},q^{b''},t^b) = \int K(q^{b'},q^{a'};t^b - t^a) \times K^\dagger(q^{b''},q^{a''};t^b - t^a)\rho(q^{a'},q^{a''},t^a)dq^{a'}dq^{a''}. \tag{12}$$

III. CLASSICAL LIMIT

Following the procedure given in the previous paper and outlined in Sec. I, we can now derive the statistical classical limit of our quantum field theory and then the deterministic classical limit. In sum, by starting out initially with an appropriate mixed state we obtain in the classical limit a phase-space trajectory $(q^c(t),p^c(t))$ satisfying the initial-value Hamiltonian system

$$\dot{q} = [q,H], \quad q(0) = q^0, \tag{13a}$$

$$\dot{p} = [p,H], \quad p(0) = p^0, \tag{13b}$$

where

$$H = \tfrac{1}{2}p^2 + V(q). \tag{14}$$

Now, form infinite *finite Fourier sums with respect to the family (2) by using the components of the infinite *finite-tuples $q^c(t),p^c(t),q^0,p^0$ as coefficients and denote the standard parts of these sums by $\phi_c(\mathbf{r},t),\pi_c(\mathbf{r},t),\phi_0(\mathbf{r}),\pi_0(\mathbf{r})$, respectively. Then the fact that $q^c(t)$ and $p^c(t)$ satisfy (13a) and (13b) implies that $\phi_c(\mathbf{r},t)$ and $\pi_c(\mathbf{r},t)$ satisfy

$$\frac{1}{c}\frac{\partial\phi}{\partial t}(\mathbf{r},t) = \pi(\mathbf{r},t), \quad \phi(\mathbf{r},0) = \phi_0(\mathbf{r}), \tag{15a}$$

$$\frac{1}{c}\frac{\partial\pi}{\partial t}(\mathbf{r},t) = -\int_\Omega \frac{\delta\mathscr{V}[\phi(\mathbf{r}',t)]}{\delta\phi(\mathbf{r},t)} d^3r', \quad \pi(\mathbf{r},0) = \pi_0(\mathbf{r}). \tag{15b}$$

Taking the functional derivative of $\mathscr{V}$ as given by (8), we have

$$\frac{\delta\mathscr{V}[\phi(\mathbf{r}',t)]}{\delta\phi(\mathbf{r},t)} = \nabla\phi(\mathbf{r}',t)\cdot\nabla\delta(\mathbf{r}-\mathbf{r}') + \frac{\delta\mathscr{U}[\phi(\mathbf{r}',t)]}{\delta\phi(\mathbf{r},t)}. \tag{16}$$

Substituting (16) into the first part of (15b) and integrating with respect to r', we obtain

$$\frac{1}{c}\frac{\partial\pi}{\partial t}(\mathbf{r},t) = \nabla^2\phi(\mathbf{r},t) - \mathscr{U}'[\phi(\mathbf{r},t)], \tag{17}$$

where we have set

$$\mathscr{U}'[\phi(\mathbf{r},t)] = \int_\Omega \frac{\delta\mathscr{U}[\phi(\mathbf{r}',t)]}{\delta\phi(\mathbf{r},t)} d^3r'. \tag{18}$$

The first part of (15a) and (17) can be combined to give

$$\nabla^2\phi - \frac{1}{c^2}\frac{\partial^2\phi}{\partial t^2} = \mathscr{U}'[\phi], \tag{19}$$

which is the classical field equation corresponding to the Lagrangian density $\mathscr{L}$ given by (1). We have therefore obtained for the classical limit of our quantum scalar field theory with self-interaction a deterministic classical field $\phi_c(\mathbf{r},t)$ satisfying the field equation (19) and the following initial conditions:

$$\phi(\mathbf{r},0) = \phi_0(\mathbf{r}), \quad \frac{1}{c}\frac{\partial\phi}{\partial t}(\mathbf{r},0) = \pi_0(\mathbf{r}). \tag{20}$$

It should be emphasized that the field ϕ is taken as a standard field and consequently the field equation (19) is also standard. It is interesting to note in this connection that if we were to start out with a renormalized Lagrangian density containing infinite counterterms, we would obtain in the classical limit a deterministic classical field satisfying a field equation similar to (19), but containing coefficients which are infinite.

IV. RELATIVISTIC INVARIANCE

Having chosen to work in the "instant form" of dynamics (as Dirac[4] calls it), we have forgone with the manifestness of relativistic invariance. Nevertheless, we must insure that our physical theories obey the principle of special relativity, i.e., the laws of physics must be invariant under changes of reference frame. This is a symmetry requirement under the inhomogeneous Lorentz transformation group. Manifest invariance is on the other hand a more special requirement which exploits the space-time symmetry of special relativity. It requires that certain quantities transform under changes of reference frame in a manner that is intimately related to the Lorentz transformation of space-time events.[5,6]

The symmetry requirement under the inhomogeneous Lorentz group can be met simply by constructing generators of the group. In other words, we can introduce relativistic symmetry into our theory by constructing the ten generators H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ which are the generators of time translations, space translations, space rotations, and pure Lorentz transformations, respectively. In order for the quantities H,P_i,J_i,K_i, for $i = 1,2,3$ to be generators of the inhomogeneous Lorentz group, they must satisfy the following Lie-bracket relations:

$$\begin{aligned}
&[P_i,H] = 0, \quad [P_i,P_j] = 0, \quad [J_i,H] = 0,\\
&[J_i,J_j] = \epsilon_{ijk}J_k, \quad [J_i,P_j] = \epsilon_{ijk}P_k,\\
&[J_i,K_j] = \epsilon_{ijk}K_k, \quad [K_i,H] = P_i,\\
&[K_i,P_j] = \delta_{ij}H, \quad [K_i,K_j] = -\epsilon_{ijk}J_k,
\end{aligned} \tag{21}$$

where ϵ_{ijk} is the Levi-Civita three-index symbol and the summation convention for repeated indices has been employed. The procedure of constructing such generators is by now a familiar task, which owes much to the advent of quantum mechanics.

As an illustration of the above idea, let us see how we can insure that our classical limit is relativistically invariant. The basic dynamical variables are the fields ϕ and π and the dynamical equations are the first part of (15a) and (17) which can be put in the form

$$\frac{1}{c}\frac{\partial\phi}{\partial t} = [\phi,H], \tag{22a}$$

$$\frac{1}{c}\frac{\partial\pi}{\partial t} = [\pi,H], \tag{22b}$$

where [,] is the field Poisson bracket defined for any pair of dynamical functionals $F[\phi,\pi]$ and $G[\phi,\pi]$ by

$$[F,G] = \int_\Omega \left\{\frac{\delta F}{\delta\phi(\mathbf{r})}\frac{\delta G}{\delta\pi(\mathbf{r})} - \frac{\delta F}{\delta\pi(\mathbf{r})}\frac{\delta G}{\delta\phi(\mathbf{r})}\right\} d^3r, \tag{23}$$

in which the "equal time" t has been, and will be, suppressed and H is the field Hamiltonian given by

$$H[\phi,H] = \int_\Omega \left\{\frac{1}{2}\pi^2(\mathbf{r}) + \frac{1}{2}[\nabla\phi(\mathbf{r})]^2 + \mathscr{U}[\phi(\mathbf{r})]\right\} d^3r. \tag{24}$$

This field Hamiltonian H is the generator of time translations and the field Poisson bracket defined by (23) is the Lie bracket in our theory. The other nine generators are as follows:

$$P_i[\phi,\pi] = \int_\Omega \pi(\mathbf{r})\partial_i\phi(\mathbf{r})d^3r, \tag{25}$$

$$J_i[\phi,\pi] = \int_\Omega -\epsilon_{ijk}x_j\pi(\mathbf{r})\partial_k\phi(\mathbf{r})d^3r, \tag{26}$$

$$K_i[\phi,\pi] = \int_\Omega -x_i\left\{\frac{1}{2}\pi^2(\mathbf{r}) + [\nabla\phi(\mathbf{r})]^2 + \mathscr{U}[\phi(\mathbf{r})]\right\} d^3r. \tag{27}$$

It is straightforward, but lengthy, to verify that the ten quantities given by (24)–(27) indeed satisfy all the relations given by (21).

At the quantum level, the dynamical equation is given by (12). In a way which is similar to that of Feynman's derivation of Schrödinger's wave equation from his path-integral formulation of quantum mechanics,[7] we obtain from (12) the following differential equation for the density matrix $\rho(q',q'',t)$:

$$i\hbar\frac{\partial\rho}{\partial t} = (H' - H'')\rho(q',q'',t), \tag{28}$$

where

$$H' = -\frac{\hbar^2}{2}\left(\frac{\partial}{\partial q'}\right)^2 + V(q'), \tag{29}$$

and

$$H'' = -\frac{\hbar^2}{2}\left(\frac{\partial}{\partial q''}\right)^2 + V(q''). \tag{30}$$

If we define the Hamiltonian matrix by

$$H(q',q'') = \left[-\frac{\hbar^2}{2}\frac{\partial}{\partial q'}\cdot\frac{\partial}{\partial q''} + V(q'')\right]\delta(q'-q''), \tag{31}$$

then (28) can be written in the form

$$i\hbar \frac{\partial \rho}{\partial t} = \int \{H(q',q)\rho(q,q'',t) - \rho(q',q,t)H(q,q'')\}dq, \tag{32}$$

which leads us to the von Neumann operator equation

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = -[\hat{\rho},\hat{H}], \tag{33}$$

where the Lie bracket on the right is now the commutator and the operator $\hat{H}$ the generator of time translations. The rest of the story is a familiar one and will not be repeated here.

ACKNOWLEDGMENTS

This work was supported by Northeast Missouri State University Faculty Research Grants, Summer 1986. Parts of the work were initially carried out at Purdue University in 1984.

[1]P. Ajanapon, Am. J. Phys. **55**, 159 (1987).

[2]An outline of nonstandard analysis is given in Sec. II of the paper listed above (Ref. 1) where many references on the subject can be found.

[3]*Contributions to Non-Standard Analysis*, edited by W. A. J. Luxemburg and A. Robinson, *Studies in Logic and the Foundations of Mathematics*, Vol. 69 (North-Holland, Amsterdam, 1972), pp. 15–39.

[4]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).

[5]L. L. Foldy, Phys. Rev. **122**, 275 (1961).

[6]D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).

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

Multipole structure of stationary gravitational fields at null infinity

Prasun K. Kundu
Department of Physics and Astronomy, Ohio University, Athens, Ohio 45701

(Received 12 January 1988; accepted for publication 16 March 1988)

The multipole structure of the gravitational field produced by a stationary, isolated source at null infinity is investigated using a Bondi-type null polar coordinate system. The results are connected with the previous works of the author on the multipole expansion of stationary, asymptotically flat, vacuum solutions of the Einstein equations in the framework of the Geroch–Hansen one-point compactification scheme. The present work strongly suggests that all such solutions are analytic in the inverse luminosity distance r^{-1} in a Bondi chart.

I. INTRODUCTION

The structure of the gravitational field produced by a bounded source has been and still remains an important area of investigation in classical general relativity.[1] Within the framework of general relativity, the gravitational field of an isolated system is mathematically modeled by the class of asymptotically flat space-times, i.e., space-times whose metric approaches the Minkowski metric in a suitable sense at large distances from the source. The limiting behavior of the metric can be studied in two distinct regimes at infinitely large spacelike and null separations from a localized source. Pioneering investigations of the structure of the far field were carried out by Bondi, van der Burg, and Metzner[2] and Sachs[3] at null infinity which demonstrated explicitly how the mass of a radiating system (the Bondi mass) is carried away by the gravitational radiation. At about the same time, a series of papers by Arnowitt, Deser, and Misner[4] (ADM) and Bergmann[5] examined the structure at spacelike infinity, which led to a notion of the total energy of the system (the ADM mass) including the sources and the gravitational field. A geometric formulation of the asymptotic structure at null infinity, due to Penrose,[6] in terms of a suitable conformal completion of the space-time, has laid the foundation of many subsequent investigations leading to significant advances in gravitational radiation theory. In particular, it has allowed a coordinate-free description of the relevant physical concepts, such as the Bondi mass of an isolated system and the Bondi–Metzner–Sachs (BMS) asymptotic symmetry group.[7,8] The main idea of Penrose's method is to rescale the space-time metric in such a way that the points that are infinitely far away in terms of the physical metric are now brought at finite distance to form a boundary of the space-time (usually denoted by $\mathscr{I}$). The study of the asymptotic structure of the physical space-time is then conveniently reduced to a study of the local geometry near $\mathscr{I}$. Geroch[9] extended the ADM work on the spacelike infinity along the lines of Penrose's conformal approach. His method was to consider the initial value formulation of the field equations and construct a conformal completion of the initial data set. In the Geroch formalism the entire spacelike infinity was represented by a single point attached to a spacelike hypersurface whose induced metric and extrinsic curvature subject to the usual constraint equations constitute the initial data for the time evolution of the gravitational field. Finally, Ashtekar and Hansen[10] gave a unified treatment of the asymptotic structure in the two regimes by introducing a further reformulation of spacelike infinity in a fully four-dimensional setting, in which it is regarded as a point i^0 attached to $\mathscr{I}$ to form a part of the space-time boundary. However, the limiting behavior of the appropriate rescaled physical fields at i^0 is considerably more intricate than their behavior on $\mathscr{I}$ even for stationary metrics. The Ashtekar–Hansen formalism has provided a natural framework for relating the various physical quantities in the two asymptotic regimes, such as the ADM and the Bondi–Sachs four-momenta,[11] and has led to a number of other interesting results dealing with conserved quantities in asymptotically flat space-time admitting Killing vectors[12] and the notion of angular momentum of stationary systems.[13]

In spite of these impressive achievements, a systematic detailed understanding of the higher-order asymptotic behavior of a general time-dependent gravitational field still remains an elusive goal. In recent years, however, further progress has been made towards a fairly complete understanding of one important special case, namely, the stationary vacuum fields. In the case of Newtonian gravity, the gravitational scalar potential satisfying Laplace's equation in the flat three-dimensional Euclidean space is completely described by the familiar multipole expansion. Because of the large coordinate freedom and the absence of a fixed background geometry, the significance of such an expansion is not immediately obvious in the context of general relativity. In order to unambiguously extract the physical properties of a particular stationary vacuum metric, Geroch[14] and Hansen[15] defined, in a manifestly coordinate independent manner, a set of quantities that can be regarded as the relativistic analogs of the usual Newtonian multipole moments. Utilizing the fact that the space-time now admits a Killing vector that is timelike at least near infinity, they introduced a three-dimensional manifold which is the set of all orbits of the Killing vector field. The gravitational field is now described by the induced metric on the orbit manifold together with a pair of scalar potentials constructed from the norm and the twist of the Killing field. A suitable conformal rescaling of the metric allows one to attach a single point Λ to the orbit manifold which then represents the entire spatial infinity. The physically required fall-off conditions governing the rate at which the metric must approach the flat Euclidean metric at large distance are correctly captured in a coordi-

 0022-2488/88/081866-13$02.50

nate independent way by demanding the existence of a conformal factor suitably vanishing at Λ, so that the appropriate rescaled potentials and three-metric define a set of smooth fields on the conformal completion of the orbit manifold. The asymptotic behavior of the physical fields at large distances is translated into the degree of smoothness of the corresponding rescaled fields near Λ. The Geroch–Hansen multipole moments are then defined recursively as values at Λ of a hierarchy of trace-free symmetric tensors constructed from these rescaled fields. The conceptual significance of these quantities lies in the recently established fact[16,17] that they completely characterize the local structure of a stationary, asymptotically flat (AF), exact solution of the full nonlinear Einstein equations outside matter. Specifically, it was shown that, for a certain intrinsic choice of the conformal factor and rather weak asymptotic flatness conditions, the rescaled fields are indeed analytic functions in a neighborhood of Λ in a suitable coordinate system. This analyticity property guarantees the existence of a convergent Taylor expansion of the rescaled fields around Λ with a nonzero radius of convergence, whose coefficients can be explicitly evaluated in terms of the Geroch–Hansen moments.

In spite of the elegance of the Geroch–Hansen approach, the fact that it requires a specially devised three-dimensional formalism suited only for stationary fields, has so far prevented one from making contact with previous extensive investigations on structure at null infinity carried out in a fully four-dimensional setting. The purpose of this paper is to present some preliminary results of a recently undertaken investigation whose eventual goal is to bring these two apparently dissimilar approaches closer together. A rather simple and straightforward way of relating the two conformal completions is given in Sec. II following a brief review of the various relevant formalisms. For simplicity, we shall restrict ourselves only to space-times with asymptotically Minkowskian topology. Thus, for instance, space-times with a nonzero angular momentum monopole moment, such as the Newman–Unti–Tamburino (NUT) metric are excluded from consideration. In Sec. III a multipole expansion of a stationary vacuum metric is presented that utilizes the particular conformal compactification scheme of Ref. 16. It is similar to an expansion previously developed by the author[18] under somewhat more restrictive assumptions. Finally, in Sec. IV the asymptotic behavior of a stationary vacuum space-time at null infinity is examined in detail in a Bondi-type null polar coordinate system and is expressed in terms of the Geroch–Hansen moments. A coordinate transformation connecting the usual Bondi coordinates defined in the physical space-time to the Riemann normal coordinates used in Sec. III is derived in Appendix A. Appendix B outlines the computation of a few of the leading multipole moments of the static axisymmetric Weyl space-times.

II. PRELIMINARIES

Let us consider a stationary space-time $(\mathscr{M}, g_{\mu\nu})$ (of Lorentz signature $+ - - -$) satisfying the vacuum field equations $R_{\mu\nu} = 0$. We assume $\mathscr{M}$ to be asymptotically flat at null infinity (more precisely, weakly asymptotically simple) in the sense of Penrose. By means of a conformal transformation

$$\hat{g}_{\mu\nu} = \Omega^2 g_{\mu\nu} \tag{2.1}$$

one can, therefore, construct the conformal completion $\hat{\mathscr{M}} = \mathscr{M} \cup \mathscr{I}$, so that the rescaled metric $\hat{g}_{\mu\nu}$ has a smooth (i.e., C^∞) extension to the boundary $\partial\hat{\mathscr{M}} = \mathscr{I}$ consisting of two disjoint pieces $\mathscr{I}^+$ and $\mathscr{I}^-$ each topologically $S^2 \times \mathbb{R}$. We shall be mostly concerned here with the behavior at $\mathscr{I}^+$, the future null infinity. The conformal factor Ω is assumed to be a scalar function regular (at least C^2) everywhere on $\mathscr{I}$, which is strictly positive on $\hat{\mathscr{M}} - \mathscr{I}$ and satisfies the conditions

$$[\Omega]_{\mathscr{I}} = 0, \quad [\Omega_{;\mu}]_{\mathscr{I}} \neq 0, \tag{2.2}$$

where the semicolon represents the action of $\hat{\nabla}_\mu$, the (torsion-free) covariant derivative operator associated with $\hat{g}_{\mu\nu}$. We also assume that the generators of $\mathscr{I}$, i.e., the integral curves of the vector field $\hat{g}^{\mu\nu}\Omega_{;\nu}$, are complete.[19] Moreover, since we are interested in stationary space-times, it is natural to assume that $\mathscr{L}_\xi \Omega = 0$, where ξ^μ is the timelike Killing vector in $\mathscr{M}$ satisfying $\mathscr{L}_\xi g_{\mu\nu} = 0$. Then the vector field ξ^μ also satisfies the Killing equation $\mathscr{L}_\xi \hat{g}_{\mu\nu} = 0$ for the conformal metric $\hat{g}_{\mu\nu}$ on $\hat{\mathscr{M}} - \mathscr{I}$ and can be extended smoothly to $\mathscr{I}$. The Ricci tensors of $g_{\mu\nu}$ and $\hat{g}_{\mu\nu}$ are related by the formula[20]

$$S_{\mu\nu} = \hat{S}_{\mu\nu} - 2\Omega^{-1}\Omega_{\mu;\nu} + \Omega^{-2}\hat{g}_{\mu\nu}\hat{g}^{\alpha\beta}\Omega_\alpha\Omega_\beta, \tag{2.3}$$

where $\Omega_\mu = \Omega_{;\mu}$, $S_{\mu\nu} = R_{\mu\nu} - \frac{1}{6}Rg_{\mu\nu}$, and $\hat{S}_{\mu\nu}$ is similarly defined from the unphysical Ricci tensor $\hat{R}_{\mu\nu}$. The field equations in empty space $S_{\mu\nu} = 0$ applied to Eq. (2.3) (more generally, an asymptotic emptiness condition $\Omega^2 S_{\mu\nu} = 0$ on $\mathscr{I}$ suffices) and evaluated at $\mathscr{I}$ then implies that $\mathscr{I}$ is null:

$$\hat{g}^{\mu\nu}\Omega_\mu\Omega_\nu = 0. \tag{2.4}$$

By readjustment of the conformal factor $\Omega' = \omega\Omega$, one can easily show that it is always possible to choose Ω so that the scalar field

$$v = \Omega^{-1}\hat{g}^{\mu\nu}\Omega_\mu\Omega_\nu \tag{2.5}$$

vanishes on $\mathscr{I}$,

$$[v]_{\mathscr{I}} = 0. \tag{2.6}$$

The relation

$$\Omega S_{\mu\nu} = \Omega\hat{S}_{\mu\nu} - 2\Omega_{\mu;\nu} + v$$

then immediately shows that Ω must satisfy the condition

$$[\Omega_{\mu;\nu}]_{\mathscr{I}} = 0, \tag{2.7}$$

which states that with this conformal structure $\mathscr{I}$ is divergence- and shear-free. It turns out that we need to impose a slightly stronger condition

$$[\Omega^{-1}v]_{\mathscr{I}} = 1 \tag{2.8}$$

on $\mathscr{I}$ in order to recover the appropriate asymptotic conditions in the orbit manifold. The norms of ξ^μ with respect to the two metrics $g_{\mu\nu}$ and $\hat{g}_{\mu\nu}$ are

$$\lambda = g_{\mu\nu}\xi^\mu\xi^\nu \tag{2.9}$$

and

$$\hat{\lambda} = \hat{g}_{\mu\nu}\xi^\mu\xi^\nu = \Omega^2\lambda. \tag{2.10}$$

For an asymptotically flat stationary metric one expects λ to remain finite on $\mathscr{I}$. Then it follows immediately that $[\hat{\lambda}]_{\mathscr{I}} = 0$, i.e., ξ^μ is null on $\mathscr{I}$ in the conformal space $(\hat{\mathscr{M}}, \hat{g}_{\mu\nu})$. Later we shall require a stronger condition

$$[\lambda]_{\mathscr{I}} = 1. \tag{2.11}$$

Next we briefly recall the basic features of the Geroch–Hansen formalism. One begins by constructing the orbit manifold $\mathscr{V} = \mathscr{M}/\mathscr{G}$, where $\mathscr{G}$ is the one-parameter timelike group of motion generated by the Killing vector field ξ^μ. The equivalence relation $\sim$ induced by the action of $\mathscr{G}$ on points of $\mathscr{M}$ defines a map

$$\pi : \mathscr{M} \to \mathscr{V} = \mathscr{M}/\mathscr{G} = \{[x] | x \in \mathscr{M}\},$$

which sends a point $x \in \mathscr{M}$ to the equivalence class $[x] \equiv \pi(x) = \{x' \in \mathscr{M} | x' \sim x\}$ representing the unique trajectory, or orbit, of ξ^μ that passes through x. Given a smooth function $\tilde{f} : \mathscr{V} \to \mathbb{R}$, the action of π induces a function $f = \tilde{f} \circ \pi = \pi^* \tilde{f}$ on $\mathscr{M}$ which has the property $\mathscr{L}_\xi f = \xi^\mu \partial_\mu f = 0$. This correspondence between functions on $\mathscr{V}$ and $\mathscr{M}$ extends in a straighforward manner to tensor fields. Geroch[21] has shown that a tensor field $T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta}$ on $\mathscr{M}$ satisfying

$$\mathscr{L}_\xi T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta} = 0,$$

$$g_{\alpha\lambda} \xi^\lambda T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta} = 0, \ldots, \xi^\delta T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta} = 0, \tag{2.12}$$

can be regarded as a tensor field on $\mathscr{V}$. In particular, the tensor field

$$h_{\mu\nu} = -g_{\mu\nu} + \lambda^{-1} g_{\mu\alpha} g_{\nu\beta} \xi^\alpha \xi^\beta, \tag{2.13}$$

satisfying

$$\mathscr{L}_\xi h_{\mu\nu} = 0, \quad h_{\mu\nu} \xi^\nu = 0, \tag{2.14}$$

defines a positive definite metric on $\mathscr{V}$. For the purpose of later discussions, it is convenient to introduce a coordinate chart on the three-manifold $\mathscr{V}$ that explicitly labels the Killing trajectories. This is simply achieved by introducing a linearly independent set of three scalar functions $f^i(x)$ $(i = 1,2,3)$ that remain constant along a particular trajectory, i.e., satisfy the equation $\mathscr{L}_\xi f^i = 0$. Their values $f^i(x_0) = c^i$ at a point x_0 unambiguously label the trajectory $\pi(x_0)$ that passes through x_0. Associated with such a solution set $f^i(x)$ there are three functions y^i: $\mathscr{V} \to \mathbb{R}^3$ such that $f^i = \pi^* y^i$ which define a coordinate chart on $\mathscr{V}$. Different choices of solution sets $f^i(x)$ and $f'^i(x)$ thus amount to different choices of coordinates in $\mathscr{V}$. Along with ξ^μ and $\gamma^i{}_\mu := \partial_\mu f^i$, the quantities $\gamma_i{}^\mu$, defined through the relations

$$g_{\mu\nu} \xi^\mu \gamma_i{}^\nu = 0, \quad \gamma^i{}_\mu \gamma_j{}^\mu = \delta^i_j, \tag{2.15}$$

constitute a complete set of projection tensors. Since $\mathscr{L}_\xi \gamma^i{}_\mu = 0$ and from the Killing equation and Eq. (2.15) it also follows that $\mathscr{L}_\xi \gamma_i{}^\mu = 0$, a tensor field $T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta}$ satisfying Eq. (2.12) gives rise to a tensor

$$T^{a\cdots b}{}_{c\cdots d} := \gamma^a{}_\alpha \cdots \gamma^b{}_\beta \gamma_c{}^\gamma \cdots \gamma_d{}^\delta T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta} \tag{2.16}$$

satisfying $\mathscr{L}_\xi T^{a\cdots b}{}_{c\cdots d} = 0$, i.e., a tensor field defined on the orbit space $\mathscr{V}$. Acting on tensor fields on $\mathscr{V}$, $\mathscr{L}_\xi$ is simply the directional derivation $\xi^\mu \partial_\mu$. The resulting formalism[22] is similar to that of Geroch; the labeling of "four-dimensional" tensor fields on $\mathscr{M}$ and "three-dimensional" tensor fields on $\mathscr{V}$ by different indices, Greek and Latin, is merely a matter of notational convenience.

The covariant derivative operator ∇_μ on $\mathscr{M}$ defines a covariant derivative ∇_m on $\mathscr{V}$ with all the usual properties. One defines

$$\nabla_m T^{a\cdots b}{}_{c\cdots d} := \gamma_m{}^\mu \gamma^a{}_\alpha \cdots \gamma^b{}_\beta \gamma_c{}^\gamma \cdots \gamma_d{}^\delta \nabla_\mu T^{\alpha\cdots\beta}{}_{\gamma\cdots\delta}. \tag{2.17}$$

The induced metric on $\mathscr{V}$

$$h_{ij} := \gamma_i{}^\mu \gamma_j{}^\nu h_{\mu\nu} \tag{2.18}$$

is compatible with ∇_m, i.e.,

$$\nabla_m h_{ij} = 0. \tag{2.19}$$

From the alternating tensor on $\mathscr{M}$ one can construct the alternating tensor on $\mathscr{V}$:

$$e_{ijk} = \lambda^{-1/2} \gamma_i{}^\mu \gamma_j{}^\nu \gamma_k{}^\rho \xi^\sigma e_{\mu\nu\rho\sigma}. \tag{2.20}$$

The twist vector ω_μ, defined by expression[23]

$$\omega_\mu := e_{\mu\nu\rho\sigma} \xi^\nu g^{\lambda[\rho} \nabla_\lambda \xi^{\sigma]}, \tag{2.21}$$

obeys $\xi^\mu \omega_\mu = 0$, $\mathscr{L}_\xi \omega_\mu = 0$ and thus defines a three-vector

$$\omega_i = \gamma_i{}^\mu \omega_\mu \tag{2.22}$$

on $\mathscr{V}$. It is related to the antisymmetric tensor

$$\phi^{ij} = \gamma^i{}_\rho \gamma^j{}_\sigma g^{\lambda[\rho} \nabla_\lambda \xi^{\sigma]} \tag{2.23}$$

through the formula

$$\omega_i = \lambda^{1/2} e_{ijk} \phi^{jk}. \tag{2.24}$$

The quantity ω_i or, equivalently, ϕ^{ij} measures the degree of failure of ξ^μ to be hypersurface orthogonal.

Finally, mainly for the purpose of future references, we give formulas for the nonzero projected tetrad components of the Riemann tensor[24] of the space-time metric $g_{\mu\nu}$ in terms of tensor fields on $\mathscr{V}$:

$$\gamma_{[i}{}^\mu \gamma_{j]}{}^\nu \gamma_{[k}{}^\rho \gamma_{l]}{}^\sigma R_{\mu\nu\rho\sigma} = {}^3R_{ijkl} + 2\lambda^{-1}(\phi_{i[k}\phi_{l]j} - \phi_{ij}\phi_{kl}), \tag{2.25}$$

$$\xi^\mu \gamma_j{}^\nu \gamma_k{}^\rho \gamma_l{}^\sigma R_{\mu\nu\rho\sigma} = \lambda^{-1} \phi_{j[k} \nabla_{l]} \lambda - \nabla_j \phi_{kl}, \tag{2.26}$$

$$\xi^\mu \gamma_j{}^\nu \xi^\rho \gamma_l{}^\sigma R_{\mu\nu\rho\sigma} = -\lambda^{1/2} \nabla_j \nabla_l (\lambda^{1/2}) + h^{mn} \phi_{jm} \phi_{ln}, \tag{2.27}$$

where ${}^3R_{ijkl}$ is the Riemann tensor of the metric h_{ij} and it is understood that $\phi_{ij} = h_{ik} h_{jl} \phi^{kl}$. Using these formulas the stationary field equations can be written down in terms of fields on $\mathscr{V}$, namely λ, ϕ^{ij}, and h_{kl}, and the projected components of the energy-momentum tensor of matter.

A major simplification of the equations is achieved in the case of the field in vacuum (or in the presence of certain special matter distributions); the relation

$$\xi^\mu \gamma_i^\nu R_{\mu\nu} = \lambda^{-1/2} h_{ik} \nabla_j (\lambda^{1/2} \phi^{jk})$$

implies that

$$\nabla_j (\lambda^{1/2} \phi^{jk}) = 0,$$

or equivalently,

$$\nabla_{[i} \omega_{j]} = 0. \tag{2.28}$$

The latter, in turn, shows that ω_i is (locally) the gradient of a scalar ω, usually called the twist potential: $\omega_i = \nabla_i \omega$. The content of the vacuum Einstein equations can now be expressed as a set of coupled second-order equations for the fields λ, ω, and h_{ij}. It is, however, convenient to reformulate them in terms of the two Hansen potentials

$$\Phi_M = \tfrac{1}{4}\lambda^{-1}(\lambda^2 + \omega^2 - 1), \tag{2.29}$$

$$\Phi_J = \tfrac{1}{2}\lambda^{-1}\omega \tag{2.30}$$

and a rescaled metric

$$\gamma_{ij} = \lambda h_{ij}. \tag{2.31}$$

The field equations then take the form[15]

$$[\gamma^{ij}D_iD_j - 2\mathscr{R}]\Phi = 0, \quad \Phi := -\Phi_M + i\Phi_J, \tag{2.32}$$

$$\mathscr{R}_{ij} = 2[D_i\Phi D_j\Phi^*] - (1 + 4|\Phi|^2)^{-1}D_i|\Phi|^2D_j|\Phi|^2, \tag{2.33}$$

where D_i is the (torsion-free) covariant derivative associated with γ_{ij}, $\mathscr{R}_{jk} = \mathscr{R}^i_{jik}$ is the Ricci tensor, and $\mathscr{R} = \gamma^{ij}\mathscr{R}_{ij}$ is the Ricci scalar. The Geroch compactification of the orbit manifold $\hat{\mathscr{V}}$ is constructed by attaching a single point Λ to $\mathscr{V}$ and performing a suitable rescaling so that the rescaled potential

$$\hat{\Phi} := -\hat{\Phi}_M + i\hat{\Phi}_J = W^{-1/2}\Phi \tag{2.34}$$

and the rescaled metric

$$\hat{\gamma}_{ij} = W^2\gamma_{ij} \tag{2.35}$$

are a set of smooth fields on the conformal completion $\hat{\mathscr{V}} = \mathscr{V}\cup\Lambda$. Asymptotic flatness is enforced by requiring the conformal factor W to be C^2 at Λ and to satisfy the conditions

$$[W]_\Lambda = 0, \quad [W_{;i}]_\Lambda = 0, \quad [W_{;ij} - 2\hat{\gamma}_{ij}]_\Lambda = 0, \tag{2.36}$$

where the semicolon represents the action of the derivative operator $\hat{D}_i$ compatible with the conformal metric $\hat{\gamma}_{ij}$. The Geroch–Hansen multipole moments are then defined as a hierarchy of symmetric trace-free tensors at Λ,

$$Q_{i_1\cdots i_s} = [P_{i_1\cdots i_s}]_\Lambda, \tag{2.37}$$

$$P = \hat{\Phi}, \quad P_{i_1} = \hat{\Phi}_{;i_1},$$

$$P_{i_1\cdots i_{s+1}} = \mathscr{C}\left[P_{(i_1\cdots i_s;i_{s+1})} - \tfrac{1}{2}s(2s-1)P_{(i_1\cdots i_{s-1}}\hat{\mathscr{R}}_{i_si_{s+1})}\right],$$

where $\mathscr{C}[\cdots]$ denotes the trace-free part of a tensor.

In order to relate the asymptotic structure of a stationary AF vacuum metric at null infinity to the behavior near Λ, perhaps the most straightforward approach is to directly construct the Geroch compactification from the Penrose construction of $\mathscr{I}^+$. A simple relationship between the two completions is immediately suggested by an examination of the situation for the flat Minkowski space-time. With the standard conformal factor $\Omega = \rho = 1/r$, the line element

$$g_{\mu\nu}\,dx^\mu\,dx^\nu = du^2 + 2\,du\,dr - r^2\,d\Sigma^2$$

leads to the conformal metric

$$\hat{g}_{\mu\nu}\,dx^\mu\,dx^\nu = \rho^2\,du^2 - 2\,du\,d\rho - d\Sigma^2$$

on $\hat{\mathscr{M}} = \mathscr{M}\cup\mathscr{I}^+$ which is degenerate near the conformal boundary $\mathscr{I}^+$ consisting of the points $\rho = 0$. The natural induced metric $\hat{h}_{ij} = \Omega^2h_{ij}$ on the orbit manifold $\hat{\mathscr{M}}/\mathscr{G}$, i.e.,

$$\hat{h}_{ij}\,dy^i\,dy^j = \hat{g}_{\mu\nu}\,dx^\mu\,dx^\nu + \hat{\lambda}^{-1}(\hat{g}_{\mu\alpha}\xi^\alpha\,dx^\mu)^2$$
$$= \rho^{-2}\,d\rho^2 + d\Sigma^2$$

is obviously singular at $\rho = 0$. This is remedied by a further rescaling

$$\hat{\gamma}_{ij} = \hat{\lambda}\hat{h}_{ij} = \Omega^4\gamma_{ij},$$

which yields a new line element

$$\hat{\gamma}_{ij}\,dy^i\,dy^j = d\rho^2 + \rho^2\,d\Sigma^2$$

that is smooth on the entire orbit space $\hat{\mathscr{M}}/\mathscr{G}$ including the boundary set $\mathscr{I}^+/\mathscr{G}$ represented by $\{\rho = 0\}$. However, it is clear that as a result of the further rescaling, the boundary $\rho = 0$ must now be interpreted as a single point (rather than a two-sphere) and is to be identified with the point Λ in the Geroch compactification scheme.

One expects that the essential features of the above construction will be carried over to a large class of general AF stationary metrics. This suggests that for a suitably chosen conformal factor Ω producing a Penrose completion one may choose

$$W = \Omega^2 \tag{2.38}$$

to construct a Geroch one-point compactification of the orbit space. It remains, however, to investigate to what extent the various asymptotic conditions needed for a multipole expansion follow from the properties of the rescaled space-time metric near $\mathscr{I}^+$. The first two conditions of Eq. (2.36), namely, $[W]_\Lambda = 0$ and $[W_{;i}]_\Lambda = 0$ readily follow from Eq. (2.2). Thus W must be at least C^1 near Λ. Next we consider the quantity

$$w := W^{-1}\hat{\gamma}^{ij}W_{;i}W_{;j}. \tag{2.39}$$

Here $\mathscr{L}_\xi W = 0$ implies $W_{;\mu} = \gamma^i_\mu W_{;i}$, so that

$$w = 4\Omega^{-1}\lambda^{-1}v. \tag{2.40}$$

From Eqs. (2.8) and (2.11) we at once deduce that

$$[w]_\Lambda = 4. \tag{2.41}$$

The differentiability requirements on the rescaled fields $\hat{\Phi}_M$, $\hat{\Phi}_J$, and $\hat{\gamma}_{ij}$ at Λ are substantially stronger and cannot as such be derived from the conditions at null infinity in a straightforward manner. In intuitive terms, differentiability at Λ implicitly contains restrictions on the angular dependence of the physical fields on $\mathscr{I}^+$ and has to be introduced separately. However, it can be easily shown that the third condition in Eq. (2.36) is a consequence of the other asymptotic conditions provided the vacuum Einstein equations are imposed. We prove the following proposition.

Proposition: With the assumptions that

(i) $[W]_\Lambda = 0, \quad [W_{;i}]_\Lambda = 0,$

(ii) $[w]_\Lambda = 4,$

(iii) the rescaled fields $\hat{\Phi}_M$, $\hat{\Phi}_J$, and $\hat{\gamma}_{ij}$ are C^2 at Λ,

(iv) the stationary vacuum equations hold in an open neighborhood of Λ,

it follows that W is C^2 near Λ and moreover

$$[W_{;ij} - 2\hat{\gamma}_{ij}]_\Lambda = 0.$$

Proof: The conformal space field equations read

$$(\hat{\Delta} - 2\hat{\mathscr{R}})\hat{\Phi} = \tfrac{15}{2}W^{-1}[\hat{\Delta}W - \tfrac{3}{2}w]\hat{\Phi}, \tag{2.42}$$

$$\hat{\mathscr{R}}_{ij} = -W^{-1}[W_{;ij} + \hat{\gamma}_{ij}(\hat{\Delta}W - 2w) - \tfrac{1}{2}|\hat{\Phi}|^2W_{;i}W_{;j}]$$

$$+2[W\hat{\Phi}_{;i}\hat{\Phi}^*_{;j}+\tfrac{1}{2}W_{;(i}|\hat{\Phi}|^2{}_{;j)}$$
$$-(1+4W|\hat{\Phi}|^2)^{-1}(W|\hat{\Phi}|^2)_{;i}(W|\hat{\Phi}|^2)_{;j}]. \quad (2.43)$$

Since $\hat{\gamma}_{ij}$ is C^2 at Λ, $\hat{\mathscr{R}}_{ij}$ is continuous there, so that $[W\hat{\mathscr{R}}_{ij}]_\Lambda=0$. Multiplying Eq. (2.43) by W and evaluating at Λ, we conclude that

$$[W_{;ij}+\hat{\gamma}_{ij}(\hat{\Delta}W-8)]_\Lambda=0,$$

which is equivalent to the desired result. It is obvious that W must be at least C^2 in a neighborhood of Λ.

III. THE MULTIPOLE EXPANSION

In this section we display explicit formulas for the metric tensor of a stationary, AF space-time satisfying the Einstein equations in vacuum in the form of convergent series expansions in inverse powers of a suitable radial coordinate. This essentially implements the program outlined in Ref. 16.

Instead of Φ_M and Φ_J it is convenient to define the field variables

$$W=m^{-2}|\Phi|^2 \quad (3.1)$$

and

$$\alpha=\arg\Phi \quad (3.2)$$

on $\mathscr{V}$, where m is a nonzero constant. The key idea of Ref. 16 was to introduce W as a preferred conformal factor in order to construct $\hat{\mathscr{V}}$. The point Λ is attached to $\mathscr{V}$ so that W is extended to a C^2 function near Λ and $[W]_\Lambda=0$ and $[W_{;i}]_\Lambda=0$. The angular momentum monopole moment $l=[\hat{\Phi}_J]_\Lambda$ is assumed to vanish so that the constant m is identified with the mass of the system: $m=-[\hat{\Phi}_M]_\Lambda$. The basic fields can now be chosen to be W, α, and the rescaled metric $\hat{\gamma}_{ij}=W^2\gamma_{ij}$ in terms of which the field equations now read

$$\hat{\Delta}W-\tfrac{3}{2}w=-\tfrac{2}{15}(\hat{\gamma}^{mn}\alpha_{;m}\alpha_{;n}+2\hat{\mathscr{R}})W, \quad (3.3)$$
$$\hat{\Delta}\alpha=0, \quad (3.4)$$

and

$$\hat{\mathscr{R}}_{ij}=-W^{-1}[W_{;ij}+\hat{\gamma}_{ij}(\hat{\Delta}W-2w)-\tfrac{1}{2}m^2W_{;i}W_{;j}]$$
$$+2m^2[W\alpha_{;i}\alpha_{;j}-m^2(1+4m^2W)^{-1}W_{;i}W_{;j}], \quad (3.5)$$

where w is given by Eq. (2.39). For later convenience we introduce the multipole moments per unit mass:

$$a_{i_1\cdots i_n}=\mathrm{Re}(Q_{i_1\cdots i_n}/m),\quad b_{i_1\cdots i_n}=\mathrm{Im}(Q_{i_1\cdots i_n}/m). \quad (3.6)$$

With our choice of the conformal factor W, the first few are given by

$$a=1,\quad b=0, \quad (3.7)$$
$$a_i=0,\quad b_i=[\alpha_{;i}]_\Lambda, \quad (3.8)$$
$$a_{ij}=-\mathscr{C}[\alpha_{;i}\alpha_{;j}+\tfrac{1}{2}\hat{\mathscr{R}}_{ij}]_\Lambda,$$
$$b_{ij}=[\alpha_{;ij}]_\Lambda, \quad (3.9)$$
$$a_{ijk}=-\mathscr{C}[3\alpha_{;(i}\alpha_{;jk)}+\tfrac{1}{2}\hat{\mathscr{R}}_{(ij;k)}]_\Lambda,$$
$$b_{ijk}=\mathscr{C}[\alpha_{;(ijk)}-\alpha_{;i}\alpha_{;j}\alpha_{;k}-\tfrac{7}{2}\hat{\mathscr{R}}_{(ij}\alpha_{;k)}]_\Lambda. \quad (3.10)$$

The moments of the static Weyl solutions up to the octupole order are computed in Appendix B.

The following crucial theorem was proved in Ref. 16.

Theorem: Suppose that the following assumptions hold in a certain neighborhood of Λ_1:

(i) W, α, $\hat{\gamma}_{ij}$ represent a solution of the stationary vacuum Einstein equations, i.e., Eqs. (3.3)–(3.5) with nonzero mass ($m\neq 0$),
(ii) W is strictly positive except at Λ and $[W]_\Lambda=0$,
(iii) there is a chart in which W and α are C^2 and $\hat{\gamma}_{ij}$ is C^4.

Then there exists a chart in which W, α, and $\hat{\gamma}_{ij}$ are analytic at Λ, i.e., they admit a convergent Taylor expansion around Λ. In particular, the Riemann normal coordinates explicitly provide such a chart that, moreover, has the property that the coefficients of the Taylor expansion can be expressed in terms of the Geroch–Hansen multipole moments.

It is also possible to express the asymptotic conditions of the analyticity theorem in terms of an arbitrarily chosen conformal factor in the spirit of the original work of Geroch and Hansen. Let $\overline{W}$ be a conformal factor which is C^2 in a neighborhood of Λ and strictly positive there except at Λ and which vanishes at Λ. Moreover, suppose that the rescaled fields

$$\overline{\Phi}=\overline{W}^{-1/2}\Phi,\quad \overline{\gamma}_{ij}=\overline{W}^2\gamma_{ij} \quad (3.11)$$

can be extended to the point Λ in such a way that $\alpha=\arg\overline{\Phi}$ is C^2 and $|\overline{\Phi}|^2$ and $\overline{\gamma}_{ij}$ are C^4 near Λ. These conditions then at once imply the conditions on the special conformal factor W needed in the proof of the analyticity theorem.

In a normal coordinate system $\{y^i\}$ centered at Λ, W, α, and $\hat{\gamma}_{ij}$ have the following expansions:

$$W=\sum_{n=2}^{\infty}\frac{1}{n!}[W_{;i_1\cdots i_n}]_\Lambda y^{i_1}\cdots y^{i_n}, \quad (3.12)$$
$$\alpha=\sum_{n=1}^{\infty}\frac{1}{n!}[\alpha_{;i_1\cdots i_n}]_\Lambda y^{i_1}\cdots y^{i_n}, \quad (3.13)$$
$$\hat{\gamma}_{ij}=\delta_{ij}-\tfrac{1}{3}[\hat{\mathscr{R}}_{ikjl}]_\Lambda y^ky^l-\tfrac{1}{6}[\hat{\mathscr{R}}_{ikjl;m}]_\Lambda y^ky^ly^m+\cdots. \quad (3.14)$$

The quantity $q=(\delta_{ij}y^iy^j)^{1/2}$ plays the role of a preferred inverse radial coordinate in the physical space-time.

We now proceed to compute a few leading terms of these expansions. From the field equations one can easily derive the relations

$$W_{;ij}=-WK_{ij}+\tfrac{1}{2}m^2W_{;i}W_{;j}+\tfrac{1}{2}\hat{\gamma}_{ij}w, \quad (3.15)$$
$$w_{;k}=m^2wW_{;k}-2K_k{}^mW_{;m}, \quad (3.16)$$

where

$$K_{ij}:=L_{ij}+m^2\Theta_{ij}, \quad (3.17)$$
$$L_{ij}:=\hat{\mathscr{R}}_{ij}-\tfrac{1}{4}\hat{\gamma}_{ij}\hat{\mathscr{R}}, \quad (3.18)$$

and

$$\Theta_{ij}:=\tfrac{1}{2}\hat{\gamma}_{ij}[W\hat{\gamma}^{mn}\alpha_{;m}\alpha_{;n}+\tfrac{1}{4}w(1+4m^2W)^{-1}]$$
$$-2[W\alpha_{;i}\alpha_{;j}-m^2(1+4m^2W)^{-1}W_{;i}W_{;j}]. \quad (3.19)$$

Evaluating Eqs. (3.15), (3.16), and (3.19) at Λ and making use of the limits

$$[W]_\Lambda=0,\quad [W_{;i}]_\Lambda=0,\quad [w]_\Lambda=4,$$

$$[W_{;ij}]_\Lambda = 2\delta_{ij}, \tag{3.20}$$

$$[w_{;k}]_\Lambda = 0, \tag{3.21}$$

$$[\Theta_{ij}]_\Lambda = \tfrac{1}{2}\delta_{ij}. \tag{3.22}$$

Next, operating with $\hat{D}_k$ on Eq. (3.15) and using Eq. (3.21) we obtain

$$[W_{;ijk}]_\Lambda = 0. \tag{3.23}$$

By repeated covariant differentiation of Eq. (3.16) and evaluation of Λ, we obtain $[w_{;kl}]_\Lambda$, $[w_{;klm}]_\Lambda$, and so on:

$$[w_{;kl}]_\Lambda = -4[K_{kl}]_\Lambda + 8m^2\delta_{kl}, \tag{3.24}$$

$$[w_{;klm}]_\Lambda = -4[K_{kl;m} + K_{km;l}]_\Lambda. \tag{3.25}$$

Proceeding similarly from Eq. (3.15) and making use of Eqs. (3.24) and (3.25) we get

$$[W_{;ijkl}]_\Lambda = -2\{[K_{ij}]_\Lambda\delta_{kl} + [K_{kl}]_\Lambda\delta_{ij}\} + 2m^2(\delta_{il}\delta_{jk} + \delta_{jl}\delta_{ik} + 2\delta_{ij}\delta_{kl}), \tag{3.26}$$

$$[W_{;ijklm}]_\Lambda = -2\{[K_{ij;k}]_\Lambda\delta_{lm} + [K_{ij;l}]_\Lambda\delta_{km} + [K_{ij;m}]_\Lambda\delta_{kl} + \delta_{ij}[K_{kl;m} + K_{km;l}]_\Lambda\}. \tag{3.27}$$

To reduce Eq. (3.26) we need $[K_{ij}]_\Lambda$. From the field equations it follows that

$$\hat{\mathscr{R}} = -8\hat{\gamma}^{mn}\alpha_{;m}\alpha_{;n}(1 + \tfrac{15}{4}m^2W) - \tfrac{15}{2}m^2w(1 + 4m^2W)^{-1}, \tag{3.28}$$

and consequently

$$[\hat{\mathscr{R}}]_\Lambda = -8b^2 - 30m^2, \quad b^2 := \delta^{ij}b_ib_j\,. \tag{3.29}$$

Furthermore, from Eq. (3.9) we get

$$\mathscr{C}[\hat{\mathscr{R}}_{ij}]_\Lambda = -2(a_{ij} + b_ib_j - \tfrac{1}{3}\delta_{ij}b^2). \tag{3.30}$$

Combining Eqs. (3.17), (3.18), (3.22), (3.29), and (3.30) we get

$$[K_{ij}]_\Lambda = -2[a_{ij} + b_ib_j + m^2\delta_{ij}], \tag{3.31}$$

and finally from Eq. (3.26),

$$[W_{;(ijkl)}]_\Lambda = 8[a_{(ij}\delta_{kl)} + b_{(i}b_j\delta_{kl)} + 2m^2\delta_{(ij}\delta_{kl)}]. \tag{3.32}$$

In order to evaluate $[W_{;(ijklm)}]_\Lambda$ it suffices to compute the quantity

$$K_{(ij;k)} = L_{(ij;k)} + m^2\Theta_{(ij;k)} = \mathscr{C}[\hat{\mathscr{R}}_{(ij;k)}] + \tfrac{3}{20}\hat{\gamma}_{(ij}\hat{\mathscr{R}}_{;k)} + m^2\Theta_{(ij;k)} \tag{3.33}$$

at Λ. But a direct calculation shows that

$$[\Theta_{ij;k}]_\Lambda = 0, \tag{3.34}$$

$$[\hat{\mathscr{R}}_{;k}]_\Lambda = -16b^ib_{ik}, \tag{3.35}$$

and

$$\mathscr{C}[\hat{\mathscr{R}}_{(ij;k)}]_\Lambda = -6\mathscr{C}[b_{(i}b_{jk)}] - 2a_{ijk} = -6b_{(i}b_{jk)} + \tfrac{12}{5}\delta_{(ij}b_{k)l}b^l - 2a_{ijk}. \tag{3.36}$$

Assembling the various results we find that

$$[W_{;(ijklm)}]_\Lambda = 20[a_{(ijk}\delta_{lm)} + 3b_{(i}b_{jk}\delta_{lm)}], \tag{3.37}$$

and finally

$$W = q^2\{1 + \tfrac{1}{3}[2m^2q^2 + (a_{ij} + b_ib_j)y^iy^j] + \tfrac{1}{6}(a_{ijk} + 3b_ib_{jk})y^iy^jy^k + O(q^4)\}. \tag{3.38}$$

Next we proceed to obtain the Taylor expansion of α. The coefficients of the two leading terms are simply

$$[\alpha_{;i}]_\Lambda = b_i, \quad [\alpha_{;ij}]_\Lambda = b_{ij}. \tag{3.39}$$

In order to compute $[\alpha_{;(ijk)}]_\Lambda$ we proceed as follows. From Eq. (3.4) it follows that

$$\mathscr{C}[\alpha_{;(ijk)}] = \alpha_{;(ijk)} - \tfrac{2}{5}\hat{\gamma}_{(ij}\hat{\mathscr{R}}_{k)}{}^m\alpha_{;m}. \tag{3.40}$$

Also

$$\mathscr{C}[\hat{\mathscr{R}}_{(ij}\alpha_{;k)}] = \hat{\mathscr{R}}_{(ij}\alpha_{;k)} - \tfrac{1}{5}[\hat{\gamma}_{(ij}\alpha_{;k)}\hat{\mathscr{R}} + 2\hat{\gamma}_{(ij}\hat{\mathscr{R}}_{k)}{}^m\alpha_{;m}]. \tag{3.41}$$

A short computation then yields

$$[\alpha_{;(ijk)}]_\Lambda = b_{ijk} - 6b_ib_jb_k + 2\delta_{(ij}b_{k)}(b^2 - 2m^2) - 7a_{(ij}b_{k)} + 2\delta_{(ij}a_{k)m}b^m. \tag{3.42}$$

Hence, to the leading few orders, the expansion of α reads

$$\alpha = b_iy^i + \tfrac{1}{2}b_{ij}y^iy^j + \tfrac{1}{6}[b_{ijk} - 6b_ib_jb_k + 2\delta_{(ij}b_{k)}(b^2 - 2m^2) - 7a_{(ij}b_{k)} + 2\delta_{(ij}a_{k)m}b^m]y^iy^jy^k + \cdots. \tag{3.43}$$

Finally, it remains to obtain the expansion of $\hat{\gamma}_{ij}$. We recall that in three dimensions the Riemann tensor can be expressed in the form

$$\hat{\mathscr{R}}_{ikjl} = 4\hat{\gamma}_{[i|[j}L_{l]|k]}.$$

The second term of the expansion (3.14) is then easily evaluated making use of the limit

$$[L_{ij}]_\Lambda = -(2a_{ij} + 2b_ib_j + \tfrac{2}{3}m^2\delta_{ij}). \tag{3.44}$$

In order to evaluate the third term we need $[L_{ij;k}]_\Lambda$. Let us consider the antisymmetrized quantity (the Bach tensor)

$$L_{i[j;k]} = K_{i[j;k]} - m^2\Theta_{i[j;k]}.$$

An explicit calculation yields

$$WK_{i[j;k]} = m^2[\tfrac{1}{2}WK_{i[j}W_{;k]} + \tfrac{1}{4}\hat{\gamma}_{i[j}W_{;k]}w - \Theta_{i[j}W_{;k]} - \hat{\gamma}_{i[j}\Theta_{k]l}\hat{\gamma}^{lm}W_{;m}].$$

Substituting the expression for Θ_{ij} from Eq. (3.19) and canceling an overall factor W, we get

$$K_{i[j;k]} = m^2\{\tfrac{1}{2}K_{i[j}W_{;k]} + 2\alpha_{;i}\alpha_{;[j}W_{;k]} - \hat{\gamma}_{i[j}W_{;k]}[\hat{\gamma}^{mn}\alpha_{;m}\alpha_{;n} - 2w(1 + 4m^2W)^{-1}] - 2\hat{\gamma}_{i[j}\alpha_{;k]}\hat{\gamma}^{mn}\alpha_{;m}W_{;n}\}. \tag{3.45}$$

The right-hand side of Eq. (3.45) is manifestly regular at Λ due to a mutual cancellation among the potentially singular W^{-1} terms. Hence, by continuity, Eq. (3.45) holds at Λ. (In Ref. 16 this fact played a crucial role in the proof of the analyticity theorem). Evaluating at Λ, we get

$$[K_{i[j;k]}]_\Lambda = [L_{i[j;k]}]_\Lambda = 0$$

in view of Eq. (3.34). Hence we have the following relations at Λ:

$$L_{ij;k} = L_{ik;j} = L_{ki;j} = L_{kj;i} = L_{jk;i}.$$

These, in turn, imply that

$$[L_{ij;k}]_\Lambda = [L_{(ij;k)}]_\Lambda.$$

But, according to Eqs. (3.34)–(3.36),

$$[L_{(ij;k)}]_\Lambda = -2a_{ijk} - 6b_{(i}b_{jk)}.$$

The leading terms of the expansion (3.14) then take the form

$$\begin{aligned}\hat{\gamma}_{ij} = \delta_{ij} &+ \tfrac{1}{3}\{\delta_{ij}[2(a_{kl} + b_k b_l)y^k y^l + 5m^2q^2] \\ &- 5m^2 y_i y_j - 4y_{(i}[a_{j)k} + b_{j)}b_k]y^k \\ &+ 2q^2(a_{ij} + b_i b_j)\} + \tfrac{1}{3}\{\delta_{ij}[a_{klm} + 3b_k b_{lm}]y^k y^l y^m \\ &- 2y_{(i}[a_{j)km} + b_{j)}b_{km} + 2b_{j)m}b_k]y^k y^m \\ &+ q^2[a_{ijm} + 2b_{(i}b_{j)m} + b_{ij}b_m]y^m\} + \cdots .\end{aligned} \tag{3.46}$$

The physical space-time metric can be readily reconstructed from the basic fields. The line element can be written as

$$ds^2 = \lambda(dt + \sigma_i\, dy^i)^2 - \lambda^{-1}\gamma_{ij}\, dy^i\, dy^j, \tag{3.47}$$

where the three-vector σ_i is related to the twist scalar ω through the equation

$$D_i\omega = \epsilon_{ijk}\gamma^{jm}\gamma^{kn}D_{[m}\sigma_{n]}, \tag{3.48}$$

with ϵ_{ijk} being the alternating tensor with respect to the metric γ_{ij}. The expansions (3.38), (3.43), and (3.46) are simplified further in the axisymmetric case. In this case it is convenient to introduce spherical coordinates (q,Θ,Ψ) in the usual manner

$$y^1 = q\sin\Theta\cos\Psi,\quad y^2 = q\sin\Theta\sin\Psi,\quad y^3 = q\cos\Theta, \tag{3.49}$$

chosen so that the axial Killing vector (which commutes with the timelike Killing vector ξ^μ) is identified with the vector field $\partial/\partial\Psi = \zeta^\mu(\partial/\partial x^\mu)$. The curl of the induced vector field on $\hat{\mathscr{V}}$,

$$\zeta^i = \gamma^i_\mu\zeta^\mu, \tag{3.50}$$

defines the axis vector

$$z^i = \hat{\epsilon}^{ijk}\zeta_{[j;k]}, \tag{3.51}$$

with $\hat{\epsilon}^{ijk}$ being the totally antisymmetric tensor on $\hat{\mathscr{V}}$. The only nonvanishing independent components of the 2^l moments are the axial components,

$$\alpha_l = (l!)^{-1}a_{i_1\cdots i_l}[z^{i_1}\cdots z^{i_l}]_\Lambda, \tag{3.52a}$$

$$\beta_l = (l!)^{-1}b_{i_1\cdots i_l}[z^{i_1}\cdots z^{i_l}]_\Lambda. \tag{3.52b}$$

We then have the expansions

$$\begin{aligned}W = q^2\{1 &+ \tfrac{2}{3}[AP_2(\cos\Theta) + (m^2 + \tfrac{1}{6}\beta_1^2)]q^2 \\ &+ [BP_3(\cos\Theta) + \tfrac{2}{5}\beta_1\beta_2\cos\Theta]q^3 + \cdots\},\end{aligned} \tag{3.53}$$

$$\begin{aligned}\alpha = \beta_1 q\cos\Theta &+ \beta_2 q^2 P_2(\cos\Theta) \\ &+ \{[\beta_3 - \tfrac{1}{5}\beta_1(2\beta_1^2 + 7\alpha_2)]P_3(\cos\Theta) \\ &- \tfrac{4}{15}\beta_1(\alpha_2 + \beta_1^2 + \tfrac{5}{2}m^2)\cos\Theta\}q^3 + \cdots,\end{aligned} \tag{3.54}$$

and

$$\begin{aligned}d\hat{l}^2 := {}&\hat{\gamma}_{ij}\, dy^i\, dy^j \\ = {}& dq^2 + \{1 + \tfrac{1}{3}(2A + \tfrac{4}{5}\beta_1^2 + 5m^2)q^2 \\ &+ (B + \tfrac{16}{15}\beta_1\beta_2)\cos\Theta\, q^3 + O(q^4)\}q^2\, d\Theta^2 \\ &+ O(q^5)dq\, d\Theta + \{1 + \tfrac{1}{3}[4AP_2(\cos\Theta) \\ &- 2A + \tfrac{4}{5}\beta_1^2 + 5m^2]q^2 + [2BP_3(\cos\Theta) \\ &- (B - \tfrac{16}{15}\beta_1\beta_2)\cos\Theta]q^3 \\ &+ O(q^4)\}q^2\sin^2\Theta\, d\Psi^2 + O(q^6)d\Theta\, d\Psi,\end{aligned} \tag{3.55}$$

where, for brevity, we have set

$$A := \alpha_2 + \tfrac{1}{3}\beta_1^2, \tag{3.56}$$

$$B := \alpha_3 + \tfrac{2}{5}\beta_1\beta_2, \tag{3.57}$$

and $P_l(\cos\Theta)$ are the usual Legendre polynomials. The coefficient of dq^2 follows simply from the fact that the angular coordinates Θ, Ψ are constant along geodesics in $\hat{\mathscr{V}}$ emanating from Λ and q is the arclength along such a geodesic. This provides a useful algebraic check of the accuracy of the computation in every order of the series expansions. Some further algebra yields the following expansions for scalars λ and ω:

$$\begin{aligned}\lambda = 1 &- 2mq + 2m^2q^2 - \tfrac{2}{3}mq^3[(A - \beta_1^2)P_2(\cos\Theta) \\ &+ (m^2 - \tfrac{1}{3}\beta_1^2)] \\ &+ mq^4\{\tfrac{2}{3}m[2(A - 2\beta_1^2)P_2(\cos\Theta) - (m^2 + \tfrac{2}{3}\beta_1^2)] \\ &- (B - \tfrac{8}{5}\beta_1\beta_2)P_3(\cos\Theta) \\ &+ \tfrac{2}{5}\beta_1\beta_2\cos\Theta\} + \cdots,\end{aligned} \tag{3.58}$$

$$\begin{aligned}\omega = 2mq^2\{&\beta_1\cos\Theta + [\beta_2 P_2(\cos\Theta) - 2m\beta_1\cos\Theta]q \\ &+ [(\beta_3 - \tfrac{6}{5}A\beta_1)P_3(\cos\Theta) \\ &- 2m\beta_2 P_2(\cos\Theta) - \tfrac{1}{3}\beta_1(\tfrac{2}{5}\alpha_2 + \tfrac{4}{5}\beta_1^2 - 5m^2)\cos\Theta]q^2 \\ &+ \cdots\}.\end{aligned} \tag{3.59}$$

In the axisymmetric case, a special solution for the one-form $\sigma_i\, dy^i$ can be chosen to be of the form

$$\sigma_i\, dy^i = \tilde{\omega}\, d\Psi. \tag{3.60}$$

An explicit integration of Eq. (3.48) yields the solution

$$\tilde{\omega} = -2m\sin^2\Theta[\beta_1 q + (m\beta_1 + \tfrac{2}{3}\beta_2\cos\Theta)q^2 + \cdots]. \tag{3.61}$$

The physical space-time metric is then explicitly given by

$$ds^2 = \lambda(dt + \tilde{\omega}\, d\Psi)^2 - \lambda^{-1}W^{-2}\, d\hat{l}^2. \tag{3.62}$$

A partial check of the lengthy algebra involved in the preceding derivations is provided by the Schwarzschild metric, for which

$$\lambda = 1 - 2m/r,\quad \omega = 0,$$

$$dl^2 = dr^2 + (1 - 2m/r)r^2\, d\Sigma^2.$$

Setting $\rho = r^{-1}$, the preferred conformal factor $W = m^{-2}\Phi_M^2$ is given by

$$W = \rho^2[(1 - m\rho)/(1 - 2m\rho)]^2.$$

Taking into account the spherical symmetry, one can readily calculate the affine parameter along a radial geodesic emanating from the point Λ, i.e., $\{\rho = 0\}$,

$$q = \int_r^\infty W\,dr = \int_0^\rho d\rho \left(\frac{1-m\rho}{1-2m\rho}\right)^2$$
$$= \rho + m\rho^2 + \tfrac{5}{3}m^2\rho^3 + 3m^3\rho^4 + \tfrac{28}{5}m^4\rho^5 + \cdots .$$

Inverting the power series we get ρ in terms of q:

$$\rho = q - mq^2 + \tfrac{1}{3}m^2q^3 + \tfrac{1}{3}m^3q^4 + \tfrac{7}{5}m^4q^5 + \cdots .$$

The leading terms of the series for $\lambda = 1 - 2m\rho$ in powers of q are seen to agree with Eq. (3.58). The series expansion for W reads

$$W = q^2[\,1 + \tfrac{2}{3}m^2q^2 + O(q^4)\,],$$

which agrees with Eq. (3.53). Finally the rescaled spatial metric $d\hat{l}^2 = W^2\,dl^2$ is given by

$$d\hat{l}^2 = ((1-m\rho)/(1-2m\rho))^4[\,d\rho^2 + (1-2m\rho)\rho^2\,d\Sigma^2]$$
$$= dq^2 + [\,1 + \tfrac{2}{3}m^2q^2 + O(q^4)\,]q^2\,d\Sigma^2$$

in agreement with Eq. (3.55).

IV. MULTIPOLE STRUCTURE AT NULL INFINITY

The space-time metric obtained in the previous section represents a general AF stationary vacuum solution of Einstein's equation and can, in principle, be taken as the starting point for a rigorous discussion of the asymptotic structure of the gravitational field of a stationary isolated system in general relativity. The main disadvantage of this form of the solution is that the coordinate system in terms of which it is expressed does not fit in very well with the standard treatments of gravitational radiation theory. A special class of phenomena of some physical interest deals with initially stationary isolated systems which start to evolve, perhaps due to a sudden onset of time dependence in their equation of state, and emit gravitational waves. The Geroch–Hansen theory would then give a satisfactory gauge-invariant description of the multipole structure of the initial stationary configuration. The subsequent dynamical evolution of the field, on the other hand, is most conveniently described in the framework of a characteristic initial value formulation in a suitable null polar coordinate system. However, aside from a successful identification of the total energy of the system (the Bondi mass), very little is known at present about the detailed relationship between the various pieces of the characteristic data and the structure of the sources. Here we take a small step towards that difficult goal by expressing the leading terms of the general asymptotic solution of the stationary vacuum Einstein equation in a Bondi–Sachs null coordinate system in terms of its Hansen multipole moments.

In a Bondi-type coordinate system (u,r,x^A) based on a family of outgoing null hypersurfaces $u = \text{const}$ the space-time metric takes the form

$$ds^2 = g_{00}\,du^2 + 2(g_{01}\,dr + g_{0A}\,dx^A)du + g_{AB}\,dx^A\,dx^B,$$
$$x^0 := u, \quad x^1 := r, \tag{4.1}$$

where $A,B,\ldots$ take the values 2,3 and the summation convention is understood. The angular coordinates $x^2 = \theta$ and $x^3 = \phi$ label a particular null geodesic lying on this hypersurface. The contravariant metric components $g^{\mu\nu}$ are given by

$$g^{00} = 0, \quad g^{0A} = 0, \quad g^{01} = e^{-2\beta}, \quad g^{11} = -Ve^{-2\beta}/r,$$
$$g^{1A} = U^Ae^{-2\beta}, \quad g^{AB} = -r^{-2}h^{AB}, \tag{4.2}$$

where the particular forms of the nonzero components are chosen for algebraic convenience. One also defines

$$g_{AB} = -r^2h_{AB}, \tag{4.3}$$

so that

$$h_{AB}h^{BC} = \delta_A{}^C. \tag{4.4}$$

The radial coordinate r is a luminosity distance parameter defined so that $\det(h_{AB}) = \sin^2\theta$. It is also convenient to display explicitly the two dynamical degrees of freedom of the gravitational field by writing the angular metric in the form

$$h_{AB}\,dx^A\,dx^B = e^{2\gamma}\cosh 2\delta\,d\theta^2 + 2\sinh 2\delta\sin\theta\,d\theta\,d\phi$$
$$+ e^{-2\gamma}\cosh 2\delta\sin^2\theta\,d\phi^2. \tag{4.5}$$

The asymptotic solutions of the vacuum Einstein equation for this form of the line element were obtained by van der Burg[25] who extended the pioneering work of Bondi *et al.* and Sachs. In a subsequent paper[26] he investigated the detailed structure of the stationary metric. In the stationary case, it can be shown that[27] the hypersurfaces $u = \text{const}$ can be so chosen that $\partial/\partial u$ coincides with the timelike Killing vector field and consequently $\partial g_{\mu\nu}/\partial u = 0$. The news tensor then vanishes at future null infinity. It is further known that by a suitable Bondi–Metzner–Sachs (BMS) supertranslation the complex asymptotic shear of the outgoing null hypersurface $u = \text{const}$ can be made to vanish. In the axisymmetric case the asymptotic expansions of γ and δ in powers of r^{-1} then take the form

$$\gamma + i\delta = G(\theta)r^{-3} + H(\theta)r^{-4} + \cdots . \tag{4.6}$$

The hypersurface equations $G_{1\mu} = 0$ (where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ is the Einstein tensor) determine, in a familiar manner, the asymptotic expansions

$$\beta = -\tfrac{3}{4}(G'^2 + G''^2)r^{-6} + \cdots, \tag{4.7}$$
$$U := U^2 + iU^3\sin\theta = 2N(\theta)r^{-3}$$
$$+ \tfrac{3}{2}\delta_2G(\theta)r^{-4} + \tfrac{8}{5}\delta_2H(\theta)r^{-5} + \cdots, \tag{4.8}$$
$$V = r - 2M(\theta) - \delta_1N'(\theta)r^{-1} - \tfrac{1}{2}\delta_1\delta_2G'(\theta)r^{-2}$$
$$- [\tfrac{1}{3}\delta_1\delta_2H'(\theta) + 3(N'^2 + N''^2)]r^{-3} + \cdots, \tag{4.9}$$

where we introduce the notations

$$N(\theta) := N'(\theta) + iN''(\theta),$$
$$G(\theta) := G'(\theta) + iG''(\theta), \tag{4.10}$$
$$H(\theta) := H'(\theta) + iH''(\theta),$$

and δ_S stands for the operator $\partial/\partial\theta + s\cot\theta$. The angular dependence of the functions $M(\theta),N(\theta),G(\theta),H(\theta)$, etc., then follow from the supplementary conditions $G_{00} = 0$, $G_{0A} = 0$ and the dynamical equations

$$G_{AB} - \tfrac{1}{2}g_{AB}g^{CD}G_{CD} = 0$$

with all the u derivatives set equal to zero:

$$M(\theta) = m = \text{const},$$
$$N(\theta) = N_0 \sin\theta, \quad N_0 = p + iJ = \text{const},$$
$$G(\theta) = G_0 \sin^2\theta, \quad G_0 := G_0' + iG_0'' = \text{const},$$
$$H(\theta) = \tfrac{1}{3}H_0 P_3^2(\cos\theta) + \tfrac{5}{2}(mG_0 - \tfrac{1}{2}N_0^2)\sin^2\theta,$$
$$H_0 := H_0' + iH_0'' = \text{const}. \qquad (4.11)$$

The series expansions of the functions U and V then become

$$U = 2N_0 \sin\theta/r^3 - 2G_0 P_2{}^1(\cos\theta)/r^4 - \tfrac{8}{3}[H_0 P_3{}^1(\cos\theta) + (mG_0 - \tfrac{1}{2}N_0^2)P_2{}^1(\cos\theta)]/r^5 + \cdots, \qquad (4.12)$$

$$V = r - 2m - 2p\cos\theta/r - 4G_0' P_2(\cos\theta)/r^2 - 2[4H_0' P_3(\cos\theta) + 2(mG_0' + J^2)P_2(\cos\theta) - (p^2 + J^2)]/r^3 - \cdots, \qquad (4.13)$$

where $P_l^m(\cos\theta)$ denote the associated Legendre functions. From Eqs. (4.12) and (4.13) one can easily derive power series expansions for λ and ω. For λ one finds

$$\lambda = g_{00} = Ve^{2\beta}/r - r^2 h_{AB} U^A U^B$$
$$= 1 - 2m/r - 2p\cos\theta/r^2 - 4G_0' P_2(\cos\theta)/r^3 - 4[2H_0' P_3(\cos\theta) + (mG_0' + J^2)P_2(\cos\theta) - \tfrac{1}{2}(p^2 + J^2)\cos 2\theta]/r^4 - \cdots. \qquad (4.14)$$

The (path independent) line integral $\omega = \int \omega_\mu \, dx^\mu$ is most easily evaluated by integrating along a null ray $u,\theta,\phi = \text{const}$,

$$\omega = -\int_r^\infty \omega_1 \, dr = -2\int_r^\infty dr\sqrt{-g}\, g^{2\alpha} g^{3\beta} g_{0[\alpha,\beta]}, \qquad (4.15)$$

where $g := \det(g_{\mu\nu})$ and the comma denotes ordinary derivative. A straightforward term by term integration then yields

$$\omega = 2J\cos\theta/r^2 + 4G_0'' P_2(\cos\theta)/r^3 + [8H_0'' P_3(\cos\theta) + 4(mG_0'' - pJ)P_2(\cos\theta)]/r^4 + \cdots. \qquad (4.16)$$

At this point it is convenient to simplify the subsequent discussions by setting $p = 0$ by means of a suitable BMS translation. Introducing a new variable $\rho = r^{-1}$ we obtain a chart (ρ,θ,ϕ) near Λ which is the point $\rho = 0$. In this chart W and α are given by

$$W = \rho^2 + 2m\rho^3 + [(4G_0'/m + \tfrac{2}{3}J^2/m^2)P_2(\cos\theta) + (5m^2 + \tfrac{1}{3}J^2/m^2)]\rho^4 + \cdots, \qquad (4.17)$$

$$\alpha = (J/m)\rho\cos\theta + [(2G_0''/m)P_2(\cos\theta) + J\cos\theta]\rho^2 + \{[(4H_0''/m) - (6JG_0'/5m^2) - (2J^3/15m^3)] \times P_3(\cos\theta) + 4G_0'' P_2(\cos\theta) + [mJ - (4JG_0'/5m^2) - (J^3/5m^3)]\cos\theta\}\rho^3 + \cdots. \qquad (4.18)$$

The rescaled metric on $\hat{\mathscr{V}}$ has the approximate form

$$d\hat{l}^2 \cong \rho^{-4} W^2 [d\rho^2 - 2U^A h_{AB}\, d\rho\, dx^B + \rho^3 V h_{AB}\, dx^A\, dx^B]. \qquad (4.19)$$

Unfortunately, the lack of smoothness of the Bondi chart at Λ precludes a direct computation of the multipole moments in this chart. However, a coordinate transformation

$$\rho = x(1 - mx) \quad (0 \leqslant x < m^{-1}), \qquad (4.20)$$

with

$$X = X^1 := x\sin\theta\cos\phi, \quad Y = X^2 := x\sin\theta\sin\phi,$$
$$Z = X^3 := x\cos\theta, \qquad (4.21)$$

improves the situation somewhat by making α, C^3 and $\hat{\gamma}_{ij}$, C^2 at Λ, i.e., $\{X^i = 0\}$. In this new quasi-Cartesian chart one finds

$$\alpha = (J/m)Z(1 - m^2x^2) + (2H_0''/m)(5Z^2 - 3x^2)Z + (3Z^2 - x^2)[(G_0''/m) - (JG_0'/m^2)Z] - (J^3/3m^3)Z^3 + O(x^4), \qquad (4.22)$$

$$\hat{\gamma}_{ij} = [1 + (12G_0'/m + 2J^2/m^2)Z^2 + (m^2 - 4G_0'/m)x^2]\delta_{ij} - 3m^2 X_i X_j + 4JX^k X_{(i}\epsilon_{j)k3} + O(x^3). \qquad (4.23)$$

By virtue of Eqs. (3.8) and (3.9) the first and second derivatives of α, upon evaluation at Λ, lead to the simple relations

$$J = m\beta_1, \qquad (4.24)$$
$$G_0'' = \tfrac{1}{2}m\beta_2, \qquad (4.25)$$

thus providing the physical interpretation of these constants in terms of the multipole moments of the metric. In order to identify the mass quadrupole moment, one needs to consider the Ricci tensor $\hat{\mathscr{R}}_{kl}$ at Λ. Since the components of $\hat{\gamma}_{ij}$ differ from those of a flat Euclidean three-metric only by terms of quadratic and higher orders, it follows at once that the Christoffel symbols vanish at Λ and, consequently, the Riemann tensor at Λ is given by the linearized expression

$$[\hat{\mathscr{R}}_{ijkl}]_\Lambda = \tfrac{1}{2}[\hat{\gamma}_{il,kj} - \hat{\gamma}_{ij,kl} + \hat{\gamma}_{kj,il} - \hat{\gamma}_{kl,ij}]_\Lambda. \qquad (4.26)$$

The Ricci tensor of the metric (4.23) is then easily evaluated at Λ:

$$[\hat{\mathscr{R}}_{kl}]_\Lambda = -2[(5m^2 + \beta_1{}^2 - 2G_0'/m)\delta_{kl} + (\beta_1{}^2 + 6G_0'/m)\delta_{k3}\delta_{l3}]. \qquad (4.27)$$

From Eq. (4.27) we see that the Ricci scalar at Λ has the value $-(30m^2 + 8\beta_1{}^2)$, in agreement with Eq. (3.29). Equation (3.9) then leads to the identification

$$G_0' = \tfrac{1}{2}m\alpha_2, \qquad (4.28)$$

generalizing a result of Bondi *et al.* obtained from an analysis of the static Weyl solutions. Furthermore, a straightforward calculation from Eqs. (3.42), (4.22), and (4.23) shows that

$$H_0'' = \tfrac{1}{4}m\beta_3. \qquad (4.29)$$

In order to determine the higher-order moments, one can, in principle, continue the procedure and introduce a sequence of progressively smoother charts in a neighborhood of Λ by means of successive judiciously chosen coordinate transformations. Fortunately, such an awkward approach is not necessary. It is possible to find, once and for all, a coordinate transformation connecting the two forms of a stationary AF vacuum metric given by Eqs. (3.47) and (4.1). Without going into the details, here we simply quote the final result for the axisymmetric case. In a neighborhood of Λ, the transformation formulas relating the (t,q,Θ,Ψ) coordinates to an asymptotically shear-free Bondi coordinate system (u,ρ,θ,ϕ) read

$$u = t - q^{-1} + 2m \ln q - \tfrac{1}{3}[2AP_2(\cos\Theta) - 4m^2 + \tfrac{1}{3}\beta_1{}^2]q - \{\tfrac{1}{2}BP_3(\cos\Theta) + \tfrac{1}{3}m[(A+\beta_1{}^2)P_2(\cos\Theta) + m^2 + \tfrac{2}{3}\beta_1{}^2] + \tfrac{1}{3}\beta_1\beta_2\cos\Theta\}q^2 + \cdots, \tag{4.30a}$$

$$\rho = q - mq^2 - \tfrac{1}{3}[2AP_2(\cos\Theta) - m^2 + \tfrac{1}{3}\beta_1{}^2]q^3 - \{\tfrac{1}{2}BP_3(\cos\Theta) - \tfrac{1}{3}m[4AP_2(\cos\Theta) + m^2 + \tfrac{2}{3}\beta_1{}^2] + \tfrac{1}{3}\beta_1\beta_2\cos\Theta\}q^4 + \cdots, \tag{4.30b}$$

$$\theta = \Theta + \tfrac{1}{3}AP_2{}^1(\cos\Theta)q^2 + \tfrac{1}{3}[BP_3{}^1(\cos\Theta) - \tfrac{2}{3}m(A-\beta_1{}^2)P_2{}^1(\cos\Theta) - \tfrac{2}{3}\beta_1\beta_2\sin\Theta]q^3 + \cdots, \tag{4.30c}$$

$$\phi = \Psi + m\beta_1 q^2 + (m\beta_2\cos\Theta - \tfrac{2}{3}m^2\beta_1)q^3 + \cdots. \tag{4.30d}$$

A derivation of these formulas is outlined in Appendix A. They provide the necessary link between the two formulations of the asymptotic solutions of the stationary vacuum Einstein equations. With further effort more terms of these expansions could be calculated if the computation of the solutions given in Sec. III are also appropriately extended to higher orders. Explicit use of these expansions confirms Eqs. (4.24), (4.28), and (4.29) and also yields the simple result

$$H_0' = \tfrac{1}{4}m\alpha_3. \tag{4.31}$$

It is not yet quite clear whether the striking simplicity of these relations found here, in fact, extends to all orders of the power series expansion or is merely an accidental feature of the low orders. It is, however, tempting to speculate that this correspondence between the moments naively read off from the Bondi expansions and the gauge-invariant Geroch–Hansen moments is perhaps a general one and points to some deep geometric property of the Bondi coordinate system.

V. CONCLUDING REMARKS

The results of the preceding sections show that all stationary AF vacuum solutions encompassed by the analyticity theorem also admit a Taylor expansion in powers of r^{-1} in a Bondi chart. It may be pointed out that the issue of convergence of these formal expansions has not been addressed in this paper and remains, for the time being, an open problem. However, the analyticity theorem guarantees convergence of the series expansions of the metric variables W, α, and $\hat{\gamma}_{ij}$ in the (t,q,Θ,Ψ) chart, and so the convergence question centers around the transformation formulas Eq. (4.30). It seems that the main difficulty lies with the formal solution to the nonlinear equation (A7) given by the series (4.30a). If this problem could be overcome, convergence of the remaining series solutions of the linear equations (A8) and (A9) given by Eqs. (4.30b)–(4.30d) is expected to follow in a relatively simple manner.

It is natural to ask whether our results can help to formulate a useful definition of the multipole moments at null infinity in the time-dependent case. The usual asymptotic fall-off conditions discussed extensively in the literature, which are imposed either on the components of the metric tensor,[2,3] or on the tetrad components of the Weyl tensor at future null infinity,[28] do not exclude incoming radiation completely. Consequently, the asymptotic solution of the vacuum Einstein equations constructed from the time evolution of a sufficiently general class of null hypersurface data is expected to possess two distinct sets of moments—one at future null infinity $\mathscr{I}^+$ and the other at past null infinity $\mathscr{I}^-$. However, the prospect of obtaining results analogous to the usual multipole expansions of the scalar and electromagnetic waves appears to be rather dim owing to the difficulty of achieving a clear separation between the incoming and the outgoing radiation in general relativity. Even if a suitable tentative definition of the moments is found, for example along the lines suggested by Janis and Newman,[29] it is questionable whether a rigorous theorem can be established to the effect that broad classes of AF time-dependent vacuum solutions, like their stationary counterparts, are uniquely characterized by their moments. Also the role of the BMS asymptotic symmetry group[7] in the present context deserves closer study. A better understanding of the multipole expansion of electromagnetism in geometric terms may shed further light into these problems.

Another noteworthy point is the existence of a hidden symmetry of the vacuum field equations in the stationary case, namely the Ehlers duality

$$\Phi \to \Phi e^{i\epsilon}, \quad \gamma_{ij} \to \gamma_{ij},$$

which, like in electromagnetism, allows one to separate the moments into electric and magnetic types in a symmetrical manner. It is not clear whether a similar symmetrical treatment would be possible in the time-dependent case, since a corresponding generalization of the duality invariance is not available and the analogy with the Maxwell equations breaks down. Some of these questions are currently under investigation.

ACKNOWLEDGMENT

The author is grateful to the Ohio University Research Committee for a grant supporting the work reported in this paper.

APPENDIX A: A COORDINATE TRANSFORMATION

In this appendix we outline a derivation of the coordinate transformation (4.30) relating the form of the space-time metric given by Eq. (3.47) to the usual Bondi form (4.1).

A Bondi-type null polar coordinate system (u,r,x^A) based on a family of outgoing null hypersurface $u = \text{const}$ is characterized by the coordinate conditions,

$$g^{\mu\nu}u_{,\mu}u_{,\nu} = 0, \tag{A1}$$

$$g^{\mu\nu}u_{,\mu}x^A{}_{,\nu} = 0, \tag{A2}$$

where the comma denotes ordinary differentiation. The vector field $l_\mu = u_{,\mu}$ normal to a hypersurface $u = u_0$ is also tangent to the null geodesics lying within the hypersurface, which, in turn, are labeled by constant values of the angular coordinates x^A. The luminosity distance r parametrizing the individual null geodesics satisfies the equation

$$\nabla_\mu (\rho^2 g^{\mu\nu} u_{,\nu}) = 0, \qquad (A3)$$

where $\rho = r^{-1}$. The future null infinity is given by $\rho = 0$. It is to be noted that, because of Eq. (A2), Eq. (A3) determines r only up to a multiplicative factor $f(x^A)$. In the (t,y^i) chart the space-time metric has the components

$$g_{tt} = \lambda, \quad g_{ti} = \lambda\sigma_i, \quad g_{ij} = -\lambda^{-1}\gamma_{ij} + \lambda\sigma_i\sigma_j. \qquad (A4)$$

The contravariant metric is then given by

$$g^{tt} = \lambda^{-1} - \lambda\gamma^{ij}\sigma_i\sigma_j, \quad g^{ti} = \lambda\gamma^{ij}\sigma_j, \quad g^{ij} = -\lambda\gamma^{ij}, \qquad (A5)$$

where $\gamma^{ij}\gamma_{jk} = \delta^i_k$. Also $\sqrt{-g} = \lambda^{-1}[\det(\gamma_{ij})]^{1/2}$.

If we make the ansatz

$$u = t - F(y^i), \qquad (A6)$$

then Eq. (A1) becomes a differential equation for the function F:

$$\lambda^2\gamma^{mn}(F_{,m} + \sigma_m)(F_{,n} + \sigma_n) = 1. \qquad (A7)$$

Equations (A2) and (A3) then take the form

$$\gamma^{mn}(F_{,m} + \sigma_m)x^A_{,n} = 0, \qquad (A8)$$

$$\{\rho^2[\det(\gamma_{ij})]^{1/2}\gamma^{mn}(F_{,n} + \sigma_n)\}_{,m} = 0. \qquad (A9)$$

We now seek a solution of Eq. (A7) in the form

$$F(q,\Theta) = q^{-1} + \widetilde{F}(\Theta)\ln q + \sum_{n=0}^{\infty} F_n(\Theta)q^n. \qquad (A10)$$

Substituting into Eq. (A7) one sees that, in order to avoid logarithmic terms, it is necessary to assume that $\widetilde{F}(\Theta) = \text{const}$. The $O(q^{-1})$ terms then imply $\widetilde{F}(\Theta) = -2m$. The $O(q^0)$ terms yield a relation between $F_1(\Theta)$ and $F_0(\Theta)$:

$$F_1(\Theta) = \tfrac{1}{2}(F_0')^2 + \tfrac{1}{3}[2AP_2(\cos\Theta) - 4m^2 + \tfrac{1}{3}\beta_1^{\,2}], \qquad (A11)$$

where the prime denotes $d/d\Theta$. Similarly, the $O(q)$ terms determine $F_2(\Theta)$ in terms of $F_1(\Theta)$ and $F_0(\Theta)$, and so on:

$$F_2(\Theta) = \tfrac{1}{2}F_0'F_1' - mF_1 + m[(A + \tfrac{1}{3}\beta_1^{\,2})P_2(\cos\Theta) - m^2 + \tfrac{1}{3}\beta_1^{\,2}] + \tfrac{1}{2}BP_3(\cos\Theta) + \tfrac{1}{3}\beta_1\beta_2\cos\Theta. \qquad (A12)$$

However, the function $F_0(\Theta)$ remains completely undetermined at this stage. This arbitrariness simply reflects the BMS supertranslation freedom in the choice of a Bondi chart. The appropriate choice of F_0 is dictated by a consideration of the asymptotic behavior of the metric tensor. Anticipating the final result, we set $F_0 = 0$. It will be shown later that this choice indeed makes the null hypersurfaces $u = \text{const}$ asymptotically shear-free.

Next we consider Eq. (A9) for $\rho(q,\Theta)$. In order to fix a possible angle dependent factor, we make the following ansatz for the form of $\rho(q,\Theta)$:

$$\rho^2 = q^2\left[1 + \sum_{n=1}^{\infty} R_n(\Theta)q^n\right]. \qquad (A13)$$

Inserting it to Eq. (A9) we obtain, after some algebra,

$$R_1 = -2m,$$
$$R_2(\Theta) = \tfrac{1}{3}[5m^2 - \tfrac{2}{3}\beta_1^{\,2} - 4AP_2(\cos\Theta)],$$
$$R_3(\Theta) = -BP_3(\cos\Theta) + 4mAP_2(\cos\Theta) - \tfrac{2}{3}\beta_1\beta_2\cos\Theta + \tfrac{2}{3}m\beta_1^{\,2},$$

and Eq. (4.30b) follows.

In order to solve Eq. (A8) for $A = 2$, it is convenient to assume an expansion of the form

$$\cos\theta = \cos\Theta\left[1 + \sum_{n=1}^{\infty} C_n(\Theta)\right]q^n. \qquad (A14)$$

Substituting and collecting like powers of q, we obtain the relations

$$C_1(\Theta) = 0, \quad C_2(\Theta) = A\sin^2\Theta,$$
$$C_3(\Theta) = \tfrac{1}{3}[-BP_3^{\,1}(\cos\Theta) + \tfrac{2}{3}m(A - \beta_1^{\,2}) \times P_2^{\,1}(\cos\Theta) + \tfrac{2}{3}\beta_1\beta_2\sin\Theta]\tan\Theta,$$

and so on. This is equivalent to Eq. (4.30c). Equation (4.30d) follows similarly. These transformation formulas determine the asymptotic behavior of a stationary vacuum metric at $\mathscr{I}^+$ in a Bondi coordinate system. A detailed examination of the asymptotic behavior of the functions γ and δ shows that $\delta = O(\rho^3)$ as expected and moreover, when $F_0 = 0$, also $\gamma = O(\rho^3)$ near $\rho = 0$. Absence of a term linear in ρ in γ and δ shows that the null hypersurfaces $u = \text{const}$ are indeed asymptotically shear-free. Furthermore, as is well known, vanishing of the coefficient of ρ^2 prevents logarithmic terms from appearing in the metric components. Its interpretation as an outgoing radiation condition has been discussed by Sachs.[3]

APPENDIX B: MULTIPOLE MOMENTS OF THE WEYL METRIC

In this appendix we compute the leading Hansen multipole moments of the static, axisymmetric Weyl metrics.

A general stationary, axisymmetric space-time admits two commuting Killing vectors $\xi^\mu\,\partial_\mu$ and $\zeta^\mu\,\partial_\mu$ whose orbits form a family of timelike two-surfaces of topology $S^1\times\mathbb{R}$ near spatial infinity. Moreover, by a theorem of Papapetrou,[30] these orbits are two-surface orthogonal if there is at least one point on the symmetry axis where the vacuum field equations $R_{\mu\nu} = 0$ are satisfied. This is always true in a physically realistic situation dealing with spatially bounded sources, since at least a portion of the symmetry axis lies in the exterior region. The space-time metric, in this case, can be expressed in the canonical form

$$ds^2 = \lambda(dt + \widetilde{\omega}\,d\overline{\Psi})^2 - \lambda^{-1} \times [e^{2\chi}(dR^2 + R^2\,d\overline{\Theta}^2) + R^2\sin^2\overline{\Theta}\,d\overline{\Psi}^2]. \qquad (B1)$$

Here $(R,\overline{\Theta},\overline{\Psi})$ is the analog of the spherical polar coordinates in the flat space. The two independent Killing vectors are $\partial/\partial t = \xi^\mu\,\partial_\mu$ and $\partial/\partial\overline{\Psi} = \zeta^\mu\,\partial_\mu$. The metric on the space of orbits of ξ^μ, namely $\mathscr{V}$, is simply

$$dl^2 = e^{2\chi}(dR^2 + R^2\,d\overline{\Theta}^2) + R^2\sin^2\overline{\Theta}\,d\overline{\Psi}^2. \qquad (B2)$$

From Eq. (B2) we note that the rotational Killing field $\zeta^i = \gamma^i_\mu\zeta^\mu$ induced on $\mathscr{V}$ by ζ^μ is two-surface orthogonal. In order to obtain a local coordinate system in a neighborhood of Λ, we set $\overline{R} = R^{-1}$, Λ being the point $\overline{R} = 0$. The rescaled metric on $\hat{\mathscr{V}} = \mathscr{V} \cup \Lambda$ is given by

$$d\hat{l}^2 = W^2\, dl^2 = W^2 \overline{R}^{-4}[e^{2\chi}(d\overline{R}^2 + \overline{R}^2\, d\overline{\Theta}^2) + \overline{R}^2 \sin^2 \overline{\Theta}\, d\overline{\Psi}^2]. \quad \text{(B3)}$$

Some elementary considerations show that the transformation relating the normal coordinates (q,Θ,Ψ) to the canonical coordinates $(\overline{R},\overline{\Theta},\overline{\Psi})$ must be of the form

$$q = q(\overline{R},\overline{\Theta}), \quad \Theta = \Theta(\overline{R},\overline{\Theta}), \quad \Psi = \overline{\Psi}. \quad \text{(B4)}$$

Let us introduce a set of quasi-Cartesian coordinates $\{\overline{X}^i\} = \{\bar{x},\bar{y},\bar{z}\}$ at Λ in the usual way,

$$\overline{X}^1 = \bar{x} := \overline{R} \sin \overline{\Theta} \cos \overline{\Psi}, \quad \overline{X}^2 = \bar{y} := \overline{R} \sin \overline{\Theta} \sin \overline{\Psi},$$
$$\overline{X}^3 = \bar{z} := \overline{R} \cos \overline{\Theta}, \quad \text{(B5)}$$

and consider a radial geodesic $\overline{X}^i(q)$ emanating from Λ in $\hat{\mathscr{V}}$. Since $\partial/\partial\overline{\Psi} = \zeta^i \partial/\partial\overline{X}^i$ is a Killing field on $\hat{\mathscr{V}}$, it follows at once that

$$\hat{\gamma}_{ij}\zeta^i \frac{d\overline{X}^j}{dq} = C, \quad \text{(B6)}$$

a constant along the geodesic. For the class of AF metrics of present interest, it is reasonable to expect that the constant C will vanish. Going back to the polar coordinates, Eq. (B6) takes the simple form $d\overline{\Psi}/dq = 0$, implying that $\overline{\Psi} = \text{const}$ along a radial geodesic emanating from Λ. Hence the relation between Ψ and $\overline{\Psi}$ must be of the form $\overline{\Psi} = \Psi + h(\Theta)$. Moreover, upon comparing the two forms of the metric $d\hat{l}^2$ given by Eqs. (3.55) and (B3), it becomes clear that one must have $h(\Theta) = 0$, i.e., $\overline{\Psi} = \Psi$. Thus as found in Sec. III no $d\Theta\, d\Psi$ term can appear in Eq. (3.55) in any order of expansion in powers of q.

After these general observations, let us specialize to the case of the static Weyl solutions. We now have

$$\lambda := e^{2\psi}, \quad \omega = 0, \quad W = (\tfrac{1}{2}m)^2 \sinh^2 2\psi, \quad \text{(B7)}$$

with ψ being an axisymmetric solution of the Laplace equation in a three-dimensional Euclidean space in which $R,\overline{\Theta},\Psi$ are the usual spherical polar coordinates. In order to write down the field equations in the conformal space, we set $\mu = R\psi$. Then the rescaled metric in a neighborhood of Λ becomes

$$d\hat{l}^2 = (\mu/m)^4[\sinh(2\overline{R}\mu)/2\overline{R}\mu]^4 \times [e^{2\chi}(d\bar{\rho}^2 + d\bar{z}^2) + \bar{\rho}^2\, d\Psi^2], \quad \text{(B8)}$$

where $\bar{\rho}^2 = \bar{x}^2 + \bar{y}^2$. The potential μ satisfies the Laplace equation,

$$\mu_{,\bar{\rho}\bar{\rho}} + \bar{\rho}^{-1}\mu_{,\bar{\rho}} + \mu_{,\bar{z}\bar{z}} = 0. \quad \text{(B9)}$$

The field equations that determine χ can be put in the form

$$\chi_{,\bar{x}} = -\bar{x}(f^2 - g^2),$$
$$\chi_{,\bar{y}} = -\bar{y}(f^2 - g^2), \quad \text{(B10)}$$
$$\chi_{,\bar{z}} = -2\bar{\rho}fg,$$

where

$$f := \mu + \bar{\rho}\mu_{,\bar{\rho}} + \bar{z}\mu_{,\bar{z}}, \quad g := \bar{\rho}\mu_{,\bar{z}} - \bar{z}\mu_{,\bar{\rho}}. \quad \text{(B11)}$$

For the class of axisymmetric solutions of Eq. (B9) of the form

$$\mu = m \sum_{l=0}^{\infty} \bar{\alpha}_l \overline{R}^{\,l} P_l(\cos \overline{\Theta}), \quad \bar{\alpha}_0 = 1, \quad \text{(B12)}$$

which are analytic functions of $\bar{x}, \bar{y}, \bar{z}$ near Λ, the quantities

$$m^{-2}(f^2 - g^2) = 1 + 4\bar{\alpha}_1\bar{z} + [(4\bar{\alpha}_1^2 + 6\bar{\alpha}_2)\bar{z}^2 - (\bar{\alpha}_1^2 + 5\bar{\alpha}_2)(\bar{x}^2 + \bar{y}^2)] + \cdots,$$
$$2m^{-2}\bar{\rho}fg = 2(\bar{x}^2 + \bar{y}^2)[\bar{\alpha}_1 + 2(\bar{\alpha}_1^2 + \bar{\alpha}_2)\bar{z} + \cdots],$$

and hence χ are also analytic functions of $\bar{x}, \bar{y}, \bar{z}$. Thus the metric given by Eq. (B8) is analytic near Λ in the $(\bar{x},\bar{y},\bar{z})$ chart. Consequently, the computation of the multipole moments can be performed in this chart without any further difficulty by directly examining the Ricci tensor $\hat{\mathscr{R}}_{ij}$ and its covariant derivatives at Λ. A straightforward but tedious computation yields the relations

$$\alpha_2 = 2\bar{\alpha}_2 - \tfrac{1}{2}\bar{\alpha}_1^2 + \tfrac{1}{3}m^2,$$
$$\alpha_3 = 2\bar{\alpha}_3 + \bar{\alpha}_1(\bar{\alpha}_1^2 + \tfrac{4}{3}m^2 - 4\bar{\alpha}_2). \quad \text{(B13)}$$

As one might intuitively expect, there is a one-parameter family of Weyl solutions (labeled by the value of $\bar{\alpha}_1$) that are diffeomorphic to a particular static axisymmetric vacuum metric with prescribed values of the multipole moments.

[1] A thorough discussion of the underlying physical concepts is given by R. Geroch, in *Asymptotic Structure of Space-Time*, edited by P. Esposito and L. Witten (Plenum, New York, 1977). See also the review articles by E. T. Newman and K. P. Tod, A. Ashtekar, and J. Winicour, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.

[2] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. R. Soc. London Ser. A **269**, 21 (1962).

[3] R. Sachs, Proc. R. Soc. London Ser. A **270**, 103 (1962).

[4] R. Arnowitt, S. Deser, and C. W. Misner, Phys. Rev. **117**, 1695 (1960); **118**, 1100 (1960); **121**, 1566 (1961); **122**, 997 (1961); in *Gravitation, An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[5] P. G. Bergmann, Phys. Rev. **124**, 274 (1961); in *Proceedings on Theory of Gravitation* (Gauthier–Villars, Paris, 1964).

[6] R. Penrose, in *Relativity, Groups and Topology: The 1963 Les Houches Lectures*, edited by B. deWitt and C. deWitt (Gordon and Breach, New York, 1964); Proc. R. Soc. London Ser. A **284**, 159 (1965); in *Battelle Recontres*, edited by B. S. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[7] R. Sachs, Phys. Rev. **128**, 2851 (1962).

[8] R. Penrose, in *Group Theory in Non-Linear Problems*, edited by A. O. Barut (Reidel, Boston, 1974); J. Winicour, J. Math. Phys. **9**, 861 (1968).

[9] R. Geroch, J. Math. Phys. **13**, 956 (1972).

[10] A. Ashtekar and R. O. Hansen, J. Math. Phys. **19**, 1542 (1978).

[11] A. Ashtekar and A. Magnon-Ashtekar, Phys. Rev. Lett. **43**, 181 (1979).

[12] A. Ashtekar and A. Magnon-Ashtekar, J. Math. Phys. **20**, 793 (1979).

[13] A. Ashtekar and M. Streubel, J. Math. Phys. **20**, 1362 (1979).

[14] R. Geroch, J. Math. Phys. **11**, 2580 (1970).

[15] R. O. Hansen, J. Math. Phys. **15**, 46 (1974).

[16] P. Kundu, J. Math. Phys. **22**, 2006 (1981).

[17] R. Beig and W. Simon, Commun. Math. Phys. **78**, 75 (1980); Proc. R. Soc. London Ser. A **376**, 333 (1981). It may be pointed out that the analyticity theorem proved by these authors in the stationary case involves assumptions that are somewhat stronger than those of Ref. 16.

[18] P. Kundu, J. Math. Phys. **22**, 1236 (1981).

[19] R. Geroch and G. T. Horowitz, Phys. Rev. Lett. **40**, 203 (1978).

[20] We have preferred to suspend the standard convention of raising and lowering of indices by the metric even at the expense of slightly complicating the form of the various equations in order to avoid any confusion about which metric is used for this purpose.

[21] R. Geroch, J. Math. Phys. **12**, 918 (1971).
[22] The formalism used here is also presented in P. G. Bergmann, *An Introduction to the Theory of Relativity* (Prentice-Hall, New York, 1942), Chap. XVII, in connection with an exposition of the Kaluza-Klein unified field theory.
[23] As usual we use the square and the round brackets to denote, respectively, antisymmetrization and symmetrization of a group of indices.
[24] Our sign conventions for the Riemann and the Ricci tensor follow that used by Geroch: $\nabla_{[\mu}\nabla_{\nu]}V_\rho = \frac{1}{2}R_{\mu\nu\rho\sigma}V^\sigma$, $R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}$.
[25] M. G. J. van der Burg, Proc. R. Soc. London Ser. A **294**, 112 (1966).
[26] M. G. J. van der Burg, Proc. R. Soc. London Ser. A **303**, 37 (1968).
[27] E. T. Newman and R. Penrose, Proc. R. Soc. London Ser. A **305**, 175 (1968).
[28] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962); **4**, 998(E) (1963); E. T. Newman and T. W. J. Unti, *ibid.* **3**, 891 (1962).
[29] A. I. Janis and E. T. Newman, J. Math. Phys. **6**, 902 (1965).
[30] A. Papapetrou, Ann. Inst. H. Poincaré A **4**, 83 (1966).

On generalized Kerr–Schild transformations

E. Nahmad-Achar
Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Apartado Postal 70-543, México D.F. 04510, Mexico

(Received 6 November 1987; accepted for publication 23 March 1988)

The geometric properties of generalized Kerr–Schild metric transformations (KST's) are studied in detail, and the energy and momentum induced by the transformation are calculated explicitly. A characterization is given of all algebraically special metrics, obtained via a KST, that satisfy the weak energy condition.

I. INTRODUCTION

Einstein's equations of general relativity simplify considerably when dealing with algebraically special metrics, that is, metrics for the Weyl conformal tensor has a repeated principal null direction. Perhaps the most important examples of these are the Kerr–Schild[1] metrics, which are of the form

$$\eta_{\alpha\beta} + n_\alpha n_\beta \,, \qquad (1)$$

where $\eta_{\alpha\beta} = \mathrm{diag}(1, -1, -1, -1)$ is the Minkowski flat metric, and η_α is a null vector with respect to $\eta_{\alpha\beta}$ (and consequently with respect to $\eta_{\alpha\beta} + n_\alpha n_\beta$). A transformation that adds the "square" of a null vector to the flat Minkowski metric, bringing it into the form (1), is called a *Kerr–Schild transformation.* In general, it is now customary to call a transformation of the form

$$g_{\alpha\beta} \to \hat{g}_{\alpha\beta} = g_{\alpha\beta} + n_\alpha n_\beta \,, \qquad (2)$$

where $g_{\alpha\beta}$ is an arbitrary metric, *a generalized Kerr–Schild transformation.* We shall refer to $g_{\alpha\beta}$ as the "background" or "old" metric, and to $\hat{g}_{\alpha\beta}$ as the "new" metric. For brevity, we shall denote such a transformation simply by KST.

The Kerr[2] metric itself, representing the space-time geometry outside a rotating black hole, is of the Kerr–Schild type and may be obtained from flat space via such a transformation.[1,3] In fact, KST's have rather curious properties; for instance, if we allow **n** to be a "timelike" Killing vector field in the direction of an extra fifth dimension, then (2) gives us the Kaluza–Klein unified theory of electromagnetic and gravitational fields, where **n** is identified with the electromagnetic vector potential. But even if we restrict ourselves to four dimensions and null vectors, these transformations have proved useful in obtaining a better understanding of the physical implications of general relativity, by providing us with a whole class of new metrics whose properties are rather easy to analyze.[4]

A known method for obtaining new metrics from old is that of complex translation along one of the coordinates. Newman and Janis[5] and Schiffer *et al.*[6] obtained in this way the Kerr geometry from Schwarzschild, and Newman *et al.*[7] obtained the Kerr–Newman metric from Reissner–Nordström. Complex translation, however, is allowed in exact general relativity when a coordinate system can be found in which Einstein's equations are linear, and this happens to be true in the algebraically special geometries of the Kerr–Schild type.[8] It is actually in trying to obtain new metrics from old ones that most of the work on KST's has been done: Robinson and Robinson[9], Hughston,[10] and Stephani[11] gave methods for obtaining new solutions with pure radiation fields starting from algebraically special vacuum metrics; Xanthopoulos[12] used KST's to obtain vacuum solutions from linearized ones; Talbot[13] obtained vacuum algebraically special space-times with zero shear; Taub[14] obtained solutions containing a null fluid with no pressure from vacuum metrics, through KST's, and also solutions containing a perfect fluid with anisotropic pressures from a Friedmann–Robertson–Walker metric.

Only Bilge and Gürses[15] have studied, independently of the author, how some of the geometric properties of the background space-time transform under this class of transformations; their work was geared mainly towards obtaining new exact solutions with, again, null-fluid distributions (this time with a nonvanishing cosmological constant).

It is not necessarily true, however, that the new metric so constructed are what one might call *physically reasonable* metrics, in the sense, for instance, that they satisfy the positive-energy (weak or dominant) condition. This is something that we want to study here. In Sec. II we will look in detail at the geometric properties of both the new and the old metrics. We show that a KST induces energy and momentum in the space-time and calculate it explicitly in Sec. III. In Sec. IV we shall characterize all the algebraically special metrics, obtained via a KST, that satisfy the weak energy condition. Section V contains a result to the effect that if two metrics are related by a KST and one of them is empty and algebraically special, then the other one is algebraically special as well. We finish in Sec. VI with a few comments.

This paper draws heavily on Chaps. 6 and 7 of the author's Ph.D. dissertation.[16]

II. GEOMETRIC PROPERTIES OF KST's

We wish here to look at generalized Kerr–Schild transformations in their own right, and to see what geometric properties are inherited from the background by the new metric. Thus consider

$$g_{\mu\nu} \to \hat{g}_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu \quad (\hat{g}^{\mu\nu} = g^{\mu\nu} - n^\mu n^\nu) \qquad (3)$$

and let $\{\mathbf{l},\mathbf{n},\mathbf{m},\bar{\mathbf{m}}\}$ be a null tetrad with respect to $g_{\mu\nu}$; the corresponding null tetrad with respect to $\hat{g}_{\mu\nu}$ is $\{\hat{\mathbf{l}} = \mathbf{l} - \frac{1}{2}\mathbf{n}, \hat{\mathbf{n}} = \mathbf{n}, \hat{\mathbf{m}} = \mathbf{m}, \hat{\bar{\mathbf{m}}} = \bar{\mathbf{m}}\}$. Note also that $\hat{l}_\alpha = \hat{g}_{\alpha\beta}\hat{l}^\beta = l_\alpha + \frac{1}{2}n_\alpha$. (We use the Newman–Penrose[17] notation throughout.)

 0022-2488/88/081879-06$02.50

Perhaps the easiest way to relate the spin coefficients of both metrics is to look at the commutators of the tetrad null vectors. For instance,

$$[\hat{\Delta},\hat{D}] = (\hat{\gamma}+\hat{\bar{\gamma}})\hat{D} + (\hat{\epsilon}+\hat{\bar{\epsilon}})\hat{\Delta} - (\hat{\tau}+\hat{\bar{\pi}})\hat{\bar{\delta}} - (\hat{\bar{\tau}}+\hat{\pi})\hat{\delta}$$
$$= (\hat{\gamma}+\hat{\bar{\gamma}})D + [(\hat{\epsilon}+\hat{\bar{\epsilon}}) - \tfrac{1}{2}(\hat{\gamma}+\hat{\bar{\gamma}})]\Delta$$
$$- (\hat{\tau}+\hat{\bar{\pi}})\bar{\delta} - (\hat{\bar{\tau}}+\hat{\pi})\delta$$

and

$$[\hat{\Delta},\hat{D}] = [\Delta, D - \tfrac{1}{2}\Delta]$$
$$= [\Delta, D] = (\gamma+\bar{\gamma})D$$
$$+ (\epsilon+\bar{\epsilon})\Delta - (\tau+\bar{\pi})\bar{\delta} - (\bar{\tau}+\pi)\delta\,, \quad (4)$$

from which we have

$$(\hat{\gamma}+\hat{\bar{\gamma}}) = (\gamma+\bar{\gamma})\,, \quad (5a)$$
$$(\hat{\tau}+\hat{\bar{\pi}}) = (\tau+\bar{\pi})\,, \quad (5b)$$
$$(\hat{\bar{\tau}}+\hat{\pi}) = (\bar{\tau}+\pi)\,, \quad (5c)$$
$$(\hat{\epsilon}+\hat{\bar{\epsilon}}) - \tfrac{1}{2}(\hat{\gamma}+\hat{\bar{\gamma}}) = (\epsilon+\bar{\epsilon})\,. \quad (5d)$$

Similarly, from the commutators we obtain 11 equations that, together with the above, can be used to work out the new spin coefficients in terms of the old ones. The result is

$$\hat{\nu} = \nu\,, \quad \hat{\lambda} = \lambda\,, \quad \hat{\sigma} = \sigma + \tfrac{1}{2}\bar{\lambda}\,,$$
$$\hat{\tau} = \tau\,, \quad \hat{\pi} = \pi\,, \quad \hat{\alpha} = \alpha + \tfrac{1}{4}\nu\,,$$
$$\hat{\beta} = \beta + \tfrac{1}{4}\bar{\nu}\,, \quad \hat{\kappa} = \kappa - \tfrac{1}{2}(\tau+\bar{\pi}) + \bar{\alpha} + \beta + \tfrac{1}{4}\bar{\nu}\,, \quad (6)$$
$$\hat{\mu} = \mu\,, \quad \hat{\gamma} = \gamma\,,$$
$$\hat{\epsilon} = \epsilon + \tfrac{1}{2}\bar{\gamma} + \tfrac{1}{4}(\mu - \bar{\mu})\,, \quad \hat{\rho} = \rho + \tfrac{1}{2}\mu\,.$$

Directly from these relations we may note the following: If **n** is tangent to a geodesic congruence in the old space-time, and the tetrad is parallely propagated along **n**, then we know that[17] $\nu = \gamma = \tau = 0$. Equations (6) then say that $\hat{\nu} = \hat{\gamma} = \hat{\tau} = 0$, i.e., the congruence will remain geodesic and the new tetrad will also be parallely propagated along **n**. Furthermore, if **n** is hypersurface orthogonal in the old metric, then $\mu = \bar{\mu}$ and, since this implies $\hat{\mu} = \hat{\bar{\mu}}$ by (6), **n** will also be hypersurface orthogonal in the new metric.

One would in principle like to relate the energy-momentum tensors of the two metrics. This, apart from being the natural thing to do, is because otherwise *any* metric obtained via a KST will satisfy Einstein's equations (just *define* its energy-momentum tensor through the field equations themselves) and will be left so general that we will not be able to extract from it almost any information at all. For example, one could ask that the energy-momentum tensors of the two metrics be equal: $\hat{T}_\alpha{}^\beta = T_\alpha{}^\beta$. It is easy to see, however, that under this identification $\hat{T}_\alpha{}^\beta$ will not in general satisfy the conservation law $\hat{T}_\alpha{}^\beta{}_{|\beta} = 0$.[16] Hereafter, a vertical bar will denote covariant differentiation with respect to the new metric $\hat{g}_{\alpha\beta}$, while a semicolon is reserved for covariant differentiation with respect to the old metric $g_{\alpha\beta}$.

Perhaps the mildest and most natural identification is that the component of the Ricci tensor along the null vector **n** be equal in both metrics, i.e., $\hat{R}_{\alpha\beta}n^\alpha n^\beta = R_{\alpha\beta}n^\alpha n^\beta$. This is mild because it is only one equation; it is natural because the vector **n** itself is singled out by the transformation, and because, as we shall see below, this is one of the consequences of choosing **n** to be geodesic. In what follows we shall study solutions of Einstein's field equations related by a KST and for which

$$\hat{R}_{nn} \equiv \hat{R}_{\mu\nu}n^\mu n^\nu = R_{\mu\nu}n^\mu n^\nu \equiv R_{nn}\,. \quad (7)$$

The actual relation between the energy-momentum tensors will be obtained explicitly in the next section. We shall see that we can express $\hat{T}_\alpha{}^\beta$ as $T_\alpha{}^\beta + L_\alpha{}^\beta$. This $L_\alpha{}^\beta$ is the most general (mixed) tensor that makes $\hat{T}_\alpha{}^\beta{}_{|\beta}$ vanish while retaining Eq. (7), and we interpret it as the *induced* energy-momentum in the new space-time by the transformation. We shall here consider what the consequences of Eq. (7) are on the spin coefficients.

Using the Ricci identities in Newman–Penrose notation (Eqs. 4.2 of Newman and Penrose,[17] hereafter denoted by NP 4.2) we can relate the components of the Weyl tensor in both metrics. However, one identity (viz. NP 4.2n) does not involve any of the Weyl components while it involves R_{nn}, and it happens to be very illuminating:

$$\delta\nu - \Delta\mu = (\mu^2 + \lambda\bar{\lambda}) + (\gamma+\bar{\gamma})\mu - \bar{\nu}\pi$$
$$+ [\tau - 3\beta - \bar{\alpha}]\nu + \Phi_{22}$$

and

$$\delta\nu - \Delta\mu = \hat{\delta}\hat{\nu} - \hat{\Delta}\hat{\mu} = (\mu^2 + \lambda\bar{\lambda}) + (\gamma+\bar{\gamma})\mu$$
$$- \bar{\nu}\pi + [\tau - 3\beta - \tfrac{3}{4}\bar{\nu} - \bar{\alpha} - \tfrac{1}{4}\nu]\nu + \Phi_{22}\,, \quad (8)$$

where we have used Eqs. (6) and (7) and the indices 1, 2, 3, 4 refer to the vectors *l*, *n*, *m*, $\bar{m}$, respectively (thus $R_{22} = R_{nn}$, etc.). Equation (8) implies

$$\nu = 0\,, \quad (9)$$

which in turn implies, using (6),

$$\bar{\nu} = 0\,. \quad (10)$$

That is, in order for $\hat{g}_{\alpha\beta}$ to be a solution of Einstein's field equations with $\hat{R}_{nn} = R_{nn}$ we must choose **n** to be a geodesic null vector in the background space-time (and by a previous result, it will automatically be geodesic in the new space-time). In their original paper, Kerr and Schild,[1] working with a flat metric $g_{\mu\nu} = \eta_{\mu\nu}$, used this fact of **n** being geodesic (which was taken as an assumption) to define another vector field **k** by $k^\mu = n^\mu/\sqrt{2H}$ where the function H is chosen so as to have $k_{\nu;\mu}k^\mu = 0$. We will not use this simplication here.

Using Eqs. (9) and (10), the relation between the spin coefficients [Eq. (6)] reduces to

$$\hat{\nu} = \nu = 0\,, \quad \hat{\lambda} = \lambda\,, \quad \hat{\mu} = \mu\,,$$
$$\hat{\gamma} = \gamma\,, \quad \hat{\tau} = \tau\,, \quad \hat{\pi} = \pi\,,$$
$$\hat{\alpha} = \alpha\,, \quad \hat{\beta} = \beta\,, \quad \hat{\sigma} = \sigma + \tfrac{1}{2}\bar{\lambda}\,, \quad (11)$$
$$\hat{\rho} = \rho + \tfrac{1}{2}\mu\,, \quad \hat{\epsilon} = \epsilon + \tfrac{1}{2}\bar{\gamma} + \tfrac{1}{4}(\mu - \bar{\mu})\,,$$
$$\hat{\kappa} = \kappa + \bar{\alpha} + \beta - \tfrac{1}{2}(\tau + \bar{\pi})\,.$$

As one might have expected, it is only those quantities related to the **l**-congruence that change with the transformation, for it is only this congruence that is different in both metrics.

One can exploit the fact that eight of the twelve (complex) spin coefficients are equal, in the remaining Ricci identities, and in this way relate the Weyl and Ricci tensors. For instance, we may obtain

$$\text{NP 4.2j:}\quad \widehat{\Psi}_4 = \Psi_4\,, \tag{12a}$$

$$\text{NP 4.2r:}\quad \widehat{\Psi}_3 = \Psi_3\,, \tag{12b}$$

$$\text{NP 4.2o:}\quad \widehat{\Phi}_{12} = \Phi_{12}\,, \tag{12c}$$

$$\text{NP 4.2p:}\quad \widehat{\Phi}_{02} = \Phi_{02} + \bar{\lambda}\,[\gamma + \bar{\gamma} + \tfrac{1}{2}(\bar{\mu} - \mu)] + \tfrac{1}{2}\bar{\Psi}_4\,, \tag{12d}$$

etc. Also, using Eqs. (6) and setting $\nu = 0$ in NP 4.2n, we get

$$\widehat{\Phi}_{22} = \Phi_{22}\,, \tag{13}$$

proving that if $\mathbf{n}$ is geodesic then $\widehat{R}_{nn} = R_{nn}$, the converse of our previous result.

III. INDUCED ENERGY AND MOMENTUM

We shall here obtain the energy-momentum that the KST induces in the space-time, directly from the field equations. Define

$$\Delta^{\alpha}{}_{\beta\gamma} \overset{\text{def}}{=} \Gamma^{\alpha}{}_{\beta\gamma} - \widehat{\Gamma}^{\alpha}{}_{\beta\gamma} = \tfrac{1}{2} g^{\alpha\delta}[g_{\beta\delta|\gamma} + g_{\gamma\delta|\beta} - g_{\beta\gamma|\delta}]\,, \tag{14}$$

where we have used the fact that the difference of Christoffel symbols is a tensor, and evaluated it in a local inertial frame with respect to $\hat{g}_{\alpha\beta}$. The difference in the Riemann tensors is

$$R^{\alpha}{}_{\beta\gamma\delta} - \widehat{R}^{\alpha}{}_{\beta\gamma\delta} = \Delta^{\alpha}{}_{\beta\delta|\gamma} - \Delta^{\alpha}{}_{\beta\gamma|\delta} + \Delta^{\alpha}{}_{\epsilon\gamma}\Delta^{\epsilon}{}_{\beta\delta} - \Delta^{\alpha}{}_{\epsilon\delta}\Delta^{\epsilon}{}_{\beta\gamma}\,, \tag{15}$$

from which, contracting in the first and third indices and using Eq. (14), we obtain

$$\widehat{R}_{\alpha\beta} - R_{\alpha\beta} = \Delta^{\gamma}{}_{\epsilon\beta}\Delta^{\epsilon}{}_{\alpha\gamma} - \Delta^{\gamma}{}_{\alpha\beta|\gamma}\,. \tag{16}$$

Define now

$$\mathfrak{R}^{\alpha}{}_{\beta} \overset{\text{def}}{=} \hat{g}^{\alpha\gamma}\widehat{R}_{\beta\gamma} - g^{\alpha\gamma}R_{\beta\gamma} = \widehat{R}^{\alpha}{}_{\beta} - R^{\alpha}{}_{\beta}\,, \tag{17}$$

which is the difference between the (mixed) Ricci tensors, i.e., *the part that was induced on the old Ricci tensor* by the KST. Using the equations above we get, after some algebra

$$\begin{aligned}\mathfrak{R}^{\alpha}{}_{\beta} = {} & -\tfrac{1}{2}(n^{\alpha}n_{\beta})^{|\delta}{}_{|\delta} + \tfrac{1}{2}[n^{\alpha}(\mu+\bar{\mu}) - 2n^{\alpha}(\gamma+\bar{\gamma})]_{|\beta} \\ & + \tfrac{1}{2}[n_{\beta}(\mu+\bar{\mu}) - 2n_{\beta}(\gamma+\bar{\gamma})]^{|\alpha} + \tfrac{1}{2}n^{\lambda}{}_{|\sigma}n^{\sigma}{}_{|\lambda}n^{\alpha}n_{\beta} \\ & - \tfrac{1}{2}n_{\epsilon|\sigma}n^{\epsilon|\sigma}n^{\alpha}n_{\beta} - \tfrac{3}{2}(\gamma+\bar{\gamma})^2 n^{\alpha}n_{\beta} \\ & + (\gamma+\bar{\gamma})(\mu+\bar{\mu})n^{\alpha}n_{\beta} + \widehat{R}^{\alpha}{}_{\lambda\delta\beta}n^{\lambda}n^{\delta} + \mathfrak{R}^{\gamma}{}_{\beta}n^{\alpha}n_{\gamma} \\ & + \tfrac{1}{2}\widehat{R}^{\alpha}{}_{\lambda}n_{\beta}n^{\lambda} - \tfrac{1}{2}\widehat{R}_{\beta\lambda}n^{\alpha}n^{\lambda} + n^{\alpha}n_{\beta}\Delta(\gamma+\bar{\gamma})\,. \end{aligned} \tag{18}$$

The induced energy-momentum is, through Einstein's equations,[18]

$$-8\pi\Upsilon^{\alpha}{}_{\beta} \overset{\text{def}}{=} \mathfrak{R}^{\alpha}{}_{\beta} - \tfrac{1}{2}\mathfrak{R}\delta^{\alpha}{}_{\beta}\,, \tag{19}$$

and using Eq. (18) in this last equation one can calculate explicitly each component in terms of spin coefficients. We choose here to evaluate Υ in the tetrad $[\hat{\mathbf{l}},\mathbf{n},\mathbf{m},\bar{\mathbf{m}}]$ of $\hat{g}_{\mu\nu}$, since this is the relevant space-time after the transformation, and in terms of the spin coefficients of $g_{\mu\nu}$, since these are the "original data," so to speak. We obtain (note that $\Upsilon^{\alpha}{}_{\beta}$ need not be symmetric; we are evaluating it in a null tetrad in the new space-time, where $\widehat{T}^{\alpha}{}_{\beta}$ is symmetric but where $T^{\alpha}{}_{\beta}$ is not in general)

$$\begin{aligned}-8\pi\Upsilon^{\hat{l}}{}_{\hat{l}} = {} & R_{\hat{l}nn\hat{l}} + \tfrac{1}{2}R_{nn} + \Delta(\epsilon+\bar{\epsilon}) + D(\mu+\bar{\mu}) - D(\gamma+\bar{\gamma}) - 2(\epsilon+\bar{\epsilon})(\gamma+\bar{\gamma}) \\ & + \bar{\pi}\bar{\tau} + \pi\tau - \bar{\rho}\bar{\mu} - \rho\mu - \bar{\sigma}\bar{\lambda} - \sigma\lambda - \bar{\delta}(\beta+\bar{\alpha}) - \delta(\bar{\beta}+\alpha) - (\gamma+\bar{\gamma})(\rho+\bar{\rho}) \\ & + (\beta+\bar{\alpha})[\bar{\tau} - \pi + 2\alpha] + (\bar{\beta}+\alpha)[\tau - \bar{\pi} + 2\bar{\alpha}] + 2(\bar{\beta}+\alpha)(\beta+\bar{\alpha})\,, \end{aligned} \tag{20a}$$

$$-8\pi\Upsilon^{n}{}_{\hat{l}} = R_{nn} + \lambda\bar{\lambda} + (\mu+\bar{\mu})(\gamma+\bar{\gamma}) - \tfrac{1}{2}(\mu^2 - \bar{\mu}^2)\,, \tag{20b}$$

$$\begin{aligned}-8\pi\Upsilon^{m}{}_{\hat{l}} = {} & R_{mnn\hat{l}} + \tfrac{1}{2}\Delta\bar{\pi} - \tfrac{1}{2}\bar{\delta}\bar{\lambda} + \tfrac{1}{2}\delta\mu - \tfrac{1}{2}\bar{\pi}\bar{\mu} + \tfrac{1}{2}\bar{\lambda}\,[\bar{\tau} - \pi + 5\alpha + \bar{\beta}\,] \\ & + \tfrac{1}{2}R_{nm} + \tfrac{1}{2}\bar{\mu}[\tau + 2\bar{\alpha} + 2\beta\,] - \delta(\gamma+\bar{\gamma}) - \tfrac{1}{2}\bar{\pi}(3\gamma+\bar{\gamma}) + \tfrac{1}{2}(\beta+\bar{\alpha})[2\gamma + 2\bar{\gamma} - \mu]\,, \end{aligned} \tag{20c}$$

$$-8\pi\Upsilon^{\bar{m}}{}_{\hat{l}} = -8\pi\bar{\Upsilon}^{m}{}_{\hat{l}}\,, \tag{20d}$$

$$-8\pi\Upsilon^{\hat{l}}{}_{n} = \lambda\bar{\lambda} - \tfrac{1}{2}(\mu^2 + \bar{\mu}^2) + (\mu+\bar{\mu})(\gamma+\bar{\gamma})\,, \tag{20e}$$

$$-8\pi\Upsilon^{n}{}_{n} = 0\,, \tag{20f}$$

$$-8\pi\Upsilon^{m}{}_{n} = -8\pi\Upsilon^{\bar{m}}{}_{n} \tag{20g}$$

$$= 0 \tag{20h}$$

$$\begin{aligned}-8\pi\Upsilon^{\hat{l}}{}_{m} = {} & R_{mnn\hat{l}} - \tfrac{1}{2}R_{nm} + \tfrac{1}{2}\Delta\bar{\pi} - \tfrac{1}{2}\bar{\delta}\bar{\lambda} + \tfrac{1}{2}\delta\mu - \tfrac{1}{2}\bar{\pi}[\bar{\mu} + 3\gamma + \bar{\gamma}] \\ & - \delta(\gamma+\bar{\gamma}) + \tfrac{1}{2}\bar{\lambda}\,[\bar{\tau} - \pi + 5\alpha + \bar{\beta}\,] + \tfrac{1}{2}\bar{\mu}[\tau + 2\bar{\alpha} + 2\beta\,] \\ & + \tfrac{1}{2}(\beta+\bar{\alpha})[2\gamma + 2\bar{\gamma} - \mu]\,, \end{aligned} \tag{20i}$$

$$-8\pi\Upsilon^{n}{}_{m} = 0\,, \tag{20j}$$

$$-8\pi\Upsilon^{m}{}_{m} = R_{mnnm} + \bar{\lambda}(\mu - \bar{\mu}) + 2\bar{\lambda}(\gamma+\bar{\gamma})\,, \tag{20k}$$

$$-8\pi\Upsilon^{\bar m}{}_m = R_{mnn\bar m} - \tfrac{1}{2}(\mu-\bar\mu)^2 - R_{nn} - \Delta(\gamma+\bar\gamma) + (\gamma+\bar\gamma)^2 - (\mu+\bar\mu)(\gamma+\bar\gamma)\,, \qquad (20l)$$

$$-8\pi\Upsilon^{\hat l}{}_{\bar m} = -8\pi\bar\Upsilon^{\hat l}{}_m\,, \qquad (20m)$$

$$-8\pi\Upsilon^{n}{}_{\bar m} = 0\,, \qquad (20n)$$

$$-8\pi\Upsilon^{m}{}_{\bar m} = -8\pi\Upsilon^{\bar m}{}_m\,, \qquad (20o)$$

$$-8\pi\Upsilon^{\bar m}{}_{\bar m} = -8\pi\Upsilon^{m}{}_m\,. \qquad (20p)$$

Equations (20) give the components of the induced energy momentum in an arbitrary tetrad field of the new space-time, in terms of the old spin coefficients. If the vector field **n** is chosen to be affinely parametrized, so that $\gamma+\bar\gamma=0$, and further, a null rotation about **n** is performed so that $\bar\beta+\alpha=0$, the above equations simplify considerably.

IV. KST's AND THE WEAK ENERGY CONDITION

In our actual universe, the energy-momentum tensor is made up of all the contributions from the different matter fields, and it is therefore complicated to describe exactly. However, there are a few conditions that are reasonable to assume, and the mildest of these is the *weak energy condition*, which assumes that, at every point p in space-time, we have

$$T_{\alpha\beta}X^\alpha X^\beta \geqslant 0 \qquad (21)$$

for any timelike vector **X** at p (cf., e.g., Hawking and Ellis[19]). This amounts to saying that the local energy density measured by any observer is positive, which seems to be reasonable physically.

In this section we shall investigate whether space-times obtained via a KST satisfy the weak energy condition (WEC) and are in this sense physically reasonable. In fact, we shall characterize all the algebraically special space-times that do; this characterization will also give a "stability" condition.

Equations (20) give the components of the induced energy-momentum tensor in a null tetrad in the new space-time. The corresponding orthonormal tetrad [**t**,**r**,**p**,**q**] is given by

$$\vec t = (1/\sqrt2)(\vec{\hat l}+\vec n),\quad \vec r = (1/\sqrt2)(\vec{\hat l}-\vec n)\,,\qquad \vec p = (1/\sqrt2)(\vec m+\vec{\bar m}),\quad \vec q = (1/\sqrt{2}i)(\vec m-\vec{\bar m}) \qquad (22)$$

so that the WEC can be stated, when the background space-time is empty (and therefore the induced energy momentum equals that of the new metric), as

$$8\pi\Upsilon_{tt} = 8\pi\Upsilon_{\alpha\beta}t^\alpha t^\beta = 4\pi[\Upsilon^{\hat l}{}_{\hat l} + \Upsilon^{\hat l}{}_n + \Upsilon^n{}_{\hat l} + \Upsilon^n{}_n] \geqslant 0\,. \qquad (23)$$

From Minkowski space, using a multiple of the Kinnersley[20] tetrad, one obtains the Kerr metric [which satisfies (23) with an equality, since it is vacuum metric], and one can repeat the process obtaining new metrics from Kerr, all of which are reasonable in the sense that (23) is satisfied. One can show that de Sitter space (the second best candidate to apply a KST to) leads as well to a new space that satisfies the WEC when the Kinnersley or other suitable tetrads are used (Nahmad-Achar[16]).

Looking at the general case, and independently of any particular tetrad field, we can obtain a necessary and sufficient condition for the new space-time to satisfy the weak energy condition, when the background space-time is empty and algebraically special. (An equivalent result for nonempty space-times would depend entirely on the energy-momentum tensor of the background metric since, in general, this may compensate for a negative amount of energy induced by the transformation.) Thus consider an algebraically special (Petrov type II) space-time, and align **n** along its repeated principal null direction. This implies, by the Goldberg–Sachs theorem, that $\nu=\lambda=0$. We can always choose **n** to be affinely parametrized, which implies $\gamma+\bar\gamma=0$. (N.B. This **n** can be made as small or large as wanted, retaining $\gamma+\bar\gamma=0$.) Now, from the definition of the Weyl and Ricci tensor components in NP notation, and the equations for the spin coefficient ϵ, we have

$$R_{1221} = \Psi_2 + \bar\Psi_2 + 2\Phi_{11} - 2\Lambda\,, \qquad (24a)$$

$$R_{22} = 2\Phi_{22}\,, \qquad (24b)$$

$$\Delta(\epsilon+\bar\epsilon) = -\tau\pi - \bar\tau\bar\pi - \Psi_2 - \bar\Psi_2 + 2\Lambda - 2\Phi_{11} \qquad (24c)$$

and using these expressions in Eqs. (20) we can write

$$-8\pi\Upsilon^{\hat l}{}_{\hat l} = D(\mu+\bar\mu) - \bar\rho\mu - \rho\mu - [\delta - \tau + \bar\pi - 3\bar\alpha - \beta](\bar\beta+\alpha) - [\bar\delta - \bar\tau + \pi - 3\alpha - \bar\beta](\beta+\bar\alpha)\,, \qquad (25a)$$

$$-8\pi\Upsilon^{\hat l}{}_n = -8\pi\Upsilon^n{}_{\hat l} = -\tfrac{1}{2}(\mu^2+\bar\mu^2)\,, \qquad (25b)$$

$$-8\pi\Upsilon^n{}_n = 0\,. \qquad (25c)$$

Plugging these expressions into the formulations of the WEC given by Eq. (23), we get the following result: *The (generalized) Kerr–Schild transformation (with* **n** *an affinely parametrized null vector field) of an algebraically special space-time satisfies the weak energy condition if and only if*

$$D(\mu+\bar\mu) - \rho\mu - \bar\rho\bar\mu - \mu^2 - \bar\mu^2 - [(\delta - \tau + \bar\pi - 3\bar\alpha - \beta)(\bar\beta+\alpha) + \text{c.c.}] \leqslant 0\,, \qquad (26)$$

where c.c. stands for complex conjugate.

When **n** is very "small," or when the KST is taken as a perturbation on the background space-time, the result above may be interpreted as follows: *In an algebraically special space-time in which* Eqs. (26) *does* not *hold, a Kerr–Schild-type metric perturbation cannot happen physically, for this will violate the weak energy condition.*

V. A FURTHER GEOMETRIC PROPERTY

We know (cf., e.g., Boyer and Lindquist[3]) that when two metrics are related by a KST, if one of them is flat then the other one is algebraically special. We shall here replace the assumption of flatness for the weaker condition

$$R_{\alpha\beta\gamma\delta}n^{\alpha}n^{\gamma}=Hn_{\beta}n_{\delta}\,, \tag{27}$$

where $\mathbf{n}$ is an affinely parametrized null vector field, and H is an arbitrary function. In other words, we shall show that if one metric is a vacuum (not necessarily flat) solution of Einstein's equations, satisfying Eq. (27), its KST-related metric is algebraically special.

Let $g_{\mu\nu}$ represent a vacuum space-time, i.e., $R_{\mu\nu}=0$. From the definition of the Weyl tensor we have, for the new metric,

$$\hat{C}^{\alpha\beta}{}_{\gamma\delta}n_{\beta}n^{\delta}=\hat{R}^{\alpha\beta}{}_{\gamma\delta}n_{\beta}n^{\delta}-\tfrac{1}{6}\hat{R}n^{\alpha}n_{\gamma}+\tfrac{1}{2}[\hat{R}_{\beta\gamma}n^{\beta}n^{\alpha}+\hat{R}^{\alpha}{}_{\delta}n_{\gamma}n^{\delta}]\,. \tag{28}$$

Since $\mathbf{n}$ is affinely parametrized, $\gamma+\bar{\gamma}=0$, and one can show that $\hat{R}^{\alpha}{}_{\beta\gamma\delta}n_{\alpha}n^{\gamma}=R^{\alpha}{}_{\beta\gamma\delta}n_{\alpha}n^{\gamma}$. Using this and Eq. (27) we can write Eq. (28) as

$$\hat{C}^{\alpha\beta}{}_{\gamma\delta}n_{\beta}n^{\delta}=(H-\tfrac{1}{6}\hat{R})n^{\alpha}n_{\gamma}+\tfrac{1}{2}(\hat{R}_{\beta\gamma}n^{\beta}n^{\alpha}+\hat{R}^{\alpha}{}_{\delta}n_{\gamma}n^{\delta}) \tag{29}$$

from which

$$\hat{C}_{\alpha\beta\delta[\gamma}n_{\sigma]}n^{\beta}n^{\delta}=\tfrac{1}{4}\hat{R}_{\beta\gamma}n_{\sigma}n^{\beta}n_{\alpha}-\tfrac{1}{4}\hat{R}_{\beta\sigma}n_{\gamma}n^{\beta}n_{\alpha}\,. \tag{30}$$

This is the key equation. Since $\mathbf{n}$ is geodesic, we know that $\hat{R}_{nn}=R_{nn}$, $\hat{R}_{nm}=R_{nm}$, $\hat{R}_{n\bar{m}}=R_{n\bar{m}}$; but the rhs of these three equations vanish, since $g_{\mu\nu}$ is empty. Therefore the only nonzero component of each of the terms on the rhs of (30) is the lll-component, and they are equal. That is,

$$\hat{C}_{\alpha\beta\delta[\gamma}n_{\sigma]}n^{\beta}n^{\delta}=0\,, \tag{31}$$

which says that $C_{\alpha\beta\gamma\delta}$ is of Petrov type II, with $\mathbf{n}$ along its repeated principal null direction. And this is what we wanted to prove.

In fact, when the KST is taken to be a perturbation

$$g_{\alpha\beta}\to\hat{g}_{\alpha\beta}=g_{\alpha\beta}+\epsilon n_{\alpha}n_{\beta}\,, \tag{32}$$

with $|\epsilon|\ll 1$ and constant, the same result derived above holds true. Since ϵ may have any desired magnitude (in fact it may be varied continuously to and from zero), and the Weyl tensor of the new space-time will only differ infinitesimally from that of the old one, the background space-time must also be algebraically special; i.e., *any vacuum metric satisfying Eq. (27) is of Petrov type II.* This last result has already been derived through other methods (cf. Kramer *et al.*,[4] Theorem 28.3), and means that the result proved above may be rephrased as follows: *If the background space-time is a vacuum algebraically special solution of Einstein's equations, the new space-time obtained from it via a KST is also algebraically special (with the same repeated principal null direction).*

There is a further property inherited by the new space-time, which comes as a corollary of the above result. Hall[21] has shown that if $R_{\alpha\beta}n^{\alpha}n^{\beta}=0$ and if the Weyl tensor is algebraically special at a point p (with repeated principal null direction $\mathbf{n}$), then the Riemann curvature of the "wave surfaces" of $\mathbf{n}$ (i.e., the two-dimensional subspaces of the tangent space at p that are spacelike and orthogonal to $\mathbf{n}$) are all equal at p. In our present case, $\hat{R}_{\alpha\beta}n^{\alpha}n^{\beta}=R_{\alpha\beta}n^{\alpha}n^{\beta}$; therefore, if we start from an empty space-time, $R_{\alpha\beta}n^{\alpha}n^{\beta}=0$ and (since algebraic specialty is inherited) at each point in the new space-time all the two-dimensional spacelike surfaces perpendicular to $\mathbf{n}$ will have the same Riemannian curvature.

VI. COMMENTS

As noted in the Introduction, practically all previous work on Kerr–Schild transformations has been directed toward the obtainment of new solutions of Einstein's equations, presumably because a larger catalog of solutions will provide a clearer picture of the consequences of general relativity. We believe that the wealth of results derived in this paper show that a great deal of understanding can be gained from studying the properties of the transformations themselves. The picture is far from complete. Perhaps the most immediate question is the following: Given an arbitrary space-time $g_{\mu\nu}$ and an affinely parametrized geodesic vector field $\mathbf{n}$, is it always possible to choose the rest of the null tetrad in such a way that Eq. (26) is satisfied? The answer is *no*, and the reason is simple: having $\mathbf{n}$ affinely parametrized means having $\nu=\gamma+\bar{\gamma}=0$; one is still free to perform a null rotation of the tetrad about $\mathbf{n}$ such that preserves these two equalities, and this can be chosen, e.g., to make $\alpha+\bar{\beta}=0$, but the other terms in Eq. (26) remain free. Space-times for which Eq. (26) can *always* be satisfied will always yield "physically reasonable" space-times after a KST. What do they share in common? How restricted is the class of these space-times?

One could also try to characterize the class of space-times for which the vanishing of the Bondi news tensor is preserved under KST's. This is interesting since in space-times in which the Bondi news vanishes identically one can introduce a four-parameter family of shear-free cross sections of $\mathscr{I}^+$, and select a canonical Poincaré subgroup of the total BMS symmetry group (Ashtekar[22]). This allows one to define angular momentum free of the supertranslation ambiguity, and this would be inherited by the new spaces obtained from these backgrounds.

ACKNOWLEDGMENTS

The author wishes to thank Roy Kerr for useful comments.

[1]R. P. Kerr and A. Schild, in *Atti del Convegno sulla Relatività Generale* (anniversary volume, fourth centenary of Galileo's birth) (Barbera, Florence, 1964).
[2]R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[3]R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).
[4]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[5]E. T. Newman and A. I. Janis, J. Math. Phys. **6**, 915 (1965).
[6]M. M. Schiffer, R. J. Adler, J. Mark, and C. Sheffield, J. Math. Phys. **14**, 52 (1973).
[7]E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and K. Torrence, J. Math. Phys. **6**, 918 (1965).
[8]M. Gürses and F. Gürsey, J. Math. Phys. **16**, 2385 (1975).
[9]I. Robinson and J. R. Robinson, Int. J. Theor. Phys. **2**, 231 (1969).
[10]L. P. Hughston, Int. J. Theor. Phys. **4**, 267 (1971).
[11]H. Stephani, J. Phys. A **12**, 1045 (1979).
[12]B. C. Xanthopoulos, J. Math. Phys. **19**, 1607 (1978).

[13]C. S. Talbot, Commun. Math. Phys. **13**, 45 (1969).
[14]A. H. Taub, Ann. Phys. (NY) **134**, 326 (1981).
[15]A. H. Bilge and M. Gürses, in *Proceedings of the XIth International Colloquium on Group Theoretical Methods in Physics,* edited by M. Sedaroglu and E. Inonu (Springer, Berlin, 1983).
[16]E. Nahmad-Achar, Ph.D. dissertation, Cambridge University, 1986.
[17]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[18]The Riemann tensor is here defined as $R^{\mu}{}_{\gamma\beta\alpha}\xi^{\gamma} = \xi^{\mu}{}_{;\alpha\beta} - \xi^{\mu}{}_{;\beta\alpha}$, and the Ricci tensor as $R_{\alpha\beta} = R^{\gamma}{}_{\alpha\gamma\beta}$. This means that Einstein's equations take the form $R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -8\pi T_{\mu\nu}$ in order to have $T_{00}>0$ as is usual.
[19]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., Cambridge, 1973).
[20]W. Kinnersley, J. Math. Phys. **10**, 1195 (1969).
[21]G. S. Hall, Z. Naturforsch. **33a**, 559 (1978).
[22]A. Ashtekar, in *Cosmology and Gravitation,* edited by P. G. Bergmann and V. de Sabbata (Plenum, New York, 1980).

States and derivations on quasi-*-algebras

Camillo Trapani
Istituto di Fisica dell'Università di Palermo, I-90123 Palermo, Italy

(Received 28 October 1987; accepted for publication 9 March 1988)

The Gel'fand–Naimark–Segal representation is extended to quasi-*-algebras, provided that a suitable choice of states is made. Derivations of quasi-*-algebras and, in particular, of left-Hilbert quasi-*-algebras are studied.

I. INTRODUCTION

Quasi-*-algebras (Refs. 1 and 2) seem to provide the natural framework for the mathematical description of several physical theories.

For instance, in the algebraic approach to statistical physics the problem of performing the thermodynamical limit of the local Heisenberg dynamics, which is not, in general, an element of the algebra of observables of the system but lies in some completion of it, has led workers to introduce in the set of observables the structure of quasi-*-algebra.

The same approach has also allowed workers to overcome the difficulties arising because of the Bogolubov inequality for a Bose gas (Ref. 3).

On the other hand, several interesting properties have been found, in quantum field theory, provided that the pointlike field is supposed to belong to the quasi-*-algebra $\mathscr{L}(\mathscr{D},\mathscr{D}')$ with $\mathscr{D}$ being the set of C^∞ vectors of the energy operator H (Ref. 4).

For these reasons it seems to be worth carrying out a systematic study of quasi-*-algebras from the mathematical point of view: states, Gel'fand–Naimark–Segal (GNS) representation, derivations, etc.

This paper is mainly concerned with some of these questions.

In Sec. III we present a generalization of the GNS construction starting from a certain family of states in a quasi-*-algebra.

Section IV is devoted to the study of derivations in quasi-*-algebras. We consider, in particular, the question of whether a derivation is spatial or not.

We introduce also the notion of a left-Hilbert quasi-*-algebra and give some properties of it.

In the Appendix a generalization to the abstract case of some propositions proved in Ref. 3 is given.

II. PRELIMINARIES AND NOTATION

The main concept we have to deal with is that of the rigged Hilbert space (RHS).

Let $\mathscr{D}$ be a dense linear manifold of Hilbert space $\mathscr{H}$. Let us endow $\mathscr{D}$ with a topology t, stronger than that induced on $\mathscr{D}$ by the Hilbert norm and let $\mathscr{D}'[t']$ be its topological dual endowed with the strong dual topology. We get in this way the familiar triplet

$$\mathscr{D}\subseteq\mathscr{H}\subseteq\mathscr{D}',$$

which is called a rigged Hilbert space.

We will denote by $\mathscr{L}^+(\mathscr{D})$ (or $C_{\mathscr{D}}$, see Refs. 5 and 6) the *-algebra of all closable operators in $\mathscr{H}$ with the properties $D(A)=\mathscr{D}, D(A^*)\supseteq\mathscr{D}$ and $A\mathscr{D}\subseteq\mathscr{D}, A^*\mathscr{D}\subseteq\mathscr{D}$. The involution in $\mathscr{L}^+(\mathscr{D})$ is defined by $A\to A^+=A^*\upharpoonright\mathscr{D}$.

An Op *-algebra is any $+$ subalgebra of $\mathscr{L}^+(\mathscr{D})$. Op *-algebras provide typical instances of RHS in the following way.

Let $\mathfrak{A}$ be an Op *-algebra on $\mathscr{D}$; let us endow $\mathscr{D}$ with a graph topology $t_{\mathfrak{A}}$ defined by the seminorms

$$\phi\to\|A\phi\|,\quad A\in\mathfrak{A},$$

$t_{\mathfrak{A}}$ is in general finer than the norm topology.

We can construct using this method the triplet

$$\mathscr{D}[t_{\mathfrak{A}}]\subseteq\mathscr{H}\subseteq\mathscr{D}'[t'_{\mathfrak{A}}],$$

where $t'_{\mathfrak{A}}$ denotes the strong dual topology of $\mathscr{D}'$. Clearly t is a projective topology and the completion $\tilde{\mathscr{D}}[\tilde{t}_{\mathfrak{A}}]$ of $\mathscr{D}[t_{\mathfrak{A}}]$ is a projective limit of Hilbert spaces as well as $\mathscr{D}'[t'_{\mathfrak{A}}]$ being an inductive limit of Hilbert spaces.

Let $(\mathscr{D},\mathscr{H},\mathscr{D}')$ be a RSH. We will denote by $\mathscr{L}(\mathscr{D},\mathscr{D}')$ the set of all continuous linear maps from $\mathscr{D}[t]$ into $\mathscr{D}[t']$ (Refs. 1 and 2).

The set $\mathscr{L}^+(\mathscr{D})$ is not, in general, a subset of $\mathscr{L}(\mathscr{D},\mathscr{D}')$ but, for instance, when $\mathscr{D}$ is a reflexive space then $\mathscr{L}^+(\mathscr{D})\subseteq\mathscr{L}(\mathscr{D},\mathscr{D}')$ and each $A\in\mathscr{L}^+(\mathscr{D})$ admits a unique extension $\hat{A}$ which maps $\mathscr{D}'$ into $\mathscr{D}'$ continuously defined by

$$\langle\hat{A}\Phi,\psi\rangle=\langle\Phi,A^+\psi\rangle,\quad\forall\Phi\in\mathscr{D}',\quad\psi\in\mathscr{D}.$$

Moreover, in this case, $\mathscr{L}(\mathscr{D},\mathscr{D}')$ is a quasi-*-algebra with distinguished *-algebra $\mathscr{L}^+(\mathscr{D})$ in the following sense (Ref. 2).

Let $\mathfrak{A}$ be a linear space and $\mathfrak{A}_0$ a *-algebra contained in $\mathfrak{A}$. We say that $\mathfrak{A}$ is a quasi-*-algebra with distinguished *-algebra $\mathfrak{A}_0$ if (i) the right and left multiplications of an element of $\mathfrak{A}$ and an element of $\mathfrak{A}_0$ are always defined and linear; and (ii) an involution $+$ is defined in $\mathfrak{A}$ with the property $(AB)^+=B^+A^+$ whenever the multiplication is defined.

A quasi-*-algebra $(\mathfrak{A},\mathfrak{A}_0)$ is said to be a topological quasi-*-algebra if a locally convex topology on $\mathfrak{A}$ is defined such that (a) the involution is continuous and the multiplications are separately continuous; and (b) $\mathfrak{A}_0$ is dense in $\mathfrak{A}$.

III. CANONICAL *-REPRESENTATIONS OF QUASI-*-ALGEBRAS

Definition 3.1: Let $\mathfrak{A}$ be a quasi-*-algebra over $\mathfrak{A}_0$ and let π_0 be a *-representation of $\mathfrak{A}_0$ in Hilbert space $\mathscr{H}$ defined on a dense domain $\mathscr{D}(\pi_0)$ (Ref. 7).

 0022-2488/88/081885-06$02.50

Let $\mathscr{D}'(\pi_0)$ be the topological dual of $\mathscr{D}(\pi_0)$ endowed with the graph topology t_{π_0} defined by the seminorms

$$f\in\mathscr{D}(\pi_0)\to\|\pi_0(A_0)f\|,\quad A_0\in\mathfrak{A}.$$

Let $\mathscr{L}(\mathscr{D}(\pi_0),\mathscr{D}'(\pi_0))$ be the linear space of all continuous linear maps from $\mathscr{D}(\pi_0)[t_{\pi_0}]$ into $\mathscr{D}'(\pi_0)$ endowed with the strong dual topology t'_{π_0}.

We say that a linear map π: $\mathfrak{A}\to\mathscr{L}(\mathscr{D}(\pi_0),\mathscr{D}'(\pi_0))$ is the canonical *-representation associated with π_0 if

(i) $\pi(A^*)=\pi(A)^+$, $\forall A\in\mathfrak{A}$,

(ii) $\pi\restriction\mathfrak{A}_0=\pi_0$,

(iii) $\pi(AB)=\pi(A)\pi(B)$, whenever $A\in\mathfrak{A}_0$ or $B\in\mathfrak{A}_0$.

Remark: For a general domain $\mathscr{D}$, $\mathscr{L}^+(\mathscr{D})$ is not a subset of $\mathscr{L}(\mathscr{D},\mathscr{D}')$. The conditions given in Definition 3.1 guarantee, however, that $\pi_0(\mathfrak{A}_0)=\pi(\mathfrak{A}_0)\subseteq\mathscr{L}(\mathscr{D}(\pi_0))$, $\pi(\mathfrak{A})\subseteq\mathscr{L}(\mathscr{D}(\pi_0),\mathscr{D}'\pi_0)$, and $\pi_0(\mathfrak{A}_0)\subseteq\pi(\mathfrak{A})$.

Let $\mathfrak{A}$ be a quasi-*-algebra over $\mathfrak{A}_0$ and π_0 a *-representation of $\mathfrak{A}_0$ with domain $\mathscr{D}(\pi_0)$. As is known (Ref. 7) there exists a minimal closed extension $\bar{\pi}_0$ of representation π_0 on the domain

$$\tilde{\mathscr{D}}(\pi_0)=\bigcap_{A\in\mathfrak{A}_0}D(\overline{\pi_0(A)}).$$

Let us assume that π_0 admits a *-representation π canonically associated with π_0. What about the closure $\bar{\pi}_0$ of π_0?

Lemma 3.2: Let $\mathfrak{A}$ be a quasi-*-algebra over $\mathfrak{A}_0$, π_0 a *-representation of $\mathfrak{A}_0$, and π the *-representation of $\mathfrak{A}$ canonically associated with it. Let $\bar{\pi}_0$ be the closure of π_0. If $\mathscr{D}'(\pi_0)$ $[t'_{\pi_0}]$ is complete, then there exists a *-representation $\bar{\pi}$ in the RHS $(\tilde{\mathscr{D}}(\pi_0),\mathscr{H},\mathscr{D}'(\pi_0))$ that is canonically associated with $\bar{\pi}_0$.

Proof: Let us first remark that also $(\tilde{\mathscr{D}}(\pi_0),\mathscr{H},\mathscr{D}'(\pi_0))$ is a RHS, since the dual of the completion of a locally convex space coincides with the dual of the space itself.

If $A\in\mathfrak{A}$ then $\pi(A)$: $\mathscr{D}(\pi_0)[t_{\pi_0}]\to\mathscr{D}'(\pi_0)[t'_{\pi_0}]$ continuously; therefore it can be extended from the completion $\tilde{\mathscr{D}}(\pi_0)$ of $\mathscr{D}(\pi_0)$ to $\mathscr{D}'(\pi_0)[t'_{\pi_0}]$. It is easy to check that $\bar{\pi}(\mathfrak{A})$ is a quasi-*-algebra over $\bar{\pi}_0(\mathfrak{A}_0)$.

We will leave open the general question as to whether a *-representation π canonically associated with a *-representation π_0 of the distinguished *-algebra $\mathfrak{A}_0$ of a quasi-*-algebra $\mathfrak{A}$ does or not exist and hope to discuss it elsewhere.

We will show that the answer is positive in the case of the GNS representation determined by a certain family of states.

We need first of all a definition of state on a quasi-*-algebra.

Definition 3.3: Let $\mathfrak{A}$ be a quasi-*-algebra with unit over $\mathfrak{A}_0$. A linear functional ω on $\mathfrak{A}$ is called a state if (i) $\omega\restriction\mathfrak{A}_0$ is a state on $\mathfrak{A}_0$; (ii) $\omega(B^*A)=\overline{\omega(A^*B)}$, $\forall A,B$ such that $A\in\mathfrak{A}_0$ or $B\in\mathfrak{A}_0$; (iii) $\forall A\in\mathfrak{A}$ there exists $K_A>0$ and $C\in\mathfrak{A}_0$ such that

$$|\omega(A^*B)|^2\leqslant K_A\omega(B^*C^*CB),\quad\forall B\in\mathfrak{A}_0.$$

We say that a state ω on $\mathfrak{A}$ is an F state if instead of (iii) the following stronger condition holds: (iii.F) $\forall A\in\mathfrak{A}$ there exist $K_A>0$ and $S,T\in\mathfrak{A}_0$ such that

$$|\omega(C_1^*AC_2)|^2\leqslant K_A\omega(C_1^*S^*SC_1)\cdot\omega(C_2^*T^*TC_2)$$

for any pair C_1,C_2 of elements of $\mathfrak{A}_0$.

A state ω will be said to be faithful if

$$\{B\in\mathfrak{A}_0:\ \omega(A^*B)=0,\ \forall A\in\mathfrak{A}\}=\{0\}.$$

Remark: In a quasi-*-algebra a product ABC, with A and C elements of $\mathfrak{A}_0$, always exists and is unambiguously defined since $\mathfrak{A}$ is semiassociative (Ref. 8).

Condition (iii.F) may appear to be more unnatural than (iii) which is simply a generalization of the Cauchy–Schwarz inequality. This opinion is also supported by the fact that (iii.F) is not in general fulfilled by states on a *-algebra. However, states on C^*-algebras always satisfy a (iii.F)-like inequality and our purpose is to generalize the C^*- situation rather than that of general *-algebras.

We have now at our disposal all ingredients to build up a GNS constrution.

Let ω be a state on the quasi-*-algebra $\mathfrak{A}$ over $\mathfrak{A}_0$ and consider the set

$$\mathscr{J}_l=\{B\in\mathfrak{A}_0:\ \omega(A^*B)=0,\ \forall A\in\mathfrak{A}\}.$$

It is simple to verify that $\mathscr{J}_l=\{B\in\mathfrak{A}_0:\ \omega(B^+B)=0\}$ and therefore it is a left ideal of $\mathfrak{A}_0$. Both $\mathfrak{A}/\mathscr{J}_l$ and $\mathfrak{A}_0/\mathscr{J}_l$ are vector spaces.

Set $\mathscr{D}=\mathfrak{A}_0/\mathscr{J}_l$ and $\mathscr{E}=\mathfrak{A}/\mathscr{J}_l$; clearly $\mathscr{D}\subseteq\mathscr{E}$. Let us denote elements of $\mathscr{E}$ as ψ_A, $A\in\mathfrak{A}$. The state ω defined a bilinear form on $\mathscr{D}\times\mathscr{E}$ separating points of $\mathscr{D}$ and $\mathscr{E}$ by $\langle\psi_A,\psi_B\rangle=\omega(B^*A)$. Therefore both the Mackey topology $\tau(\mathscr{D},\mathscr{E})$ on $\mathscr{D}$ and the $\tau(\mathscr{E},\mathscr{D})$ on $\mathscr{E}$ are Hausdorff.

Let us define a map π from $\mathfrak{A}$ into the set $L(\mathscr{D},\mathscr{E})$ of linear maps from $\mathscr{D}$ into $\mathscr{E}$ by π: $A\to\pi(A)$, where $\pi(A)$: $\psi_B\in\mathscr{D}\to\psi_{AB}\in\mathscr{E}$. It is easily seen that $\pi_0=\pi\restriction\mathfrak{A}_0$ is a *-representation of $\mathfrak{A}_0$. Moreover $\forall A\in\mathfrak{A}$, $\pi(A)$ is continuous from $\mathscr{D}$ into $\mathscr{E}$ for their respective Mackey topologies.

The bilinear form $\langle\cdot,\cdot\rangle$ when both elements are chosen in $\mathscr{D}$ defines a scalar product in $\mathscr{D}$; let us denote by $\mathscr{H}(\pi)$ the Hilbert space that is the completion of $\mathscr{D}$ with respect to the norm $\|\psi_B\|=\langle\psi_B,\psi_B\rangle^{1/2}$.

We can now endow $\mathscr{D}$ with the graph topology t_π defined by the seminorms

$$\psi_B\to\|\pi(A_0)\psi_B\|,\quad A_0\in\mathfrak{A}_0,$$

and $\mathscr{D}'$ with the strong dual topology t'_π.

Let $F\in\mathscr{E}$; then $F\equiv\psi_R$ for some $R\in\mathfrak{A}$. Then, by making use of (iii) of Definition 3.3, we get

$$\begin{aligned}|F(\psi_B)&=|\langle\psi_B,\psi_R\rangle|=|\omega(R^*B)|\\&\leqslant K_R\,\omega(B^*C^*CB)^{1/2}\\&=K_R\|\psi_{CB}\|=K_R\|\pi(C)\psi_B\|,\end{aligned}$$

for some $C\in\mathfrak{A}_0$. This implies $\mathscr{E}\subseteq\mathscr{D}'$. Since $\mathscr{D}\subseteq\mathscr{E}\subseteq\mathscr{D}'$, $\mathscr{E}$ is dense in $\mathscr{D}'[t']$.

To complete the proof we need only to prove that $\pi(A)\in\mathscr{L}(\mathscr{D},\mathscr{D}')$, $\forall A\in\mathfrak{A}$.

It is just here that condition (iii.F) plays a role.

Let $\mathscr{M}$ be a bounded subset of $\mathscr{D}[t_\pi]$ and let $\hat{\mathscr{M}}=\{C\in\mathfrak{A}_0:\ \psi_C\in\mathscr{M}\}$. Then we have, for $\psi_B\in\mathscr{D}$

$$\sup_{\psi_C\in\mathscr{U}} |\langle\pi(A)\psi_B,\psi_C\rangle|$$

$$= \sup_{C\in\mathscr{U}}|\omega(C^*AB)|$$

$$\leqslant \sup_{C\in\mathscr{U}} K_A^{1/2}(C^*S^*SC)^{1/2}\omega(B^*T^*BT)^{1/2}$$

$$= K_A^{1/2}\sup_{\psi_C\in\mathscr{U}}\|\pi(S)\psi_C\|\,\|\pi(T)\psi_B\| = K'\|\pi(T)\psi_B\|,$$

for some $S,T\in\mathfrak{A}_0$.

The vector ψ_E, where E denotes the unit of $\mathfrak{A}$, is cyclic in the following sense:

$$\overline{\{\pi(A)\psi_E: A\in\mathfrak{A}\}}^{t'_\pi} = \mathscr{D}'.$$

It is clear that π is canonically associated with the nonclosed GNS representation $\pi_0 = \pi \upharpoonright \mathfrak{A}_0$.

Then we have proved the following generalization of the GNS theorem.

Proposition 3.4: Let ω be an F state on the quasi-*-algebra with unit $\mathfrak{A}$ over $\mathfrak{A}_0$. Then there exists a cyclic representation π which is the canonical *-representation of $\mathfrak{A}$ associated with the strongly cyclic GNS representation of $\mathfrak{A}_0$ generated by $\omega \upharpoonright \mathfrak{A}_0$.

Remarks: (a) Each of the operators $\pi(A)$, $A\in\mathfrak{A}$, constructed above, can be extended to a continuous map $\tilde{\pi}(A)$ form the completion $\tilde{\mathscr{D}}[\tilde{t}_\pi]$ of $\mathscr{D}[t_\pi]$ to the completion $\tilde{\mathscr{D}}'[\tilde{t}'_\pi]$ of $\mathscr{D}'[t'_\pi]$. The triplet $\tilde{\mathscr{D}},\mathscr{H},\tilde{\mathscr{D}}'$ is still a RHS when $\tilde{\mathscr{D}}$ is endowed with the Mackey topoloty $\tau(\tilde{\mathscr{D}}\tilde{\mathscr{D}}')$ that is finer than t_π. The representation $\tilde{\pi}$ defined in this way is not, however, canonically associated with π in the sense of Definition 3.1, since $\tilde{\mathscr{D}}'$ is not the dual of $\tilde{\mathscr{D}}[\tilde{t}_\pi]$.

(b) Extensions of the GNS construction to partial algebraic structures have also been considered by other authors (Refs. 9 and 10). The point of view is, however, different.

IV. DERIVATIONS ON QUASI-*-ALGEBRAS

We will now extend to quasi-*-algebras the notion of derivation.

Definition 4.1: Let $\mathfrak{A}$ be a quasi-*-algebra with distinguished *-algebra $\mathfrak{A}_0$. A *-derivation δ of $\mathfrak{A}$ is a linear map from $\mathfrak{A}$ into $\mathfrak{A}$ satisfying the following properties:

(i) $\delta(A)\in\mathfrak{A}_0$, $\forall A\in\mathfrak{A}_0$;
(ii) $\delta(A^*) = \delta(A)^*$, $\forall A\in\mathfrak{A}$;
(iii) $\delta(AB) = \delta(A)B + A\delta(B)$,
whenever $A\in\mathfrak{A}_0$ or $B\in\mathfrak{A}_0$.

Clearly if δ is a *-derivation of $\mathfrak{A}$ then $\delta \upharpoonright \mathfrak{A}_0$ is a *-derivation of $\mathfrak{A}_0$ in the usual sense.

A *-derivation is said to be spatial if there exists an element $H = H^*\in\mathfrak{A}_0$ such that

$\delta(A) = i[A,H]$, $\forall A\in\mathfrak{A}$.

As is known in a *-algebra $\mathfrak{A}$, a derivation δ and an infinitesimally invariant state ω [i.e., $\omega(\delta(A)) = 0$, $\forall A\in\mathfrak{A}$] give rise to a symmetric operator H via this GNS construction (see, for instance, Refs. 11 and 12). An analogous result holds true for quasi-*-algebras. In what follows, if ω is an F state on $\mathfrak{A}$, π will denote the nonclosed GNS representation of $\mathfrak{A}$ constructed in Proposition 3.4 whose domain is the set

$$\mathscr{D}_0(\pi) = \{\pi(A_0)\psi_E, A_0\in\mathfrak{A}_0\}.$$

Proposition 4.2: Let $\mathfrak{A}$ be a quasi-*-algebra with unit E over $\mathfrak{A}_0$ and let δ be a *-derivation on $\mathfrak{A}$. Assume that $\mathfrak{A}$ possesses an F state ω such that $\omega(\delta(A)) = 0$, for each $A\in\mathfrak{A}$.

Then there exists an operator $H = H^+\in\mathscr{L}^+(\mathscr{D}(\pi))$ such that

$$\delta^\omega(\pi(A)) = i[H,\pi(A)], \quad \forall A\in\mathfrak{A},$$

where δ^ω is the *-derivation induced by δ on the quasi-*-algebra $\pi(\mathfrak{A})$ by defining $\delta^\omega(\pi(A)) = \pi(\delta(A))$, for $A\in\mathfrak{A}$.

Proof: It is completely analogous to the statement for C^*-algebras given in Ref. 11. We give just a sketch of it for the reader's convenience.

The operator H is defined by

$$H\pi(A)_E = \pi(\delta(A))\psi_E, \quad A\in\mathfrak{A}_0.$$

The relations

$$\langle\pi(B)\psi_E,\pi(\delta(A))\psi_E\rangle$$
$$= \omega(\delta(A^*)B) = \omega(\delta(A^*B)) - \omega(A^*\delta(B))$$
$$= \langle\pi(\delta(B))\psi_E,\pi(A)\psi_E\rangle = 0$$

if $\pi(A)\psi_E = 0$, show that H is well defined and symmetric. now for $A\in\mathfrak{A}$ and $B\in\mathfrak{A}$ we get

$$\pi(\delta(A))\pi(B)\psi_E = \pi(\delta(AB))\psi_E - \pi(A)\pi(\delta(B))\psi_E$$
$$= [H,\pi(A)]\pi(B)\psi_E.$$

The following proposition concerning derivations of quasi-*-algebras in $\mathscr{L}(\mathscr{D},\mathscr{D}')$ generalizes Theorem 8 of Ref. 11 and Proposition 3.2 of Ref. 12.

Lemma 4.3: Let $\mathfrak{A}\subseteq\mathscr{L}(\mathscr{D},\mathscr{D}')$ be a quasi-*-algebra with distinguished *-algebra $\mathfrak{A}_0 = \mathfrak{A}\cap\mathscr{L}^+(\mathscr{D})$, satisfying the following conditions.

(i) For some $\phi\in\mathscr{D}$ the vector state $\omega_\phi(A) = \langle A\phi,\phi\rangle$, $A\in\mathfrak{A}$, is an F state on $\mathfrak{A}$.

(ii) $\mathfrak{A}_0$ contains all finite rank operators of $\mathscr{L}^+(\mathscr{D})$.

Then, if δ is a *-derivation of $\mathfrak{A}$, there exists an element $H = H^+$ of $\mathscr{L}^+(\mathscr{D})$ such that

$$\delta(A) = i[A,H], \quad \forall A\in\mathfrak{A}.$$

Proof: We give only a sketch of the proof that is analogous to that of Theorem 8 of Ref. 11.

Let ϕ be an element of $\mathscr{D}$ such that the associated vector state is an F state and let P be the projection onto the subspace generated by ϕ. Let us define a *-derivation δ_p on $\mathfrak{A}$, setting

$$\delta_p(A) = \delta(A) - i[A,X_p], \quad \forall A\in\mathfrak{A},$$

where $X_p = i(\delta(P)P - P\delta(P))$. By a simple calculation one finds $\langle\delta_p(A)\phi,\phi\rangle = 0$, $\forall A\in\mathfrak{A}$. By applying Proposition 4.2 it follows that there exists an operator $H_p = H_p^+ \in\mathscr{L}^+(\mathscr{D}_0(\pi))$ such that

$$\delta(A) = i[A,H_p], \quad \forall A\in\mathfrak{A}.$$

Therefore

$$\delta(A) = i[A,H_p + X_p], \quad \forall A\in\mathfrak{A}.$$

By setting $H = H_p + X_p$, δ has the desired form.

The domain $\mathscr{D}_0(\pi)$ is (isomorphic to) $\mathfrak{A}_0$ since the state $\omega_\phi(A) = \langle A\phi,\phi\rangle$ is faithful.

Now $\mathfrak{A}_0$ contains the set $[\mathscr{D}]$ of the rest classes obtained from the equivalence $\phi \equiv \psi$ if and only if there exists $\lambda \in \mathbb{C}$, $|\lambda| = 1$ such that $\phi = \lambda\psi$. The linear hull of $[\mathscr{D}]$ exactly corresponds to finite rank operators and then to $\mathscr{D}$. Moreover, $H\mathscr{D} \subseteq \mathscr{D}$ and thus H belongs to $\mathscr{L}^+(\mathscr{D})$.

Lemma 4.4: Let $\mathfrak{A} \subseteq \mathscr{L}(\mathscr{D},\mathscr{D}')$ be a quasi-*-algebra over $\mathfrak{A}_0 \subseteq \mathscr{L}^+(\mathscr{D})$ and assume that the graph topology $t_{\mathfrak{A}_0}$ of $\mathscr{D}$ is metrizable and barrelled; then each vector state is an F state.

Proof: The sesquilinear form $F_A(\phi,\psi) = \langle A\phi,\psi\rangle$ on $\mathscr{D}$ is separately continuous and therefore, by the assumption on $\mathscr{D}$, jointly continuous with respect to $t_{\mathfrak{A}_0}$; then there exist $S,T \in \mathfrak{A}_0$ such that

$$|F_A(\phi,\psi)| \leqslant K_A \|S\phi\| \, \|T\psi\|,$$

and then

$$\begin{aligned}|\omega_\phi(C_1^+ AC_2)|^2 &= |F_A(C_2\phi,C_1\psi)|^2 \\ &\leqslant K_A\omega_\phi(C_2^+ S^+SC_2)\omega_\phi(C_1^+ T^+TC_1).\end{aligned}$$

We can give now the main result of this section whose proof follows easily from Lemmas 4.3 and 4.4.

Proposition 4.5: Let T be a densely defined self-adjoint operator in $\mathscr{H}$ and let $\mathscr{D} = \mathscr{D}^\infty(T) = \cap_{n>0} D(T^n)$ be endowed with the graph topology t_T.

If δ is a *-derivation of $\mathscr{L}(\mathscr{D},\mathscr{D}')$, then there exists an operator $H = H^+ \in \mathscr{L}^+(\mathscr{D})$ such that

$$\delta(A) = i[A,H], \quad \forall A \in \mathscr{L}(\mathscr{D},\mathscr{D}').$$

V. HILBERT QUASI-*-ALGEBRAS

As is known many results can be obtained for derivations on left-Hilbert algebras. Some of them have been extended to left EW*-algebras by Inoue and Ota in Ref. 12.

We will introduce here the notion of left-Hilbert quasi-*-algebra (LHQ) where the concepts of RHS and quasi-*-algebra are put together.

Definition 5.1: Let $\mathscr{D}[t] \subseteq \mathscr{H} \subseteq \mathscr{D}'[t']$ be a RHS. We say that $(\mathscr{D},\mathscr{H},\mathscr{D}')$ is a left-Hilbert quasi-*-algebra (shortly, LHQ-*-algebra) if $\mathscr{D}'[t']$ is a topological quasi-*-algebra with distinguished *-algebra $\mathscr{D}$ satisfying the following conditions:

(i) $\langle\phi,\psi\rangle = \overline{\langle\phi^*,\psi^*\rangle}, \quad \forall\phi\in\mathscr{D}, \quad \psi\in\mathscr{D}'$,

(ii) $\langle\chi,\phi\cdot\psi\rangle = \overline{\langle\psi,\phi^*\chi\rangle}, \quad \forall\psi,\chi\in\mathscr{D}, \quad \phi\in\mathscr{D}'$,

(iii) $\langle\chi,\psi\cdot\phi\rangle = \langle\psi^*\chi,\phi\rangle, \quad \forall\psi,\chi\in\mathscr{D}, \quad \phi\in\mathscr{D}'$.

There are several familiar examples of LHQ *-algebras. Think, for instance, of the space $\mathscr{S}(\mathbb{R}^n)$ of tempered distributions of $\mathbb{R}^n$; the Schwartz space $\mathscr{S}(\mathbb{R}^n)$ of fast decreasing C^∞ functions is, clearly, the distinguished *-algebra (the involution is simply the complex conjugation). Another example is the pair $(\mathscr{D}',\mathscr{D})$ considered in Theorem 2.3 of Ref. 3.

Remark: In Ref. 13, Inoue, in view of a generalization of Tomita–Takesaki theory, has introduced the concept of "left-Hilbert-partial *-algebra." As any quasi-*-algebra is a partial *-algebra, so any LHQ-*-algebra $(\mathscr{D}',\mathscr{D})$ with $\mathscr{D}\cdot\mathscr{D}$ dense in $\mathscr{D}'$ belongs to the class considered by Inoue.

Now, let $\mathscr{D}'$ be a LHQ-*-algebra; for $\phi\in\mathscr{D}'$, we can define two linear maps $L(\phi)$ and $R(\phi)$ from $\mathscr{D}$ into $\mathscr{D}'$ by

$$L(\phi)\chi = \phi\chi, \quad \forall\chi\in\mathscr{D}, \quad R(\phi)\chi = \chi\phi, \quad \forall\chi\in\mathscr{D}.$$

Lemma 5.2: For $\phi\in\mathscr{D}'$, both $L(\phi)$ and $R(\phi)$ are elements of $\mathscr{L}(\mathscr{D},\mathscr{D}')$. In particular, if $\phi\in\mathscr{D}$ then $L(\phi)$ and $R(\phi)$ are elements of $\mathscr{L}^+(\mathscr{D})$. Moreover

$$L(\phi)^+ = L(\phi^*), \quad R(\phi)^+ = R(\phi^*).$$

Proof: $L(\phi)$ and $R(\phi)$ are elements of $\mathscr{L}(\mathscr{D},\mathscr{D}')$ since the multiplication is, by hypothesis, continuous from $\mathscr{D}'[t']$ into itself. Now for $\psi,\chi\in\mathscr{D}$ we have

$$\langle L(\phi)\psi,\chi\rangle = \langle\phi\cdot\psi,\chi\rangle = \langle\psi,\phi^*\chi\rangle = \langle\psi,L(\phi^*)\chi\rangle.$$

The proof for $R(\phi)$ is analogous.

By means of the maps L and R defined above it is possible to define two *-representations of $\mathscr{D}'$ in the following way:

$$\pi_R(\phi)\chi = R(\phi)\chi, \quad \forall\chi\in\mathscr{D},$$

$$\pi_L(\phi)\chi = L(\phi)\chi, \quad \forall\chi\in\mathscr{D}.$$

These two representations are called, respectively, the right- and left-regular representations of $\mathscr{D}'$. They are, as is easily seen, canonically associated, in the sense of Definition 3.1, with the right- and left-regular representations of $\mathscr{D}$. Moreover π_R and π_L are faithful if either $\mathscr{D}$ has a unit or $\mathscr{D}\cdot\mathscr{D}$ is dense $\mathscr{D}[t]$.

Our purpose is now to characterize *-derivations of a certain class of LHQ-*-algebras.

Definition 5.3: A LHQ-*-algebra $(\mathscr{D}',\mathscr{D})$ is said to be of class S if $\mathscr{L}^+(\mathscr{D}) \subseteq \mathscr{L}(\mathscr{D},\mathscr{D}')$.

From now on we will deal only with LHQ-*-algebras of class S.

Let $(\mathscr{D}',\mathscr{D})$ be a LHQ-*-algebra of class S and assume that it possesses a unit $e\in\mathscr{D}$. If δ is a *-derivation of $\mathscr{D}'$ with the property $\langle e,\delta\phi\rangle = 0$, $\forall\phi\in\mathscr{D}'$, then $\delta_0 = \delta \upharpoonright \mathscr{D} \in \mathscr{L}^+(\mathscr{D})$ and consequently δ coincides with the extension $\hat{\delta}_0$ of δ_0 to $\mathscr{D}'$ defined as the unique map from $\mathscr{D}'$ into $\mathscr{D}'$ satisfying the relation

$$\langle\psi,\hat{\delta}_0\phi\rangle = -\langle\delta_0\psi,\phi\rangle, \quad \forall\psi\in\mathscr{D}, \quad \phi\in\mathscr{D}'.$$

Therefore $\delta\in\mathscr{L}(\mathscr{D},\mathscr{D}')$ and is a continuous derivation.

Conversely if δ_0 is a *-derivation of the *-algebra $\mathscr{D}$ with the property $\langle e,\delta_0\phi\rangle = 0$, $\forall\phi\in\mathscr{D}$, then $\delta_0\in\mathscr{L}^+(\mathscr{D})$ and $\delta_0^+ = -\delta_0$. Hence δ_0, as well as each element of $\mathscr{L}^+(\mathscr{D})$, can be extended to a map δ defined all over $\mathscr{D}'$. It is easy to prove that $\delta\phi^* = (\delta\phi)^*$ $\forall\phi\in\mathscr{D}'$ and $\delta(\phi\psi) = \delta\phi\cdot\psi + \phi\delta\psi$ whenever $\phi\in\mathscr{D}$ or $\psi\in\mathscr{D}$. Therefore δ is a *-derivation of $\mathscr{D}'$ and is continuous, since $\delta\in\mathscr{L}(\mathscr{D},\mathscr{D}')$.

Lemma 5.4: Let $(\mathscr{D}',\mathscr{D})$ be a LHQ-*-algebra. A linear map δ from $\mathscr{D}'$ into $\mathscr{D}'$ is a *-derivation if, and only if, the following three conditions are fulfilled:

(i) $\delta_0 = \delta \upharpoonright \mathscr{D}$ maps $\mathscr{D}$ into itself,

(ii) $\delta\phi^* = (\delta\phi)^*$, $\forall \phi \in \mathscr{D}'$,

(iii) $[L(\phi),\delta_0] = L(\phi)\delta_0 - \delta L(\phi)$

$$= -L(\delta\phi), \quad \forall \phi \in \mathscr{D}'.$$

Proof: We need only to prove that (iii) is equivalent to (iii) of Definition 4.1.

Let δ be a derivation and $\psi \in \mathscr{D}$; then

$$\begin{aligned}[L(\phi),\delta_0]\psi &= L(\phi)\delta_0\psi - \delta L(\phi)\psi \\ &= L(\phi)\delta_0\psi - \delta(\phi\cdot\psi) \\ &= L(\phi)\delta_0\psi - \delta\phi\cdot\psi - \phi\cdot\delta\psi \\ &= -\delta\phi\cdot\psi = -L(\delta\phi)\psi.\end{aligned}$$

Conversely if (iii) holds and $\phi \in \mathscr{D}'$, $\psi \in \mathscr{D}$ we have

$$\begin{aligned}\delta(\phi\cdot\psi) = \delta L(\phi)\psi &= (L(\phi)\delta_0 + L(\delta\phi))\psi \\ &= \phi\cdot\delta_0\psi + \delta\phi\cdot\psi.\end{aligned}$$

Let now δ be a *-derivation of $\mathscr{D}'$ such that $\delta_0 = \delta \upharpoonright \mathscr{D} \in \mathscr{L}^+(\mathscr{D})$.

It is then possible to define a *-derivation Δ of $\mathscr{L}(\mathscr{D},\mathscr{D}')$ putting

$$\Delta(A) = [\delta_0,A], \quad \forall A \in \mathscr{L}(\mathscr{D},\mathscr{D}').$$

Let us put $L(\mathscr{D}') = \{L(\phi),\ \phi \in \mathscr{D}'\} \subseteq \mathscr{L}(\mathscr{D},\mathscr{D}')$. Clearly $L(\mathscr{D}')$ is a quasi-*-algebra with distinguished *-algebra $L(\mathscr{D}) = \{L(\phi),\ \phi \in \mathscr{D}\} \subseteq \mathscr{L}^+(\mathscr{D})$. We want to show that $\Delta \upharpoonright L(\mathscr{D}')$ acts in the same way as δ on $\mathscr{D}'$. We have, in fact, for $\phi \in \mathscr{D}'$, $\psi \in \mathscr{D}$,

$$\begin{aligned}\Delta(L(\phi))\psi &= [\delta_0,L(\phi)]\psi \\ &= \delta(L(\phi)\psi) - L(\phi)\delta\psi \\ &= \delta(\phi\psi) - \phi\cdot\delta\psi = L(\delta\phi)\psi.\end{aligned}$$

On the other hand, if Δ is a *-derivation of $\mathscr{L}(\mathscr{D},\mathscr{D}')$ such that $\Delta L(\mathscr{D}') \subseteq L(\mathscr{D}')$, if L is invertible, it is possible to define a *-derivation δ of $\mathscr{D}'$ by

$$\delta\phi = L^{-1}\Delta(L(\phi)), \quad \forall \phi \in \mathscr{D}'.$$

In Lemma 4.3 we gave conditions for a *-derivation of a quasi-*-algebra $\mathfrak{A} \subseteq \mathscr{L}(\mathscr{D},\mathscr{D}')$ to be spatial. We will show now that a *-derivation of $L(\mathscr{D}')$ is, in some sense, spatial under lighter assumptions.

Proposition 5.5: Let Δ be a *-derivation of $L(\mathscr{D}')$ with $(\mathscr{D}',\mathscr{D})$ a LHQ-*-algebra such that either $\mathscr{D}\cdot\mathscr{D}$ is dense in $\mathscr{D}[t]$ or $\mathscr{D}$ has a unit e.

Then if we set

$$\delta\phi = L^{-1}\Delta(L(\phi)), \quad \forall \phi \in \mathscr{D}',$$

we get

$$\Delta(L(\phi)) = \delta L(\phi) - L(\phi)\delta_0, \quad \forall \phi \in \mathscr{D}',$$

where $\delta_0 = \delta \upharpoonright \mathscr{D}$.

Proof: If δ is a defined as above, then δ is a *-derivation of $\mathscr{D}'$ and $L(\delta\phi) = \Delta(L(\phi))$. Then (iii) of Lemma 5.4 implies the statement.

We conclude this section with the following corollary.

Corollary 5.6: Let δ be a *-derivation of $\mathscr{D}'$ with $\mathscr{D}$ as in Proposition 5.5 and $\delta_0 = \delta \upharpoonright \mathscr{D} \in \mathscr{L}^+(\mathscr{D})$. Then

$$\delta\phi = L^{-1}[\delta_0,L(\phi)], \quad \forall \phi \in \mathscr{D}'.$$

ACKNOWLEDGMENTS

It is a pleasure to thank Professor G. Epifanio for useful discussions and a critical reading of the manuscript.

The research for the paper was supported by Fondi MPI-Università di Palermo.

APPENDIX: CANONICAL COMMUTATION RELATIONS

The relation $-L(\delta\phi) = [L(\phi),\delta_0]$ that we have seen to be valid for any derivation δ of $\mathscr{D}'$ generalizes to the abstract case the usual canonical commutation relation (CCR)

$$\left[f(x),\ -i\frac{d}{dx}\right] = i\frac{df}{dx}.$$

For this reason we will call it CCR, too.

We will generalize here the situation described in Sec. III of Ref. 3 to build up the CCR quasi-*-algebra starting from a LHQ-*-algebra $(\mathscr{D}',\mathscr{D})$ of class S with unit.

We will omit all proofs which are slight modifications of those given in Ref. 3.

Let us denote with $\mathscr{A}$ the linear space of all formal polynomials in the variable X with coefficients in $L(\mathscr{D}')$, i.e., if $f \in \mathscr{A}$ then

$$f = \sum_{k=0}^{n} L(\phi_k)X^k, \quad \phi_k \in \mathscr{D}'.$$

With $\mathscr{A}_0$ we will denote the subspace of $\mathscr{A}$ of all $f = \Sigma_K L(\phi_k)X^k$ with $\phi_k \in \mathscr{D}$.

Let δ be a *-derivation of $\mathscr{D}$ with $\delta \in \mathscr{L}^+(\mathscr{D})$. Then we start with the following representation of $\mathscr{A}$ in $\mathscr{L}(\mathscr{D},\mathscr{D}')$,

$$\pi(f) = \sum_{k=0}^{n} L(\phi_k)\delta^k,$$

for $f = \Sigma_{k=0}^{n} L(\phi_k)X^k$. We set $\mathfrak{A} = \pi(\mathscr{A})$, $\mathfrak{A}_0 = \pi(\mathscr{A}_0)$.

Lemma A.1: (i) π is a bijection of $\mathscr{A}$ onto $\mathfrak{A}$. (ii) $\mathfrak{A}_0 = \pi(\mathscr{A}_0) \subseteq \mathscr{L}^+(\mathscr{D})$ is an Op*-algebra. (iii) For $A \in \mathfrak{A}$ and $B \in \mathfrak{A}_0$, $AB, BA \in \mathfrak{A}$.

We will denote with τ_∞ the topology of uniformly bounded convergence on $\mathscr{L}(\mathscr{D},\mathscr{D}')$ defined by the seminorms

$$\|A\|_{\mathscr{M}} = \sup_{\phi,\psi \in \mathscr{M}} |\langle A\phi,\psi\rangle|, \quad \mathscr{M} \text{ bounded in } \mathscr{D}[t].$$

Lemma A.2: If in $\mathscr{D}[t]$ the multiplication is jointly continuous, $(\mathfrak{A},\mathfrak{A}_0)$ is a topological quasi-*-algebra with respect to τ_∞.

The proof, as in Lemma 3.3 of Ref. 3, lies on the following two facts: (a) $\mathscr{D}$ is dense in $\mathscr{D}'[t']$; and (b) since the multiplication in $\mathscr{D}[t]$ is jointly continuous, if $\mathscr{M}$ and $\mathscr{N}$ are bounded subsets of $\mathscr{D}[t]$ then $\mathscr{M}\cdot\mathscr{N}$ is also.

An example of the above situation is that described in Ref. 3, §3. But several examples of the same kind may be constructed making use of the notion of CCR framework we will introduce now.

Definition A.3: A CCR framework is a triplet $(\delta,{}^*, L)$, where (i) δ is a self-adjoint operator in Hilbert space $\mathscr{H}$; (ii) * is an involution in $\mathscr{D}^\infty(\delta) = \bigcap_{n>0} D(\delta^n)$ with the property $\langle\phi,\psi\rangle = \langle\psi^*,\phi^*\rangle$, $\forall\phi,\psi\in\mathscr{D}^\infty(\delta)$; (iii) L is a linear map from $\mathscr{D}^\infty(\delta)\to\mathscr{L}^+(\mathscr{D}^\infty(\delta))$ satisfying the following requirements:

(iii.a) $L(\phi^*) = (L(\phi))^+, \quad \forall\phi\in\mathscr{D}^\infty(\delta)$,
(iii.b) $\delta(L(\phi)\psi) = L(\delta\phi)\psi + L(\phi)\delta\psi, \quad \forall\phi,\psi\in\mathscr{D}^\infty(\delta)$,
(iii.c) $(L(\phi)\psi)^* = L(\psi^*)\phi^*, \forall\phi,\psi\in\mathscr{D}^\infty(\delta)$.

Remark: Starting with L and * we can define a linear map R from $\mathscr{D}^\infty(\delta)$ into $\mathscr{L}^+(\mathscr{D}^\infty(\delta))$ by $R(\phi)\psi = (L(\psi^*)\phi^*)^*, \quad \forall\psi\in\mathscr{D}^\infty(\delta)$.

Proposition A4: (a) If $(\delta,*,L)$ is a CCR framework then $\mathscr{D}^\infty(\delta)$ is a topological *-algebra with respect to the graph topology t_δ defined by the seminorms

$$\phi\to\|\phi\|_k = \|\delta^k\phi\|, \quad k=1,2,\dots,$$

when the multiplication $\phi\cdot\psi$ of two elements of $\mathscr{D}^\infty(\delta)$ is defined as

$$\phi\cdot\psi = L(\phi)\psi.$$

(b) δ is a *-derivation of the *-algebra $\mathscr{D}^\infty(\delta)$.

(c) The multiplication in $\mathscr{D}^\infty(\delta)$ satisfies the condition

$$\|\phi\psi\|_k \leqslant C_k\|\phi\|_k\|\psi\|_k,$$

where the C_k are certain constants; thus the multiplication is jointly continuous.

Proof: (a) and (b) are straightforward; (c) can be proved as Lemma 2.1 of Ref. 3.

Let $\mathscr{D}'[t']$ be the topological dual of $\mathscr{D}^\infty(\delta)[t_\delta]$, endowed with the strong dual topology t'. Since $\mathscr{D}^\infty(\delta)$ is Fréchet and reflexive, $\mathscr{L}^+(\mathscr{D}^\infty(\delta))\subseteq\mathscr{L}(\mathscr{D}^\infty(\delta),\mathscr{D}')$. Moreover $\mathscr{D}'$ has the structure of quasi-*-algebra defined as follows.

The map L can be extended from $\mathscr{D}^\infty(\delta)$ to $\mathscr{D}'$. For $F\in\mathscr{D}'$ and $\phi,\psi\in\mathscr{D}^\infty(\delta)$ we can put

$$\langle L(F)\phi,\psi\rangle = \langle F,L(\psi)\phi^*\rangle = \langle F,\psi\phi^*\rangle.$$

We will show that L maps $\mathscr{D}'$ into $\mathscr{L}(\mathscr{D}^\infty(\delta),\mathscr{D}')$.

By the definition itself we get that $L(F)\phi$ is a linear form on $\mathscr{D}^\infty(\delta)$. Moreover it is continuous; in fact,

$$|\langle L(F)\phi,\psi\rangle| = |\langle F,\psi\phi^*\rangle| \leqslant M\|\psi\phi^*\|_k \leqslant MC_k\|\psi\|_k\|\phi^*\|_k.$$

Let us now show that $L(F)\in\mathscr{L}(\mathscr{D}^\infty(\delta),\mathscr{D}')$. We have, for any bounded subset $\mathscr{M}$ of $\mathscr{D}^\infty(\delta)$,

$$\sup_{\psi\in\mathscr{M}}|\langle L(F)\phi,\psi\rangle| \leqslant MC_k(\sup_{\psi\in\mathscr{M}}\|\psi\|_k)\|\phi^*\|_k = \widetilde{M}\|\phi^*\|_k.$$

Proposition A.5: $\mathscr{D}'[t']$ is a topological quasi-*-algebra with distinguished *-algebra $\mathscr{D}^\infty(\delta)$ when the left and right multiplications are defined, for $F\in\mathscr{D}'$ and $\phi\in\mathscr{D}^\infty(\delta)$, as

$$F\cdot\phi = L(F)\phi, \quad \phi\cdot F = R(F)\phi,$$

where $\langle R(F)\phi,\psi\rangle = \langle F,\phi^*\psi\rangle$. Moreover $(\mathscr{D}',\mathscr{D}^\infty(\delta))$ is a LHQ-*-algebra of class S.

The conclusion of the previous discussion is the following proposition.

Proposition A.6: A CCR framework $(\delta,*,L)$ gives rise to a CCR quasi-*-algebra in the sense of Lemma A.2.

[1]G. Lassner, Wiss. Z. Karl Marx Univ. Leipzig Math. Naturwiss. Reihe **30**, 572 (1981).
[2]G. Lassner, Physica **124A**, 471 (1984).
[3]G. Lassner, G. A. Lassner, and C. Trapani, J. Math. Phys. **28**, 174 (1987).
[4]G. Epifanio and C. Trapani, Ann. Inst. H. Poincaré **46**, 175 (1987).
[5]G. Lassner, Rep. Math. Phys. **3**, 279 (1972).
[6]R. Ascoli, G. Epifanio, and A. Restivo, Commun. Math. Phys. **18**, 291 (1970); Riv. Mat. Univ. Parma **3**, 21 (1974).
[7]R. T. Powers, Commun. Math. Phys. **21**, 85 (1971).
[8]J.-P. Antoine and F. Mathot, Ann. Inst. H. Poincaré **46**, 299 (1987).
[9]J.-P. Antoine, in *Spontaneous Symmetry Breakdown and Related Topics* (Karpacz, 1985), edited by L. Michel, J. Mozrzymas, and A. Pelalski (World Scientific, Singapore, 1985), p. 247.
[10]J.-P. Antoine and A. Inoue, "Positive invariant sesquilinear forms on partial *-algebras," preprint UCL-IPT 87-35,1987.
[11]O. Bratteli and D. W. Robinson, Commun. Math. Phys. **42**, 253 (1975).
[12]A. Inoue and S. Ota, Trans. Am. Math. Soc. **261**, 567 (1980).
[13]A. Inoue, "A generalization of the Tomita-Takesaki theory to partial Op*-algebras," preprint, 1987.

The Cauchy problem for stringy gravity

Yvonne Choquet-Bruhat
Mécanique Relativiste, Université P. et M. Curie, 4 Place Jussieu, 75252 Paris Cedex 05, France

(Received 3 February 1988; accepted for publication 20 April 1988)

The Cauchy problem and the propagation of discontinuities for stringy gravity—that is, equations with Gauss–Bonnet terms—in arbitrary dimensions are studied.

I. INTRODUCTION

The effective Lagrangian of the low energy limit of the supersymmetric string action[1] contains, in addition to the scalar curvature of a Riemannian connection on a d-dimensional manifold ($d = 10$ in the current model which offers anomaly cancellation), quadratic and higher order invariants in the curvature tensor. If these terms appear in the so-called Gauss–Bonnet combination the corresponding field equations $A_{\alpha\beta} = 0$ are like the Einstein equations $S_{\alpha\beta} = 0$, second-order partial differential equations for the metric[2–4]: the terms $L(p)$ of degree p would be, if the manifold had dimension $2p$, the quantity corresponding to the closed form that represents the Euler class of the manifold[5] (up to a constant factor)

$$L(p) = \mathscr{C}^{\alpha_1\cdots\alpha_{2p}}_{\beta_1\cdots\beta_{2p}} R_{\alpha_1\alpha_2}{}^{\beta_1\beta_2} \cdots R_{\alpha_{2p-1}\alpha_{2p}}{}^{\beta_{2p-1}\beta_{2p}}.$$

The linearization of the field equations $A_{\alpha\beta} = 0$, called equations of stringy gravity, around flat space are identical with the linearized Einstein equations. In the second variation, around flat space, the equations $A_{\alpha\beta} = 0$ and $S_{\alpha\beta} = 0$ have the same differential operator for the second variation: the propagators used in quantization by a perturbative approach around flat space are the same in Einstein and stringy gravity.[4,6]

Various interesting, nonperturbative results have also been obtained for the equations of stringy gravity: plane wave solutions[3] (the same as for Einstein equations), spherically symmetric solutions and their stability,[6] applications to a nonlinear model of electromagnetism via Kaluza–Klein,[7] and study of characteristics.[8,9]

In this self-contained paper we consider the Cauchy problem for the equations of stringy gravity, in arbitrary dimension. We show that it is a well posed problem with constraints in an analytic framework, but that the general propagation is governed by a cone of order $2d$, which need not be convex, not even real. The properties of the generic solutions of $A_{\alpha\beta} = 0$ may be very different from those of $S_{\alpha\beta} = 0$, and there is no proof that a limit theorem will hold when the polynomial corrections tend to zero.

II. EQUATIONS

Let V be a d-dimensional C^∞ manifold, with a metric g of hyperbolic signature $(-,+,+,\ldots,+)$. This metric is said to represent "stringy gravity" if it satisfies a system of partial differential equations of the type

$$A^{\ \beta}_{\alpha} \equiv K_0\delta^\beta_\alpha + \sum_{p=1}^{\infty} K_p\, \mathscr{C}^{\beta\ \lambda_1\cdots\lambda_{2p}}_{\alpha\ \mu_1\cdots\mu_{2p}} \times R_{\lambda_1\lambda_2}{}^{\mu_1\mu_2}\cdots R_{\lambda_{2p-1}\lambda_{2p}}{}^{\mu_{2p-1}\mu_{2p}} = 0, \tag{2.1}$$

where the K_p are constants, the $\mathscr{C}^{\beta\ \lambda_1\cdots\lambda_{2p}}_{\alpha\ \mu_1\cdots\mu_{2p}}$ completely antisymmetric Kronecker symbols, and $R_{\alpha\beta}{}^{\lambda\mu}$ the Riemann curvature tensor of g. The sum is indeed finite: all terms with $2p+1>d$ are identically zero; for $d=4$ the only nonzero terms are for $p=1$. For arbitrary d the term in $p=1$ is proportional to the Einstein tensor:

$$\mathscr{C}^{\beta\ \lambda_1\lambda_2}_{\alpha\ \mu_1\mu_2} R_{\lambda_1\lambda_2}{}^{\mu_1\mu_2} \equiv -4(R^{\ \beta}_{\alpha} - \tfrac{1}{2}g^\beta_\alpha R) \equiv -4S^{\ \beta}_{\alpha}.$$

We rescale (2.1) to write it under the form

$$A^{\ \beta}_{\alpha} \equiv S^{\ \beta}_{\alpha} + \Lambda g^\beta_\alpha + \chi B^{\ \beta}_{\alpha} = 0, \tag{2.2}$$

with $P<d/2$ being some positive integer,

$$B^{\ \beta}_{\alpha} \equiv \sum_{p=2}^{P} k_p\, \mathscr{C}^{\beta\ \lambda_1\cdots\lambda_{2p}}_{\alpha\ \mu_1\cdots\mu_{2p}} R_{\lambda_1\lambda_2}{}^{\mu_1\mu_2}\cdots R_{\lambda_{2p-1}\lambda_{2p}}{}^{\mu_{2p-1}\mu_{2p}}.$$

We have by the symmetries of the Riemann tensor

$$A_{\alpha\beta} \equiv A_{\beta\alpha}.$$

Equations (2.2) are, in local coordinates, a system of $d(d+1)/2$ second-order partial differential equations for the $d(d+1)/2$ unknowns $g_{\alpha\beta}$.

These equations are invariant by diffeomorphisms of V—i.e., change of local coordinates—and they satisfy the d identities

$$\nabla_\alpha A^{\ \alpha}_{\beta} \equiv 0; \tag{2.3}$$

indeed each coefficient of K_p in (2.1) satisfies such an identity.[2]

III. CAUCHY PROBLEM. CONSTRAINTS

Equations (2.2) are, like the equations of ordinary general relativity, an underdetermined system: their characteristic determinant is identically zero, at most of rank $d(d+1)/2 - d = d(d-1)/2$, due to (2.3). They are also an overdetermined one: the unknowns and their first derivatives cannot be given arbitrarily on a $(d-1)$-dimensional submanifold of V; the Cauchy data must satisfy constraints.

To make a geometric analysis of the Cauchy problem we use, as in ordinary gravity, a $(d-1)+1$ decomposition of the metric and, in the case considered here, of the full Riemann tensor.

Let $U = S\times I$ be a local slicing of an open set $U\subset V$ by $(d-1)$-dimensional spacelike manifolds $S_t = S\times\{t\}$. The metric reads, in adapted coordinates $x^0 = t$, x^i coordinates on S,

$$ds^2 = -\alpha^2(dx^0)^2 + g_{ij}(dx^i + \beta^i dx^0)(dx^j + \beta^j dx^0).$$

If the shift β is zero (a choice that is always possible) a simple calculation gives the identities

$$R_{ijh}{}^k \equiv \overline{R}_{ijh}{}^k + K_j{}^k K_{ih} - K_i{}^k K_{hj}, \tag{3.1a}$$

$$R_{ijh}{}^0 \equiv -(1/\alpha)(\overline{\nabla}_j K_{ih} - \overline{\nabla}_i K_{jh}), \tag{3.1b}$$

 0022-2488/88/081891-05$02.50

$$R_{0h}{}^0 \equiv -(1/\alpha)\partial_0 K_{ih} + K_{ij}K_h{}^j - (1/\alpha)\overline{\nabla}_i\,\partial_h\alpha\,, \quad (3.1c)$$

where $\overline{\nabla}$ and $\overline{R}_{ijh}{}^k$ are the Riemannian covariant derivative and curvature tensor of the metric $\bar{g} = (g_{ij})$ induced on S_t by $g = (g_{\alpha\beta})$, and $K = (K_{ij})$ is the extrinsic curvature of S_t, that is,

$$K_{ij} = -(1/2\alpha)\partial_0 g_{ij}\,. \quad (3.2)$$

Formulas (3.1) and (3.2) can be written in arbitrary coordinates, or in intrinsic notation, by making the following substitution:

$$(g_{ij}) \to \pi g\,, \quad \alpha^{-1}\,\partial_0 \to \pi\mathscr{L}_n\,,$$

where $\mathscr{L}_n$ is the Lie derivative with respect to the unit normal $n = (\alpha^{-1}, -\alpha^{-1}\beta^i)$ to S_t and π the projection operator on S_t, and

$$(R_{ijh}{}^k) \to \pi\,\mathrm{Riem}(g) = (\gamma^{\alpha'}_{\alpha}\gamma^{\beta'}_{\beta}\gamma^{\lambda'}_{\lambda}\gamma^{\mu}_{\mu'}\,R_{\alpha'\beta'\lambda'}{}^{\mu'})\,,$$

$$(\alpha R_{ijh}{}^0) \to \pi n\cdot\mathrm{Riem}(g) = \gamma^{\alpha'}_{\alpha}\gamma^{\beta'}_{\beta}\gamma^{\lambda'}_{\lambda}n_\mu\,R_{\alpha'\beta'\lambda'}{}^{\mu}\,,$$

$$(R_{0j}{}^0) \to n\otimes n_b\cdot\mathrm{Riem}(g) = (n^\beta n_\mu R_{\alpha\beta\lambda}{}^\mu)\,,$$

$$\gamma^{\alpha'}_{\alpha} = \delta^{\alpha'}_{\alpha} + n^{\alpha'}n_\alpha\,.$$

We remark in the formulas (3.1) that the derivative $\partial^2_{00}\alpha$ appears nowhere and that the derivative $\partial_0 K_{ij}$, therefore the derivative $\partial^2_{00}g_{ij}$, appears only in $R_{0h}{}^0$. As a consequence the quantities A^0_0 and A^0_i are determined on a slice S_t by the values on S_t of the first derivatives of the metric; they give constraints on the Cauchy data, namely, in the coordinates we have adopted,

$$A^0_0 \equiv S^0_0 + \Lambda g^0_0 + \chi \sum_{p=2}^{P} k_p\,\mathscr{E}^{i_1\cdots i_{2p}}_{j_1\cdots j_{2p}} \times R_{i_1i_2}{}^{j_1j_2}\cdots R_{i_{2p-1}i_{2p}}{}^{j_{2p-1}j_{2p}}\,,$$

$$A^0_i \equiv S^0_i - \chi \sum_{p=2}^{P} 2pk_p\,\mathscr{E}^{i_1\cdots i_{2p}}_{j_2\cdots j_{2p}} \times R_{i_1i_2}{}^{0j_2}R_{i_3i_4}{}^{j_3j_4}\cdots R_{i_{2p-1}i_{2p}}{}^{j_{2p-1}j_{2p}}\,.$$

We know from Einstein's equations that

$$S^0_0 \equiv -\tfrac{1}{2}(\overline{R} - K_i{}^jK_j{}^i + (K^h_h)^2)\,,$$

$$\alpha S^0_i \equiv \overline{\nabla}_h K_i{}^h - \partial_i K^h_h\,.$$

Using (3.1) the other terms in A^0_0 and αA^0_i can also be expressed in terms only of the geometric elements $\bar{g}_t$ and K_t on S_t.

The intrinsic Cauchy data on a slice S_0 for stringy gravity are, as for usual gravity, a metric and a symmetric two-tensor on S_0, satisfying the constraints,

$$A_{\perp\perp} \equiv A_{\alpha\beta}n^\alpha n^\beta = 0\,, \quad \text{"Hamiltonian" constraint}\,,$$

$$A^{\parallel}_{\perp} = (A_{\perp i}) = \pi^\beta_i n_\alpha A^\alpha{}_\beta = -\alpha A^0_i = 0\,, \quad \text{"momentum" constraint}\,.$$

An analytic solution of the equations $A_{ij} = 0$ on $U = S\times I$ that satisfies the constraints on S_0 satisfies the equations $A_{\alpha\beta} = 0$ in a neighborhood of S_0, for any analytic choice of lapse and shift, due to the identities

$$\nabla_\alpha A^\alpha_\beta = 0\,,$$

which are then a first-order homogeneous system of the Cauchy–Kovalevski type for the d quantities A^0_α.

The same is true if, instead of $A_{ij} = 0$, we consider the equations

$$\widetilde{A}_{ij} \equiv A_{ij} - [1/(d-2)]g_{ij}A^\alpha_\alpha \equiv R_{ij} + [2/(2-d)]\Lambda g_{ij} + B_{ij} - [1/(d-2)]g_{ij}B^\alpha_\alpha = 0\,.$$

IV. EVOLUTION. ANALYTIC CASE

The equations $A_{ij} = 0$, or $\widetilde{A}_{ij} = 0$, are, when the lapse is given as well as the shift (here taken to be zero), a system of nonlinear $d(d-1)/2$ partial differential equatioins for the unknown g_{hk}. As remarked by Aragone, they are linear in the second derivatives $\partial^2_{00}g_{hk}$. They are of the Cauchy–Kovalevski type in a neighborhood in $S\times\mathbb{R}$ of the manifold $S_0 = S\times\{0\}$, for the Cauchy data $\bar{g}$ and K, if they can be solved with respect to the $\partial^2_{00}g_{hk}$, that is, if the determinant of the coefficients of these derivatives is nonzero.

We denote by $\simeq$ equality modulo the addition of terms that contain no second derivatives $\partial^2_{00}g_{hk}$. We have

$$R_{ij} \simeq (1/2\alpha^2)\partial^2_{00}g_{ij}\,,$$

$$B_{ij} \simeq \sum_{p=2}^{P} g_{jl}k_p(2p)^2\,\mathscr{E}^{l\;l_2l_3\cdots l_{2p}}_{i\;m_2m_3\cdots m_{2p}} \times R_{l_20}{}^{m_20}R_{l_3l_4}{}^{m_3m_4}\cdots R_{l_{2p-1}l_{2p}}{}^{m_{2p-1}m_{2p}} \simeq (1/2\alpha^2)X^{hk}_{ij}\,\partial^2_{00}\,g_{hk}$$

with

$$X^{hk}_{ij} = g_{jl}g^{mk}\sum_{p=2}^{P} k_p(2p)^2\,\mathscr{E}^{l\;h\;l_3\cdots l_{2p}}_{i\;m\;m_3\cdots m_{2p}} \times R_{l_3l_4}{}^{m_3m_4}\cdots R_{l_{2p-1}l_{2p}}{}^{m_{2p-1}m_{2p}}\,.$$

The system $\widetilde{A}_{ij} = 0$ is of the Cauchy–Kovalevski type, for the unknown g_{ij} and an arbitrarily given α, if the determinant of the matrix M with elements (a capital index is a pair of ordered indices)

$$M^J_I \equiv M^{hk}_{ij} = (1/2\alpha^2)(\delta^J_I + \chi Y^{hk}_{ij})\,,$$

$$Y^{hk}_{ij} \equiv X^{hk}_{ij} - [1/(d-2)]g_{ij}g^{lm}X^{hk}_{lm}\,,$$

$$g^{lm}X^{hk}_{lm} \equiv g^{mk}\sum_{p=2}^{P} k_p(2p)^2(d-2p)\,\mathscr{E}^{h\;l_3\cdots l_{2p}}_{m\;m_3\cdots m_{2p}} \times R_{l_3l_4}{}^{m_3m_4}\cdots R_{l_{2p-1}l_{2p}}{}^{m_{2p-1}m_{2p}}\,,$$

is nonidentically zero.

We have, $\mathbb{1}$ denoting the unit matrix and $D = d(d-1)/2$,

$$\det M = (1/2\alpha^2)^D\det(\mathbb{1} + \chi Y)$$

and

$$\det(\mathbb{1} + \chi Y) = 1 + \chi a_1 + \chi^2 a_2 + \cdots + \chi^D a_D\,,$$

with

$$a_1 = \delta^I_J Y^J_I = \mathrm{tr}\, Y = \delta^i_h \delta^i_k Y^{hk}_{ij}$$
$$= X^{ij}_{ij} - [1/(d-2)] g_{ij} g^{lm} X^{ij}_{lm}$$
$$= \sum_{p=2}^{P} C_p k_p (2p)^2 \mathscr{E}^{l_3 \cdots l_{2p}}_{m_3 \cdots m_{2p}}$$
$$\times R_{l_3 l_4}{}^{m_3 m_4} \cdots R_{l_{2p-1} l_{2p}}{}^{m_{2p-1} m_{2p}},$$
$$a_2 = \mathscr{E}^{I_1 I_2}_{J_1 J_2} Y^{J_1}_{I_1} Y^{J_2}_{I_2} = (\mathrm{tr}\, Y)^2 - Y^{hk}_{ij} Y^{ij}_{hk},$$
$$\vdots$$
$$a_D = \mathscr{E}^{I_1 \cdots I_D}_{J_1 \cdots J_D} Y^{I_1}_{J_1} \cdots Y^{I_D}_{J_D} = \det Y.$$

The term a_q is a polynomial in the components $R_{ij}{}^{kl}$ of the Riemann tensor of g which can be expressed on S_0, using (3.1), in terms of the Cauchy data $\bar{g}$ and K.

Theorem: Let $(S,\bar{g},K)$ be an analytic initial data set, satisfying the constraints such that $\det(\mathbb{1} + \chi Y)_s \neq 0$. There exists an analytic space-time (V,g) taking these initial data and solution of the equations of stringy gravity.

The (analytic) lapse is arbitrary: to different choices of lapse correspond locally isometric space-times.

Proof: The Cauchy–Kovalevski theorem.

Remark: If Λ, or if K and the Riemann tensor of $\bar{g}$ are small enough, then $\det(\mathbb{1} + \chi Y)_S \neq 0$.

V. CHARACTERISTICS AS POSSIBLE WAVE FRONTS

In the previous section we considered an initial data set $(S,\bar{g},K)$ and showed that it is possible to find an evolution of these initial data (at least in the analytic case) if they are such that $\det(\mathbb{1} + \chi Y)|_S \neq 0$. The lapse and, if we like it, also the shift were arbitrary given functions. By its very construction the metric g is analytic across S_0; in particular its second derivatives admit no discontinuity across S_0.

In order to study possible propagation of stringy gravity, and eventually get rid of the analyticity hypothesis in the solution of the Cauchy problem, we now look for the possible significant discontinuities of the second derivatives of the metric across a $(d-1)$-dimensional submanifold S of a given space-time (V,g) solution of the equations of stringy gravity. Such hypersurfaces are called wave fronts.

We know by general analysis in nonlinear partial differential equations that these hypersurfaces will also be the hypersurfaces of constant phase of the high frequency waves determined by asymptotic expansion.[10]

We know (cf. Lichnerowicz[11]) that the significant discontinuities of the second derivatives of the metric—that is, those that cannot be removed by a C^2 by pieces change of coordinates—are the discontinuities in $\partial^2_{00} g_{ij}$ if S has local equation $x^0 = \mathrm{const}$. The calculations made in the previous section show that these discontinuities can occur across S if and only if

$$\det M|_S = 0. \tag{5.1}$$

We now express this condition in arbitrary coordinates, where the equation of S is $f(x^\alpha) = 0$.

The lapse function α relative to the slicing $S \times \mathbb{R}$ is now such that

$$1/\alpha^2 = -g^{\lambda\mu} \partial_\lambda f \partial_\mu f; \tag{5.2}$$

it is infinite for isotropic (null) slices.

If S is nonisotropic for the metric g, we denote by $\pi^\alpha{}_\beta$ the projection tensor on S.

$$\pi^\alpha{}_\beta = \delta^\alpha_\beta + n^\alpha n_\beta, \quad n^\alpha n_\alpha = -1 \text{ (or } +1). \tag{5.3}$$

In the coordinates of Sec. IV,

$$n^0 = 1/\alpha, \quad n_0 = -\alpha, \quad n_i = n^i = 0, \tag{5.4a}$$
$$\pi^0_0 = \pi^i_0 = \pi^0_i = 0, \quad \pi^i_j = \delta^i_j. \tag{5.4b}$$

The quantities relative to the slicing $S \times \mathbb{R}$, $R_{ij}{}^{kh}$ in the coordinates of Sec. IV, are the components of a tensor ρ with components in arbitrary coordinates, with $f(x^\alpha) = \mathrm{const}$ determining the slicing,

$$\rho_{\alpha\beta}{}^{\lambda\mu} = \pi^{\alpha'}_\alpha \pi^{\beta'}_\beta \pi^\lambda_{\lambda'} \pi^\mu_{\mu'} R_{\alpha'\beta'}{}^{\lambda'\mu'}. \tag{5.5}$$

That is, due to the antisymmetries in the Riemann tensor,

$$\rho_{\alpha\beta}{}^{\lambda\mu} = R_{\alpha\beta}{}^{\lambda\mu} + n^\rho n_{[\alpha} R_{\rho\beta]}{}^{\lambda\mu} + n_\sigma n^\lambda R_{\alpha\beta}{}^{\sigma\mu} + n_\sigma n^\rho$$
$$\times (n_{[\alpha} n^\lambda R_{\rho\beta]}{}^{\sigma\mu} + n_\alpha n^\mu R_{\rho\beta}{}^{\lambda\sigma} + n_\beta n^\lambda R_{\alpha\rho}{}^{\sigma\mu}).$$

Thus

$$\rho_\alpha{}^\lambda = \rho_{\alpha\beta}{}^{\lambda\beta} = \pi^{\alpha'}_\alpha \pi^{\beta'}_\beta \pi^\lambda_{\lambda'} R_{\alpha'\beta'}{}^{\lambda'\mu'}$$
$$= R^\lambda_\alpha + R_\alpha{}^{\lambda'} n^\lambda n_{\lambda'} + R^\lambda_{\alpha'} n^{\alpha'} n_\alpha$$
$$+ R_{\alpha\beta'}{}^{\lambda\mu'} n^{\beta'} n_{\mu'} + R^{\lambda'}_{\alpha'} n^{\alpha'} n_{\lambda'} n_\alpha n^\lambda,$$
$$\rho = \rho^\alpha_\alpha = R + 2n^\alpha n^\lambda R_{\alpha\lambda}.$$

In arbitrary coordinates, and the slicing $f(x^\alpha) = \mathrm{const}$, condition (5.1) can be expressed with the invariant quantities associated with the matrix M. It will read

$$(-g^{\lambda\mu} \partial_\lambda f \partial_\mu f)^D (1 + \chi a_1 + \cdots + \chi^D a_D) \neq 0,$$

where

$$a_1 = \sum_{p=2}^{P} C_p k_p (2p)^2 \mathscr{E}^{\lambda_3 \cdots \lambda_{2p}}_{\mu_3 \cdots \mu_{2p}}$$
$$\times \rho_{\lambda_3 \lambda_4}{}^{\mu_3 \mu_4} \cdots \rho_{\lambda_{2p-1} \lambda_{2p}}{}^{\mu_{2p-1} \mu_{2p}},$$

while a_q is an invariant polynomial in the tensor ρ, of order $(P-1)q$. It can be seen using the expression of ρ and the antisymmetry of the $\mathscr{E}$ tensor that a_q is only a polynomial of order $2q$ in n. We find, for instance, when $P = 2$,

$$a_1 = -64(d-1)(d-3)(d-4)/(d-2) k_2 S^{\alpha\beta} n_\alpha n_\beta$$
$$(\text{using } n^\alpha n_\alpha = -1)$$

(the result found by Aragone[9]) and a_2 is of the form, with C_0, C_1, C_2 numbers depending on d,

$$a_2 = C_0 (\rho^\alpha_\alpha)^2 + C_1 \rho^\lambda_\alpha \rho^\alpha_\lambda + C_2 \rho_{\alpha\beta}{}^{\lambda\mu} \rho^{\alpha\beta}{}_{\lambda\mu},$$

which, using antisymmetries and $n^\alpha n_\alpha = -1$, reduces to a polynomial of degree 4 in n.

We obtain the equation for the wave fronts by replacing $n_\alpha n_\beta$ in a_q by

$$n_\alpha n_\beta = -\frac{\partial_\alpha f \partial_\beta f}{g^{\lambda\mu} \partial_\lambda f \partial_\mu f}$$

and we see that the hypersurface S, $f = \mathrm{const}$, can be a wave front if

$$\Delta \equiv (-g^{\lambda\mu} \partial_\lambda f \partial_\mu f)^D + \chi(-g^{\lambda\mu} \partial_\lambda f \partial_\mu f)^{D-1} b_1(\nabla f)$$
$$+ \chi^2 (-g^{\lambda\mu} \partial_\lambda f \partial_\mu f)^{D-2} b_2(\nabla f)$$

$$+ \cdots + \chi^D b_D(\nabla f) = 0,$$

where $b_q(\nabla f)$ is a homogeneous polynomial of degree $2q$ in ∇f, whose coefficients vanish when the curvature tensor vanishes.

The wave front cone at a point of V is a cone in the cotangent space, obtained by replacing $\partial_\lambda f$ by a covariant vector ξ_λ, of degree $2D$. By taking the parameter χ small—or the curvature small—it is possible to insure that this cone remains in a region close to the null cone of the metric.

Proposition: If the curvature is bounded there exists a number ϵ such that if $|\chi| < \epsilon$ the cone $\Delta = 0$ lies between two second-order cones with equations in an orthonormal frame of the metric g:

$$(X_0)^2 = \lambda_1 \Sigma (X_i)^2, \quad X_0^2 = \lambda_2 \Sigma (X_i)^2, \quad \lambda_1 < 1 < \lambda_2.$$

Proof: In an orthogonal frame the equation of the null cone C_0 of the metric g is

$$-g^{\lambda\mu}\xi_\lambda \xi_\mu \equiv \xi_0^2 - \Sigma \xi_1^2 = 0. \qquad (C_0)$$

Consider a cone

$$\xi_0^2 - \lambda \Sigma \xi_i^2 = 0, \quad \lambda > 0. \qquad (C_\lambda)$$

If $\xi \in C_\lambda$ then

$$\Delta(\xi) \equiv (\lambda - 1)^D (\Sigma \xi_i)^D + \chi P(\lambda(\Sigma \xi_i^2)^{1/2}, \xi_i),$$

with P a homogeneous polynomial in X_i, of degree $2D$, with coefficient polynomials in the curvature.

Fix, for instance, $|\lambda| < 2$; if the curvature is bounded there exists a constant K such that

$$|P(\lambda(\Sigma \xi_i^2)^{1/2}, \xi_i)| < K(\Sigma \xi_i^2)^D;$$

thus

$$((\lambda - 1)^D - \chi K)(\Sigma \xi_i^2)^D < \Delta(\xi) < ((\lambda - 1)^D + \chi K)(\Sigma \xi_i^2)^D.$$

Therefore taking $\lambda = 1 + a$, $a > 0$, and

$$\chi < a^D / K$$

implies $\Delta(\xi) > 0$, *a fortiori* $\Delta(\xi) \neq 0$.

The same result, $\Delta(\xi) \neq 0$, is obtained by taking $\lambda = 1 - b$, $b > 0$, and

$$\chi < b^D / K,$$

which implies $\Delta(\xi) < 0$ if D is odd, and $\Delta(\xi) > 0$ if D is even.

However, there is no reason to consider that the product of the null cone and the cone $g^{\lambda\mu}\xi_\lambda\xi_\mu - \chi b_1(\xi) = 0$ approximates the full cone C.

VI. HARMONIC COORDINATES

Coordinates are harmonic if the metric satisfies the conditions

$$F^\lambda \equiv g^{\alpha\beta}\Gamma_\alpha{}^\lambda{}_\beta = 0. \qquad (6.1)$$

It is well known that

$$R_{\alpha\beta} \equiv -\tfrac{1}{2} g^{\lambda\mu} \partial_\lambda{}^2{}_\mu g_{\alpha\beta} + g_{\alpha\lambda}\partial_\beta F^\lambda + g_{\beta\lambda}\partial_\alpha F^\lambda + h_{\alpha\beta}(g, \partial g), \qquad (6.2)$$

where $H_{\alpha\beta}$ depends only on the metric and its first derivative. We set

$$R_\alpha{}^{(h)}{}_\beta \equiv -\tfrac{1}{2} g^{\lambda\mu} \partial^2_{\lambda\mu} g_{\alpha\beta} + h_{\alpha\beta} \qquad (6.3)$$

and

$$\tilde{A}_\alpha{}^{(h)}{}_\beta \equiv R_\alpha{}^{(h)}{}_\beta + \chi \tilde{B}_{\alpha\beta}, \qquad (6.4)$$

where we do not truncate $B_{\alpha\beta}$ by the use of (6.1).

We deduce from the conservation identities that a solution of $\tilde{A}_\alpha{}^{(h)}{}_\beta = 0$ satisfies the homogeneous wave equations in F^λ,

$$\nabla^\alpha \partial_\alpha F^\lambda = 0. \qquad (6.5)$$

On the other hand a hyperbolic metric $g_{\alpha\beta}$ solution of $\tilde{A}_\alpha{}^{(h)}{}_\beta = 0$ that satisfies the constraints on S_0 ($x^0 = 0$) also satisfies

$$\partial_0 F^\lambda|_{x^0=0} = 0. \qquad (6.6)$$

We deduce from this remark the following proposition.

Proposition: A solution of $A_\alpha{}^{(h)}{}_\beta = 0$ that satisfies the constraints on S_0 and

$$F^\lambda|_{S_0} = 0 \qquad (6.7)$$

satisfies $F^\lambda = 0$ in all the future of S_0, determined by the isotropic cone of the metric, under only mild regularity hypothesis (as necessary for the uniqueness theorem for the wave equations), and hence satisfies $A_{\alpha\beta} = 0$.

The system $\tilde{A}_\alpha{}^{(h)}{}_\beta = 0$ is of the Cauchy–Kovalevski type. Its characteristic determinant is nonzero except on a cone, the characteristic cone. The elements of this determinant are

$$P^{\rho\sigma}_{\alpha\beta} = \frac{\partial \tilde{A}_\alpha{}^{(h)}{}_\beta}{\partial(\partial^2_{\lambda\mu} g_{\rho\sigma})} \xi_\lambda \xi_\lambda = -\frac{1}{2} g^{\lambda\mu}\xi_\lambda\xi_\mu \delta^\rho_\alpha \delta^\sigma_\beta + \chi \frac{\partial \tilde{B}_{\alpha\beta}}{\partial(\partial^2_{\lambda\mu} g_{\rho\sigma})} \xi_\lambda \xi_\mu, \qquad (6.8)$$

where rows and columns are numbered by ordered pairs of indices $(\alpha\beta)$ and $(\rho\sigma)$.

Proposition: The characteristic determinant of the system

$$\tilde{A}_\alpha{}^{(h)}{}_\beta = 0 \qquad (6.9)$$

is, with $C = d^{-d(d+1)/2}$,

$$\det P^{(h)} = C(g^{\lambda\mu}\xi_\lambda\xi_\mu)^d \det M.$$

The elements of the determinant $P^{(h)}$, given by (6.8), are at a point $x \in V$ the components of a mixed tensor, though $\tilde{A}_\alpha{}^{(h)}{}_\beta$ is not a tensor on V. The determinant is a scalar which we can compute in an arbitrary coframe. We choose a coframe such that $\xi_0 = 1$, $\xi_i = 0$, $g^{00} = -\alpha^2$, $g_{0i} = 0$; then

$$\tilde{P}^{\rho}_{\alpha}{}^{\sigma}_{\beta} = \frac{\partial \tilde{A}_\alpha{}^{(h)}{}_\beta}{\partial(\partial^2_{00} g_{\rho\sigma})} = -\frac{1}{2}\delta^\rho_\alpha \delta^\sigma_\beta + \chi \frac{\partial \tilde{B}_{\alpha\beta}}{\partial(\partial^2_{00} g_{\rho\sigma})},$$

we know that the second derivatives $\partial^2_{00} g_{0\sigma}$ do not appear in $\tilde{B}_{\alpha\beta}$, therefore

$$\det P^{(h)} = (-1/2\alpha^2)^{d(d+1)/2} \det(\mathbf{1}_D + \chi Y),$$

where Y is the same $D \times D$ matrix, $D = d(d-1)/2$, as in the previous section, and $\mathbf{1}_D$ is the $D \times D$ matrix.

Writing the stringy gravity in harmonic coordinates as $A_\alpha{}^{(h)}{}_\beta = 0$ introduces the isotropic cone as a spurious wave front cone, but preserves the true one.

Without further information on this cone—reality, simplicity, convexity—it is not easy to give more results on the general Cauchy problem for the classical system of partial differential equations of stringy gravity.

[1]M. Green, J. Schwarz, and E. Witten, *String Theory* (Cambridge U.P., Cambridge, 1987).
[2]D. Lovelock, J. Math. Phys. **12**, 498 (1971).
[3]D. Boulware and S. Deser, Phys. Rev. D **30**, 707 (1984).
[4]B. Zumino, Phys. Rep. **137**, 109 (1986).
[5]Y. Choquet-Bruhat and C. DeWitt, *Analysis Manifolds and Physics*, revised ed. (North-Holland, Amsterdam, 1982), p. 394; M. Dubois-Violette and J. Madore, Orsay preprint, 1986.
[6]D. Boulware and S. Deser, Phys. Lett. B **55**, 2656 (1985).
[7]R. Kerner, C. R. Acad. Sci. **304** (II), 621 (1987); B. Giorgini and R. Kerner, to appear in Class. Quantum Grav.
[8]N. Deruelle (private communication).
[9]C. Aragone, "The characteristics of stringy gravity," in *Silarg 5 Proceedings*, edited by M. Novello (World Scientific, Singapore, in press).
[10]Y. Choquet-Bruhat, Commun. Math. Phys. **12**, 16 (1969).
[11]A. Lichnerowicz, "Théories Relativistes de la Gravitation et de l'Electromagnétisme," Masson, 1955.

The nonlinear desingularization phenomenon in the Abelian Higgs model for isotropic solutions

Y. Y. Chen
Department of Mathematics, California State University, San Bernardino, California 92407-2397

(Received 2 December 1987; accepted for publication 20 April 1988)

In this paper, the occurrence of a linearization phenomenon in the Abelian Higgs model for isotropic solutions will be proved. In this linearizing process, as the parameter $\lambda \to \infty$, the smooth solutions of the nonlinear differential equations will degenerate to the Green's function of the related linearized equation, which is an inhomogeneous linear elliptic equation.

I. INTRODUCTION

The system of equations to be considered in this paper has arisen from the Abelian Higgs model.[1] This system contains a positive parameter λ and a vortex number n which is an integer. We will demonstrate that as λ approaches infinity, the smooth isotropic solutions, with fixed vortex number n and characterized by a variational principle, will degenerate to a Green's function of a related linear problem. Mathematically, such a limiting process, in which the smooth solutions develop a singularity, is named "nonlinear desingularization" by Berger.[2]

In the Abelian Higgs equations, the unknown functions are a complex-valued function Φ (Higgs field) defined on $\mathbb{R}^2$ and a real valued one-form A (Abelian gauge field) defined on $\mathbb{R}^2$. The equations can be written as

$$D_A\,{}^*D_A\Phi = (\lambda/2)V'(|\Phi|)\Phi/|\Phi|\,, \tag{1}$$

$$d\,{}^*\,dA = (i/2)^*(\Phi\,\overline{D_A\Phi} - \overline{\Phi}D_A\Phi)\,, \tag{2}$$

where $i = \sqrt{-1}$, $D_A\Phi = (d - iA)\Phi$ is the covariant derivative of Φ with respect to A, and $V(\text{potential}) \in C^2(\mathbb{R})$ is nonnegative symmetric about the origin and is assumed to possess the following properties: $V(1) = 0$, $V(t) > 0$ on $[0,1)$, and $V''(1) > 0$.[3] [One of the simplest and most popular examples is $V = (1 - \mathbb{R}^2)^2$. The same results presented by this paper are also valid for $V = (1 - \mathbb{R}^2)^3$, which does not satisfy the condition $V''(1) > 0$. The proofs are almost the same and even simpler.]

The Euclidean action associated with the Abelian Higgs model is

$$I_\lambda(\Phi,A) = \frac{1}{2}\int_{\mathbb{R}^2} |dA|^2 + |D_A\Phi|^2 + \lambda V(|\Phi|)\,. \tag{3}$$

The isotropic solutions of (1) and (2) are defined as the smooth finite action solutions that can be written, in terms of polar coordinates (r,θ) on $\mathbb{R}^2\setminus\{0\}$, as

$$\Phi(r,\theta) = R(r)e^{in\theta}\,, \tag{4}$$

$$A(r,\theta) = S(r)d\theta\,, \tag{5}$$

together with the boundary conditions

$$R(0) = S(0) = 0\,, \tag{6}$$

$$R(r) \to 1, \quad S(r) \to n \quad \text{as } r \to \infty\,, \tag{7}$$

where R is a non-negative function and n is a nonzero integer called the vortex number.

Straightforward computation[3] shows that (Φ,A) with form (4) and (5) satisfies (1) and (2) on $\mathbb{R}^2\setminus\{0\}$ if and only if (R,S) satisfies

$$-R''(r) - (1/r)R'(r) + (1/r^2)(n - S(r))^2R(r) + (\lambda/2)V'(R(r)) = 0\,, \tag{8}$$

$$-S''(r) + (1/r)S'(r) - (n - S(r))R^2(r) = 0\,, \tag{9}$$

in $(0,\infty)$. The corresponding Euclidean action is

$$I_\lambda(R,S) = \frac{1}{2}\int_{\mathbb{R}^2}\left[\frac{1}{r}S'(r)\right]^2 + [R'(r)]^2 + \frac{1}{r^2}[n - S(r)]^2R^2(r) + \lambda V(R(r))\,. \tag{10}$$

Remark: From (8) and (9) it is easily seen that if (R,S) is a solution for given n then $(R, -S)$ is a solution for $-n$. Therefore, we only consider the positive integer n in this paper.

The existence of the isotropic solutions has been proved by Plohr.[3] In the previous paper written with Berger[4] we studied the system (1) and (2) with special V, where $V(R) = (1 - R^2)^2$. We proved the existence of the solutions with form (4) and (5) and investigated the nonlinear desingularization phenomenon. For the special V, it was proved that $R(r)$ and $S(r)$ are strictly increasing functions on $[0,\infty)$. However, for the general V considered here, we lose the monotonic property of R that causes all the difficulties in the proof of the occurrence of the nonlinear desingularization.

In Sec. II, we will first set the function spaces which are the same as in the previous paper.[4] We will then discuss the existence and the properties of the smooth isotropic solutions obtained by the variational principle. In Sec. III we will study the nonlinear desingularization phenomenon for smooth isotropic solutions.

II. EXISTENCE, PROPERTIES

The function spaces of R and S are defined by

C_R = the set of real-valued radially symmetric nonnegative functions $R(|x|)$ defined on $\mathbb{R}^2$ such that $1 - R \in W_{1,2}(R^2)$;

C_S = the set of real-valued radially symmetric functions $S(|x|)$ defined on $\mathbb{R}^2$, such that $(1/r)S \in L_2^{\text{loc}}(\mathbb{R}^2)$ with $(1/r)S' \in L_2(\mathbb{R}^2)$, where the derivative S' is in the distributional sense.

 0022-2488/88/081896-05$02.50

We now establish the properties of V and I_λ.

Lemma 1: Suppose that V has the properties stated in the Introduction. Then

(i) $V'(1)=0$;

(ii) there is $\alpha>0$ such that $V(t)\geqslant\alpha(1-t^2)^2$ on $[0,1]$;

(iii) there is $0<\delta<1$ such that

$$V''(t)>0,\quad V'(t)<0\quad \text{on } [\delta,1)\,,$$

$$\inf_{[0,\delta]} V(t) = V(\delta)\,.$$

Proof: (i) Since $V(1)=0$, $V\geqslant 0$ on $\mathbb{R}$ and $V\in C^2(\mathbb{R})$, we have $V'(1)=0$.

(ii) From $V(t)>0$ on $[0,1)$, we have $V(t)/(1-t)^2>0$ on $[0,1)$. According to (i) and $V''(1)>0$,

$$V(t)=\tfrac{1}{2}V''(1)(t-1)^2+\circ(t-1)^2\,;$$

thus $V(t)/(1-t)^2\to\frac{1}{2}V''(1)>0$ as $t\to 1$. Therefore, there is $\alpha>0$ such that

$$V(t)\geqslant\alpha(1-t^2)^2\quad \text{on } [0,1]\,.$$

(iii) $V''(1)>0$ and $V\in C^2$ imply $V''(t)>0$ on $[\tilde\delta,1]$ for an appropriate constant $\tilde\delta>0$; thus $V'(t)<0$ on $[\tilde\delta,1)$ because $V'(1)=0$. Hence V is strictly decreasing on $[\tilde\delta,1]$. Therefore, there exists $\delta\geqslant\tilde\delta$ which satisfies the conclusion.

Lemma 2: For every $(R,S)\in C_R\oplus C_S$, we can define a modified function $\tilde R$ of R such that $\tilde R\in C_R$, $\tilde R\leqslant 1$ on $\mathbb{R}^2$ and $I_\lambda(\tilde R,S)\leqslant I_\lambda(R,S)$.

Proof: Refer to Ref. 3, Lemma II.3.3.

Utilizing Lemmas 1 and 2, we can obtain the existence and some properties of the isotropic solutions by the similar argument used in the previous paper.[4] Here we only state the results without proof.

Theorem 3 (Existence): (i) The infimum of I_λ over $C_R\oplus C_S$ is attained, say at (R,S); (ii) $R(r)$, $S(r)\in C^\infty(0,\infty)\cap C[0,\infty)$ and (R,S) satisfies (6)–(9); (iii) (Φ,A) with form (4) and (5) is a C^2 smooth solution of (1) and (2). [By C^2 smooth we mean that the components of Φ and A, Φ_1, Φ_2, A_1, and A_2 are elements of $C^2(\mathbb{R}^2)$.]

Theorem 4 [Properties of $(R(r),S(r))$]:

(i) $0\leqslant S<n,\quad S'\geqslant 0\quad$ on $[0,\infty)$,

$$\sup_{(0,\infty)}\left|\frac{1}{r}S\right|\leqslant\int_{\mathbb{R}^2}\left(\frac{1}{r}S'\right)^2;$$

(ii) $0\leqslant R\leqslant 1\quad$ on $[0,\infty)$;

(iii) $R\to 1,\quad S\to n$ exponentially as $r\to\infty$;

(iv) $\displaystyle\int_{\mathbb{R}^2}\frac{1}{r}S'=2\pi n,$

where n is the integer in (4).

III. THE NONLINEAR DESINGULARIZATION PHENOMENON FOR ISOTROPIC SOLUTIONS

In this section we will study the asymptotic behavior of the isotropic solutions $\Phi_\lambda=R_\lambda e^{in\theta}$, $A_\lambda=S_\lambda\,d\theta$, as $\lambda\to\infty$. We denote $F_\lambda={}^*dA_\lambda$ (note that it is a gauge invariant), which has representation in $(0,\infty)$:

$$F_\lambda(r)=(1/r)S'_\lambda(r)\,.$$

It can be proved[4] that $F_\lambda\in C^2(\mathbb{R}^2)$ and satisfies

$$-\Delta F_\lambda+F_\lambda=T_\lambda(R_\lambda,S_\lambda)\,, \tag{11}$$

$$F_\lambda\to 0\quad \text{as } |x|\to\infty\,, \tag{12}$$

where $T_\lambda\in C^1(\mathbb{R}^2)$ and

$$T_\lambda=(2/r)R_\lambda R'_\lambda(n-S_\lambda)+(1/r)(1-R^2_\lambda)S'_\lambda\quad \text{in } (0,\infty)\,. \tag{13}$$

Let G be the Green's function of the linear equation

$$-\Delta G+G=\delta\quad (\delta \text{ is the Dirac delta function}),$$

$$G\to 0\quad \text{as } |x|\to\infty\,.$$

We will demonstrate that for fixed n, as $\lambda\to\infty$,

$$F_\lambda\to 2\pi nG\quad \text{in } W_{1,p}(\mathbb{R}^2),\quad 1\leqslant p<2$$

and

$$T_\lambda\to 2\pi n\delta\quad \text{in an integral sense.}$$

We will first establish the exponential decay for T_λ and F_λ.

Lemma 5: For every $\lambda>0$, as $r\to\infty$,

$$T_\lambda(r),F_\lambda(r)\to 0\quad \text{exponentially}\,.$$

Proof: If we can show that $R'_\lambda(r),F_\lambda(r)\in W_{1,2}[1,\infty)$, then we have $R'_\lambda(r),F_\lambda(r)\to 0$ as $r\to\infty$ (cf. Ref. 1, p. 84). Thus according to Theorem 4 (iii) we have $T_\lambda\to 0$ exponentially at ∞, and applying the standard theory of exponential decay at ∞ (cf. Ref. 1, p. 84) to (11), we obtain $F_\lambda\to 0$ exponentially at ∞. In fact, since $(R_\lambda,S_\lambda)\in C_R\oplus C_S$, $R'_\lambda,F_\lambda\in L_2[1,\infty)$. From (8) and (9) R''_λ and F'_λ satisfy

$$R''_\lambda=-(1/r)R'_\lambda+(1/r^2)(n-S_\lambda)^2R_\lambda+(\lambda/2)V'(R_\lambda)\,,$$

$$F'_\lambda=-(1/r)(n-S_\lambda)R^2_\lambda\,.$$

By Theorem 4 (iii), $(1/r^2)(n-S_\lambda)^2R_\lambda$, $(1/r)(n-S_\lambda)R^2_\lambda\in L_2[1,\infty)$. By Lemma 1 and Theorem 4 (ii),

$$\begin{aligned}|V'(R_\lambda)|^2&=|V'(R_\lambda)-V'(1)|^2\\&=|V''(\xi)(R_\lambda-1)|^2\\&\leqslant C(1-R^2_\lambda)^2\leqslant(C/\alpha)V(R_\lambda)\,.\end{aligned}$$

The proof is complete.

We will now investigate the main result of this paper.

Theorem 6:

(i) $\displaystyle\int_{\mathbb{R}^2}T_\lambda=2\pi n,\quad$ for all $\lambda>0$;

(ii) for given $\epsilon>0$, $\displaystyle\lim_{\lambda\to\infty}\int_{|x|>\epsilon}|T_\lambda|=0$;

(iii) $F_\lambda\to 2\pi nG\quad$ in $W_{1,p}(\mathbb{R}^2)$ as $\lambda\to\infty$, where $1\leqslant p<2$.

Proof of (i): From (12), Lemma 5, and Theorem 4 (iv),

$$\int_{\mathbb{R}^2}T_\lambda=\int_{\mathbb{R}^2}F_\lambda-\int_{\mathbb{R}^2}\Delta F_\lambda=2\pi n-\int_{\mathbb{R}^2}\Delta F_\lambda\,.$$

By Theorem 4 (ii) and (iii), $F'_\lambda=-(1/r)(n-S_\lambda)R^2_\lambda$ approaches zero exponentially at infinity; thus the Green's formula gives $\int_{\mathbb{R}^2}\Delta F_\lambda=0$.

Proof of (ii): We will prove that

(I) $\lim_{\lambda\to\infty}\int_{\mathbb{R}^2}\left|\frac{1}{r}(1-R_\lambda^2)S_\lambda'\right|=0$;

(II) $\lim_{\lambda\to\infty}\int_{|x|>\epsilon}\left|\frac{2}{r}R_\lambda R_\lambda'(n-S_\lambda)\right|=0$.

(I): For any $\alpha\neq 0$,

$$I_\lambda(R_\lambda(\alpha r),S_\lambda(\alpha r))$$
$$=\frac{\alpha^2}{2}\int_{\mathbb{R}^2}\left(\frac{1}{r}S_\lambda'\right)^2+\frac{1}{2}\int_{\mathbb{R}^2}(R_\lambda')^2+\frac{1}{r^2}(n-S_\lambda)^2R_\lambda^2+\frac{\lambda}{2\alpha^2}\int_{\mathbb{R}^2}V(R_\lambda).$$

Thus

$$\frac{d}{d\alpha}I_\lambda(R_\lambda(\alpha r),S_\lambda(\alpha r))=\alpha\int_{\mathbb{R}^2}\left(\frac{1}{r}S_\lambda'\right)^2-\frac{\lambda}{\alpha^3}\int_{\mathbb{R}^2}V(R_\lambda).$$

Since $(R_\lambda(\alpha r),S_\lambda(\alpha r))\in C_R\oplus C_S$, the minimality of (R_λ,S_λ) implies

$$\left.\frac{d}{d\alpha}(R_\lambda(\alpha r),S_\lambda(\alpha r))\right|_{\alpha=1}=0,$$

i.e.,

$$\int_{\mathbb{R}^2}\left(\frac{1}{r}S_\lambda'\right)^2=\lambda\int_{\mathbb{R}^2}V(R_\lambda). \tag{14}$$

On the other hand, by Lemma 5 and (9) we can use integration by parts to obtain

$$\int_{\mathbb{R}^2}\left(\frac{1}{r}S_\lambda'\right)^2=-2\pi\int_0^\infty\left(\frac{1}{r}S_\lambda'\right)\left(\frac{1}{r}S_\lambda'\right)'r^2\,dr=-\pi\int_0^\infty[(n-S_\lambda)^2]'R_\lambda^2\,dr,$$

and according to Theorem 4, we obtain

$$\int_{\mathbb{R}^2}\left(\frac{1}{r}S_\lambda'\right)^2\leqslant-\pi(n-S_\lambda(r))^2|_0^\infty=\pi n^2. \tag{15}$$

Combining (14) and (15), we have

$$\lim_{\lambda\to\infty}\int_{\mathbb{R}^2}V(R_\lambda)=0. \tag{16}$$

Now Lemma 1 (ii) and Theorem 4 (ii) imply

$$\lim_{\lambda\to\infty}\int_{\mathbb{R}^2}\left|\frac{1}{r}(1-R_\lambda^2)S_\lambda'\right|<\lim_{\lambda\to\infty}\left[\int_{\mathbb{R}^2}\left(\frac{1}{r}S_\lambda'\right)^2\right]^{1/2}\left[\int_{\mathbb{R}^2}(1-R_\lambda^2)^2\right]^{1/2}\leqslant\frac{\sqrt{\pi}n}{\sqrt{\alpha}}\lim_{\lambda\to\infty}\left[\int_{\mathbb{R}^2}V(R_\lambda)\right]^{1/2}=0.$$

(II): The proof rests on the following two lemmas concerning the limiting behavior of R_λ.

Lemma 7: For given $\epsilon>0$, $R_\lambda(r)\to 1$ uniformly on $[\epsilon,\infty)$ as $\lambda\to\infty$.

Proof: It is sufficient to prove that for given $\epsilon>0$ and $\delta<\gamma<1$, where δ is the scalar obtained in Lemma 1, there exists $\Lambda>0$ such that $\lambda\geqslant\Lambda$ implies $R_\lambda(r)\geqslant\gamma$ on $[\epsilon,\infty)$.

We first show that there is $\Lambda_1>0$ such that

$$\sup_{[\epsilon/2,\epsilon]}R_\lambda(r)\geqslant\gamma\quad\text{whenever }\lambda\geqslant\Lambda_1.$$

In fact, if for any $\Lambda_1>0$ there is $\lambda\geqslant\Lambda_1$ such that $\sup_{[\epsilon/2,\epsilon]}R_\lambda(r)<\gamma$, then Lemma 1 and Theorem 4 imply

$$\int_{\mathbb{R}^2}V(R_\lambda)\geqslant\int_{\epsilon/2<|x|<\epsilon}\alpha(1-R_\lambda^2)^2>\alpha(1-\gamma^2)^2\frac{3\pi\epsilon^2}{4},$$

which contradicts (16).

Let $m=\inf_{[\delta,\gamma]}|V'(t)|>0$. Let

$$\Lambda=\max\left\{\Lambda_1,\frac{8n^2}{m\epsilon^2},\frac{4n^2}{[V(\delta)-V(\gamma)]\epsilon^2}\right\}.$$

We will now prove that $R_\lambda\geqslant\gamma$ on $[\epsilon,\infty)$ for all $\lambda\geqslant\Lambda$. We can argue by contradiction. Suppose there is both $\lambda\geqslant\Lambda$ and $r_1\in[\epsilon,\infty)$ such that $R_\lambda(r_1)<\gamma$. Since $\Lambda\geqslant\Lambda_1$ and $R_\lambda\in C^\infty(0,\infty)$, there is $r_0\in[\epsilon/2,r_1)$ such that $R_\lambda(r_0)=\gamma$. We define a modified function $\widetilde{R}_\lambda$ of R_λ by

$$\widetilde{R}_\lambda(r)=\begin{cases}R_\lambda(r), & r\in[0,r_0],\\ \max\{\gamma,R_\lambda(r)\}, & r\in(r_0,\infty).\end{cases}$$

We claim $\widetilde{R}_\lambda\in C_R$. We only need to check $\widetilde{R}_\lambda'\in L_2(\mathbb{R}^2)$. Clearly, $\widetilde{R}_\lambda'=R_\lambda'$ if $r\in[0,r_0]$, $\widetilde{R}_\lambda'=[\max\{\gamma,R_\lambda(r)\}]'$ if $r\in(r_0,\infty)$. Consider $\max\{\gamma,R_\lambda(r)\}$ as a composite function $f\circ R_\lambda$, where f is defined by $f(x)=x$ if $x\geqslant\gamma$, $f(x)=\gamma$ if $x<\gamma$. Thus f is piecewise smooth on $\mathbb{R}$ with $f'\in L_\infty(\mathbb{R})$. Applying the chain rule[5] to $f\circ R_\lambda$, we obtain $|(f\circ R_\lambda)'|\leqslant|R_\lambda'|$. Therefore $\widetilde{R}_\lambda'\in L_2(\mathbb{R}^2)$. Since $(\widetilde{R}_\lambda,S_\lambda)\in C_R\oplus C_S$, $I_\lambda(\widetilde{R}_\lambda,S_\lambda)\geqslant I_\lambda(R_\lambda,S_\lambda)$. Denote

$$E_1=\{r\in(r_0,\infty),\ \delta\leqslant R_\lambda(r)<\gamma\},$$
$$E_2=\{r\in(r_0,\infty),\ R_\lambda(r)<\delta\}.$$

We now prove that on $E_1\cup E_2$,

$$[(n-S_\lambda)^2/r^2](R_\lambda^2-\widetilde{R}_\lambda^2)+\lambda(V(R_\lambda)-V(\widetilde{R}_\lambda))>0.$$

Since E_1 is nonempty, the continuity of R_λ and S_λ implies $I_\lambda(\widetilde{R}_\lambda,S_\lambda)<I_\lambda(R_\lambda,S_\lambda)$. This contradiction concludes the proof. Indeed, for any $r\in E_1$, since $\delta\leqslant R_\lambda(r)<\gamma=\widetilde{R}_\lambda(r)<1$, $\lambda\geqslant\Lambda$ and $0\leqslant S_\lambda<n$,

$$\lambda[V(R_\lambda(r))-V(\widetilde{R}_\lambda(r))]=\lambda V'(\xi)(R_\lambda(r)-\widetilde{R}_\lambda(r))\geqslant\lambda m(\widetilde{R}_\lambda(r)-R_\lambda(r))$$
$$>\frac{\lambda}{2}m(\widetilde{R}_\lambda^2(r)-R_\lambda^2(r))\geqslant\frac{n^2}{(\epsilon/2)^2}(\widetilde{R}_\lambda^2(r)-R_\lambda^2(r))\geqslant\frac{(n-S_\lambda(r))^2}{r^2}(\widetilde{R}_\lambda^2(r)-R_\lambda^2(r)).$$

For any $r\in E_2$, Lemma 1 (iii) implies

$$\lambda[V(R_\lambda(r))-V(\widetilde{R}_\lambda(r))]\geqslant\lambda[V(\delta)-V(\widetilde{R}_\lambda(r))]=\lambda(V(\delta)-V(\gamma)).$$

Because $\lambda\geqslant\Lambda$ and $0\leqslant\widetilde{R}_\lambda^2(r)-R_\lambda^2(r)<1$,

$$\lambda[V(R_\lambda(r))-V(\widetilde{R}_\lambda(r))]\geqslant\frac{(n-S_\lambda(r))^2}{r^2}(\widetilde{R}_\lambda^2(r)-R_\lambda^2(r)).$$

The proof of Lemma 7 is complete.

Lemma 8: For given $\epsilon > 0$ and $\delta < \gamma < 1$, there exists $\Lambda > 0$ such that $\lambda \geqslant \Lambda$ implies $R'_\lambda \geqslant 0$ on $[\epsilon, \infty)$.

Proof: According to Lemma 7, for given $\epsilon > 0$ and $\delta < \gamma < 1$, we can find a $\Lambda > 0$ such that $\lambda \geqslant \Lambda$ implies $R_\lambda \geqslant \gamma$ on $[\epsilon, \infty)$. We will now prove that if $R_\lambda \geqslant \gamma$ on $[\epsilon, \infty)$ then $R'_\Lambda \geqslant 0$ on $[\epsilon, \infty)$, which confirms the result. We can again argue by contradiction. Assume that there is both $\lambda \geqslant \Lambda$ and $r_1 \in [\epsilon, \infty)$, such that $R'_\lambda(r_1) < 0$. We denote $D_1 = \{r \in [0, r_1],\ R'_\lambda(r) = 0, \text{ and } R''_\lambda(r) \leqslant 0\}$. Since $R_\lambda(0) = 0$, $R_\lambda(r) \geqslant 0$ on $[0, \infty)$, $R_\lambda(r) \in C^2[0, \infty)$, and $R'_\lambda(r_1) < 0$, D_1 is a nonempty bounded closed set; thus D_1 attains its smallest upper bound, say at r_0. Clearly, $0 < r_0 < r_1$, $R'_\lambda(r_0) = 0$, $R''_\lambda(r_0) \leqslant 0$, and $R_\lambda(r) \geqslant R_\lambda(r_1) \geqslant \gamma$ on $[r_0, r_1]$. We define

$$\widetilde{R}_\lambda(r) = \begin{cases} R_\lambda(r), & r \in [0, r_0], \\ \max\{R_\lambda(r_0), R_\lambda(r)\}, & r \in (r_0, \infty). \end{cases}$$

The same argument in Lemma 7 gives $\widetilde{R}_\lambda \in C_R$. We denote $D_2 = \{r \in (r_0, \infty),\ R_\lambda(r) < \widetilde{R}_\lambda(r) = R_\lambda(r_0)\}$. We claim that for $r \in D_2$,

$$(1r^2)(n - S_\lambda)^2 (R_\lambda^2 - \widetilde{R}_\lambda^2) + \lambda\,[V(R_\lambda) - V(\widetilde{R}_\lambda)] > 0,$$

or equivalently,

$$\frac{1}{r^2}(n - S_\lambda)^2 < -\lambda \frac{V(R_\lambda) - V(\widetilde{R}_\lambda)}{R_\lambda^2 - \widetilde{R}_\lambda^2}. \tag{17}$$

Thus the continuity of R_λ, $\widetilde{R}_\lambda$, and S_λ imply $I_\lambda(\widetilde{R}_\lambda, S_\lambda) < I_\lambda(R_\lambda, S_\lambda)$, which will contradict the minimality of I_λ at (R_λ, S_λ). Indeed, since (R_λ, S_λ) satisfies (8),

$$(1/r_0^2)(n - S_\lambda(r_0))^2 R_\lambda(r_0) + (\lambda/2) V'(R_\lambda(r_0))$$
$$= R''_\lambda(r_0) + (1/r_0) R'_\lambda(r_0) \leqslant 0. \tag{18}$$

For $r \in D_2$, since $\gamma \leqslant R_\lambda(r) < \widetilde{R}_\lambda(r) = R_\lambda(r_0)$, Lemma 1 implies

$$-\lambda \frac{V(R_\lambda(r)) - V(\widetilde{R}_\lambda(r))}{R_\lambda^2(r) - \widetilde{R}_\lambda^2(r)} = -\lambda \frac{V'(\xi)}{R_\lambda(r) + \widetilde{R}_\lambda(r)} \geqslant -\frac{\lambda V'(R_\lambda(r_0))}{2R_\lambda(r_0)}.$$

Moreover, according to Lemma 4 (i) and (18), we have

$$-\frac{\lambda V'(R_\lambda(r_0))}{2R_\lambda(r_0)} \geqslant \frac{1}{r_0^2}(n - S_\lambda(r_0))^2 \geqslant \frac{1}{r^2}(n - S_\lambda(r))^2.$$

Inequality (17) is achieved. The proof of Lemma 8 is complete.

Now we are in the position to prove (II).

Using Lemma 8 and Lemma 7, for $\lambda \geqslant \Lambda$,

$$\int_{|x| > \epsilon} \left| \frac{2}{r} R_\lambda R'_\lambda (n - S_\lambda) \right| = \int_{|x| > \epsilon} \frac{2}{r} R_\lambda R'_\lambda (n - S_\lambda)$$
$$\leqslant 2\pi n R_\lambda^2(r)\big|_\epsilon^\infty = 2\pi n (1 - R_\lambda^2(\epsilon)) \to 0 \quad \text{as } \lambda \to \infty.$$

Proof of (iii): Since F_λ satisfies (11) and (12),

$$F_\lambda(x) = \int_{\mathbf{R}^2} G(x - y) T_\lambda(y) dy.$$

Since $\int_{\mathbf{R}^2} T_\lambda(y) dy = 2\pi n$,

$$F_\lambda(x) - 2\pi n G(x) = \int_{\mathbf{R}^2} (G(x - y) - G(x)) T_\lambda(y) dy.$$

Thus

$$\|F_\lambda(x) - 2\pi n G(x)\|_{W_{1,p}(\mathbf{R}^2)}$$
$$\leqslant \left\{ \int_{\mathbf{R}^2} \left| \int_{\mathbf{R}^2} (G(x - y) - G(x)) T_\lambda(y) dy \right|^p dx \right\}^{1/p} + \sum_{i=1}^2 \left\{ \int_{\mathbf{R}^2} \left| \int_{\mathbf{R}^2} \left[\frac{\partial}{\partial x_i} (G(x - y) - G(x)) \right] \times T_\lambda(y) dy \right|^p dx \right\}^{1/p}$$
$$\leqslant \int_{\mathbf{R}^2} |T_\lambda(Y)| \left\{ \|G(x - y) - G(x)\|_{L_p(\mathbf{R}^2)} + \sum_{i=1}^2 \left\| \frac{\partial}{\partial x_i} (G(x - y) - G(x)) \right\|_{L_p(\mathbf{R}^2)} \right\} dy$$
$$= \int_{\mathbf{R}^2} |T_\lambda(Y)| \widetilde{G}(y) dy,$$

where

$$\widetilde{G}(y) = \|G(x - y) - G(x)\|_{L_p(\mathbf{R}^2)} + \sum_{i=1}^2 \left\| \frac{\partial}{\partial x_i} (G(x - y) - G(x)) \right\|_{L_p(\mathbf{R}^2)}.$$

In Ref. 4 we established the L_p estimates for $\widetilde{G}(y)$. We obtained that for $1 \leqslant p < 2$,

(a) $0 \leqslant \widetilde{G}(y) < C$ on $\mathbf{R}^2$, where C is a constant;
(b) $0 \leqslant \widetilde{G}(y) \leqslant C|Y|^\alpha$ on $|y| \leqslant 1$,

where C and α are positive constants.

Combining (I) and (a) we obtain

$$\lim_{\lambda \to \infty} \int_{\mathbf{R}^2} \left| \frac{1}{r} (1 - R_\lambda^2) S'_\lambda \right| \widetilde{G}(y) dy = 0.$$

Combining (II) and (a) we obtain

$$\lim_{\lambda \to \infty} \int_{|y| > \epsilon} \left| \frac{2}{r} R_\lambda R'_\lambda (n - S_\lambda) \right| \widetilde{G}(y) dy = 0.$$

To conclude the proof we show that there are constants $C > 0$ and $\Lambda > 0$, such that for all $0 < \epsilon \leqslant 1$ and all $\lambda \geqslant \Lambda$, we have

$$\int_{|y| < \epsilon} \left| \frac{2}{r} R_\lambda R'_\lambda (n - S_\lambda) \right| \widetilde{G}(y) dy \leqslant C|\epsilon|^{\alpha/2}.$$

According to (b) it is sufficient to show that there exists a constant $C > 0$, such that for all $\lambda \geqslant \Lambda$,

$$|R'_\lambda(r)| \leqslant C/r, \quad \text{on } [0,1]. \tag{19}$$

In fact, using the same method introduced by Plohr[3] (Chap. II, Sec. 4), and noticing that the S in Plohr's paper is $-S$ in this paper, we obtain the representation of R'_λ on $[0,1]$:

$$R'_\lambda(r_0) = \frac{1}{2r_0}\int_0^{r_0}\left[\frac{1}{r^2}S_\lambda(r)(S_\lambda(r)-2n)R_\lambda(r) + V'(R_\lambda(r))\right]\left[\frac{r}{r_0}\right]^n r\,dr - \frac{1}{2r_0}\int_{r_0}^{3}\left[\frac{1}{r^2}S_\lambda(r)(S_\lambda(r)-2n)R_\lambda(r) + V'(R_\lambda(r))\right]\left[\frac{r_0}{r}\right]^n \chi(r) r\,dr - \frac{n}{2r_0}\int_2^3\left[R_\lambda(r) + \frac{rR'_\lambda(r)}{n}\right]\left[\frac{r_0}{r}\right]^n \chi'(r)dr$$

where χ is a C^∞ function on $[0,\infty)$, such that $\chi(r)=1$ for $r\leqslant 2$, $\chi(r)=0$ for $r>3$, and $0\leqslant\chi(r)\leqslant 1$ on $[0,\infty)$. Since $0\leqslant R_\lambda(r)\leqslant 1$ and $V\in C^2(-\infty,\infty)$, we have

$$\left|\frac{1}{2r_0}\left\{\int_0^{r_0} V'(R_\lambda(r))\left[\frac{r}{r_0}\right]^n r\,dr - \int_{r_0}^3 V'(R_\lambda(r))\left[\frac{r_0}{r}\right]^n \chi(r) r\,dr\right\}\right| \leqslant \frac{1}{2r_0}\sup_{[0,1]}|V'(t)|\left[\frac{r_0^2}{2} + \frac{9-r_0^2}{2}\right] \leqslant \frac{C}{r_0}, \tag{20}$$

where C is a constant independent of r_0. Since Theorem 4 and inequality (15) imply

$$\sup_{(0,\infty)}|(1/r)S_\lambda(r)| \leqslant \pi n^2,$$

therefore,

$$\left|\frac{1}{2r_0}\left\{\int_0^{r_0}\frac{1}{r}S_\lambda(r)(S_\lambda(r)-2n)R_\lambda(r)\left[\frac{r}{r_0}\right]^n r\,dr - \int_{r_0}^3\frac{1}{r}S_\lambda(r)(S_\lambda(r)-2n)R_\lambda(r)\left[\frac{r_0}{r}\right]^n\chi(r) r\,dr\right\}\right| \leqslant \frac{9\pi n^3}{2r_0}, \tag{21}$$

because $0\leqslant S_\lambda(r)<n$, $0\leqslant R_\lambda(r)\leqslant 1$, and $0\leqslant\chi(r)\leqslant 1$. According to Lemma 8, there is $\Lambda>0$, such that $\lambda\geqslant\Lambda$ implies $R'_\lambda\geqslant 0$ on $[2,\infty)$, therefore,

$$\left|\frac{n}{2r_0}\int_2^3\left[R_\lambda(r) + \frac{rR'_\lambda(r)}{n}\right]\left[\frac{r_0}{r}\right]^n\chi'(r)dr\right| \leqslant \frac{n}{2r_0}\sup_{[2,3]}|\chi'(r)|\left[1 + \frac{3}{n}\int_2^3 R'_\lambda(r)dr\right] \leqslant \frac{c}{r_0}. \tag{22}$$

Combining (20)–(22) we obtain (19).

ACKNOWLEDGMENT

I would like to thank M. S. Berger for suggesting the problem and the paths to their solutions.

[1]A. Jaffe and C. Taubes, *Vortices and Monopoles* (Birkhäuser, Boston, 1980).
[2]M. S. Berger and L. E. Fraenkel, Commun. Math. Phys. 77, 149 (1980).
[3]B. J. Plohr, Ph.D. thesis, Princeton University, 1980.
[4]M. S. Berger and Y. Y. Chen, "Symmetric vortices for the Ginzburg–Landau equations and the nonlinear desingularization phenomenon," to be published in J. Funct. Anal.
[5]S. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, Berlin, 1983), 2nd ed.

One loop effective measure for sigma models coupled with gravity

E. Guadagnini
Dipartimento di Fisica dell'Università di Pisa, Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Pisa, Italy

(Received 8 December 1987; accepted for publication 13 April 1988)

The one loop effective measure for a two-dimensional generalized sigma model with torsion in the presence of an arbitrary gravitational background is computed. The consequences of the result on the relation between the conformal anomalies and the divergences of the model are considered.

I. INTRODUCTION

Generalized sigma models in two dimensions have recently received much attention because of their relation with string theory.[1] The requirement of conformal invariance of the sigma model is connected with the low energy equations of motion of the string.[1,2] In practice, one is interested in the transformation properties of the partition function (or vacuum energy) of the model under Diff and Weyl transformations in the presence of a gravitational background.

For a generic interacting sigma model, i.e., for a model defined on a nonflat target space (eventually also with torsion), the best one can do is to use perturbation theory, that is, to make a loop expansion. Even at the one loop level, a *complete* computation of the conformal anomalies of an interacting sigma model in the presence of an arbitrary gravitational background has not been done. In performing such a computation one meets essentially two kinds of difficulties. First, the Lagrangian contains infinitely many interaction vertices with two space-time derivatives and, second, some care is needed in the choice of the regularization that might introduce an explicit breaking of the conformal symmetries in which one is interested.

I have shown recently[3] how the above problems can be bypassed by using the Schwinger proper-time regulation.[4] In Ref. 3 a free bosonic model in the presence of an arbitrary gravitational background has been considered. In the present paper the same method is applied to study the Diff and Weyl anomalies of an interacting sigma model at the one loop level.

The computational work consists of evaluating the one loop effective measure $\mathscr{M}_{ij}(x)$ of the theory

$$\mathscr{M}_{ij}(x) = \langle x,i|e^{-\epsilon H(p,q)}|x,j\rangle\,|_{\epsilon\to 0}, \tag{1.1}$$

where $H(p,q)$ is just the operator entering in the quadratic part of the Lagrangian in background and the indices i and j label the internal degrees of freedom.

The one loop effective measure (1.1) is the fundamental quantity that enters in all the one loop computations of any kind of anomaly. In practice, anyone who has computed a one loop anomaly has, in one way or another, computed $\mathscr{M}_{ij}(x)$ or part of it. The one loop effective measure contains all the information concerning the symmetries of the theory and describes also the one loop divergences that, in fact, are related in certain models to the breaking of the dilatation symmetry.

In the present paper I will concentrate on the computation of $\mathscr{M}_{ij}(x)$ with the more general form of $H(p,q)$ relevant for a two-dimensional bosonic sigma model in an arbitrary gravitational background. I consider then the relation between the conformal anomalies and the divergences of the model at one loop. The implications of the result on the low energy physics of the string will be discussed elsewhere.[5]

II. THE ONE LOOP EFFECTIVE MEASURE

In this section we fix the notation and formulate the problem. The material contained in the present section represents a short review of some well-known aspects of the field theory. Our main purpose will be to show the crucial role played by the effective measure (1.1) in the anomaly computations.

We are interested in a two-dimensional generalized sigma model in presence of an arbitrary gravitational background. A typical action is

$$S=\frac{1}{2}\int d^2x(\sqrt{g}g^{\mu\nu}G_{ij}(\phi)\partial_\mu\phi^i\partial_\nu\phi^j - i\epsilon^{\mu\nu}K_{ij}(\phi)\partial_\mu\phi^i\partial_\nu\phi^j), \tag{2.1}$$

where $g^{\mu\nu}(x)$ and $g(x)$ are the inverse and the determinant of the classical two-dimensional metric $g_{\mu\nu}(x)$ (with Euclidean signature). The interactions of the scalar field $\phi^i(x)$ ($i=1,2,...,D$) with itself are described by the two functions $G_{ij}(\phi)=G_{ji}(\phi)$ and $K_{ij}(\phi) = -K_{ji}(\phi)$.

At the tree level the symmetries of the theory are those displayed in the classical action (2.1), whereas at one loop one has to consider the effective action $\Gamma[\phi]$,

$$\Gamma[\phi] = \Gamma_0[\phi] + \hbar\Gamma_1[\phi] + \cdots . \tag{2.2}$$

As is well known, $\Gamma_0[\phi]$ is just the classical action and $\Gamma_1[\phi]$ can be obtained from the knowledge of all the one particle irreducible diagrams at one loop with dropped external propagators. The easiest way of computing $\Gamma_1[\phi]$ is to use the so-called background field method. In fact, this method can be used to find $\Gamma[\phi]$ at every order of the loop expansion. The recipe is the following.

(i) Make the shift $\phi^i(x)\to\phi^i(x)+\pi^i(x)$ in the classical action. (Note that this is *not* a normal coordinate expansion.)

(ii) Consider now $S[\phi^i+\pi^i]$ as the action for the "quantum" field π^i, and consider $\phi^i(x)$ as *arbitrary* classical functions.

(iii) Compute the connected, one particle irreducible vacuum-to-vacuum diagrams for the π field. The result is precisely $\Gamma[\phi]$.

 0022-2488/88/081901-08$02.50

A proof that the above procedure gives the right result can be found in Ref. 6. If you like a more direct proof in terms of single Feynman diagrams you can easily verify that the statistics of the diagrams is indeed correct.

So, let us compute now $\Gamma_1[\phi]$. According to the above rules, Γ_1 is just given by one-half of the logarithm of the determinant of the operator H entering in the quadratic part of the action $S[\phi+\pi]$ in the π field. That is, in short-hand notation,

$$\frac{1}{2}\ln \operatorname{Det} H = \frac{1}{2}\ln \operatorname{Det}\frac{\delta^2 S[\phi]}{\delta\phi^i\,\delta\phi^j}. \tag{2.3}$$

In order to give a precise meaning to the above expression, one has to introduce a regularization. We use the Schwinger's proper-time method,[4] and define the regularized one loop effective action Γ_1 as

$$\Gamma_1 = -\frac{1}{2}\int_\epsilon^\infty \frac{d\tau}{\tau}\operatorname{Tr}[e^{-\tau H}], \tag{2.4}$$

where ϵ is the ultraviolet cutoff (to be sent to zero) and Tr means the trace operation in both the internal indices i and j $(i,j=1,2,\dots,D)$ and the space-time coordinates.

With the action (2.1), the H operator is

$$H_{ij}(p,q) = p_\mu \hat g^{\mu\nu}(q)A_{ij}(q)p_\nu + i[p_\mu B^\mu_{ij}(q) + B^\mu_{ij}(q)p_\mu] + C_{ij}(q), \tag{2.5}$$

where

$$p_\mu \equiv -i\frac{\partial}{\partial q^\mu}, \tag{2.6}$$

$$\hat g^{\mu\nu}(q) = \sqrt{g(q)}\,g^{\mu\nu}(q), \tag{2.7}$$

and

$$A_{ij}(q) = A_{ji}(q) = G_{ij}(\phi(q)), \tag{2.8}$$

$$B^\mu_{ij} = -B^\mu_{ji} = \tfrac{1}{2}\hat g^{\mu\nu}(\partial_i G_{jk} - \partial_j G_{ik})\partial_\nu\phi^k - i\epsilon^{\mu\nu}T_{kij}\,\partial_\nu\phi^k, \tag{2.9}$$

$$C_{ij} = C_{ji} = \tfrac{1}{2}\hat g^{\mu\nu}\,\partial_i\partial_j G_{kl}\,\partial_\mu\phi^k\,\partial_\nu\phi^l - \tfrac{1}{2}\partial_\mu[\hat g^{\mu\nu}(\partial_i G_{jk} + \partial_j G_{ik})\partial_\nu\phi^k] - (i/2)\epsilon^{\mu\nu}(\partial_i T_{klj} + \partial_j T_{kli})\partial_\mu\phi^k\,\partial_\nu\phi^l, \tag{2.10}$$

$$T_{ijk} = \tfrac{1}{2}[\partial_i K_{jk} + \partial_k K_{ij} + \partial_j K_{ki}]. \tag{2.11}$$

There is an important point to be noted here concerning the dependence of Γ on the gravitational background. The action (2.1) does not depend on all the components of the classical metric $g_{\mu\nu}(x)$. The action only depends on $\hat g^{\mu\nu} = \sqrt{g}g^{\mu\nu}$. If the metric $g_{\mu\nu}$ is not involved in the regularization procedure (like in the Schwinger proper-time method), then Γ also is a functional of $\hat g^{\mu\nu}$. Consider now an infinitesimal Weyl transformation

$$\Delta_\sigma g_{\mu\nu}(x) = 2\sigma(x)g_{\mu\nu}(x). \tag{2.12}$$

Since $\hat g^{\mu\nu}(x)$ is invariant under a transformation (2.12), it follows that the regularized effective action Γ is invariant under Weyl transformations.

Let us consider now an infinitesimal Diff transformation

$$\Delta_V g^{\mu\nu} = \nabla_\mu V_\nu(x) + \nabla_\nu V_\mu(x), \tag{2.13}$$

$$\Delta_V\phi^i(x) = V^\mu(x)\partial_\mu\phi^i(x), \tag{2.14}$$

where the covariant derivative ∇_μ is defined in terms of the Christoffel connection

$$\Gamma^\mu_{\nu\rho} = \tfrac{1}{2}g^{\mu\sigma}(\partial_\nu g_{\sigma\rho} + \partial_\rho g_{\sigma\nu} - \partial_\sigma g_{\nu\rho}). \tag{2.15}$$

The classical action (2.1) is invariant under the transformations (2.13) and (2.14), but what about the one loop effective action Γ_1? All we need to know is the transformation property of the H operator (2.5) under the transformations (2.13) and (2.14). Since H is related to the second derivative of the classical action (which is invariant) with respect to ϕ^i, H transforms covariantly with respect to the transformation (2.14); that is, H must transform as the product

$$\frac{\delta}{\delta\phi^i}\otimes\frac{\delta}{\delta\phi^j}$$

transforms. This means, in particular, that the functional

$$\mathscr{F} = \int d^2x\,\phi^i(x)H_{ij}\phi^j(x)$$

is invariant under Diff transformations. Therefore,

$$\Delta_V\mathscr{F} = \int d^2x[V^\nu(\partial_\nu\phi^i)H_{ij}\phi^j + \phi^i(\Delta_V H_{ij})\phi^j + \phi^i H_{ij}V^\nu(\partial_\nu\phi^j)] = \int d^2x\phi^i[\Delta_V H_{ij} + H_{ij}V^\nu\partial_\nu - \partial_\nu V^\nu H_{ij}]\phi^j$$

must vanish, and hence

$$\Delta_V H_{ij}(p,q) = ip_\nu V^\nu(q)H_{ij}(p,q) - iH_{ij}(p,q)V^\nu(q)p_\nu. \tag{2.16}$$

A direct computation, done by making use of the explicit dependence of H on $\phi^i(x)$ and $g_{\mu\nu}(x)$, shows that Eq. (2.16) is indeed satisfied.

Note that the covariant properties of H with respect to any transformation that is a symmetry of the classical action is a general feature which does not depend on the particular form of the action nor on the fact that we are in two dimensions.

From definition (2.4) and Eq. (2.16) it follows that

$$\Delta_V\Gamma_1 = \frac{1}{2}\operatorname{Tr}[\partial_\lambda V^\lambda e^{-\epsilon H}] = \frac{1}{2}\int d^2x\,\partial_\lambda V^\lambda(x)\operatorname{tr}\langle x|e^{-\epsilon H}|x\rangle, \tag{2.17}$$

where tr means the trace in the target space indices $i,j=1,2,\dots,D$. In terms of the effective measure (1.1), one has

$$\Delta_V\Gamma_1 = \frac{1}{2}\int d^2x\,\partial_\lambda V^\lambda(x)\operatorname{tr}\mathscr{M}(x). \tag{2.18}$$

Equation (2.18) shows how the effective measure controls the Diff anomaly.

To find the one loop divergences of the model is even simpler; a logarithmic derivative of Γ in the cutoff gives

$$\epsilon \frac{d\Gamma_1}{d\epsilon} = \frac{1}{2}\,\mathrm{Tr}[e^{-\epsilon H}] = \frac{1}{2}\int d^2x\,\mathrm{tr}\,\mathcal{M}(x). \quad (2.19)$$

Again, it is the effective measure that gives the one loop divergences.

It is clear that the behavior of Γ_1 under an arbitrary symmetry transformation is always expressed in terms of the effective measure. We leave the proof to the reader; the crucial property that one has to note is the covariant behavior of the H operator under any symmetry transformation.

The expressions (2.18) and (2.19) depend on tr $\mathcal{M}$. But if one considers a symmetry transformation that acts nontrivially on the target space indices, then the other components of $\mathcal{M}_{ij}(x)$ are also involved. In Sec. III the whole effective measure (1.1) is computed.

I would like to add a few remarks on the global properties of the sigma model two-dimensional space-time manifold in view of an application of the present formalism in the string theory context. At the tree string level, i.e., with the world sheet topology of the sphere, the conformal (Diff and Weyl) transformation properties of the partition function $\mathcal{Z}$ of the sigma model determine all the constraints the interacting string states must satisfy. When string loops are considered, that is, with world sheets of higher genus, the knowledge of the dependence of $\mathcal{Z}$ on the Diff Weyl transformation parameters is not sufficient to obtain the complete expressions of the string amplitude constraints (one has to integrate also over the moduli parameters). So we are mainly interested here in world sheets with the topology of the sphere. The connection of such a compact manifold with the sigma model space-time we are considering can be obtained in the following way. One can select a point P on the surface and project the surface on the plane with the point P identified with infinity. Then, the set of the metrics $\{g_{\mu\nu}(x)\}$ relevant for our problem is identified with the orbit of the Diff and Weyl groups of transformations acting on some reference metric $h_{\mu\nu}(x)$. For instance, one can choose $h_{\mu\nu}(x)$ to be the standard metric obtained by a stereographic projection of the sphere on the plane

$$h_{\mu\nu}(x) = \delta_{\mu\nu} 4R^4/(R^2+x^2)^2. \quad (2.20)$$

In what follows, we assume that the metric $g_{\mu\nu}(x)$ is related to the reference metric $h_{\mu\nu}(x)$, Eq. (2.20), by a Diff and/or Weyl transformation. Moreover, in order to keep the distinguished point P on the surface fixed, the infinitesimal parameters of the Diff transformations are assumed to vanish at infinity. This last assumption is just connected with the requirement of maintaining the boundary conditions fixed so that the transformations that one considers act on the states of the same Hilbert space.

III. THE PERTURBATIVE EXPANSION

With the general form (2.5) of the H operator, the computation of the effective measure is not straightforward because the quadratic part of H in the momenta p_μ depends on the position operators q^ν. This problem can be solved by using the method adopted in Ref. 3. One has

$$\langle x,i|e^{-\epsilon H(p,q)}|x,j\rangle = \langle 0,i|e^{-\epsilon H(p,q+x)}|0,j\rangle, \quad (3.1)$$

where

$$q^\mu|0\rangle = 0. \quad (3.2)$$

Consider now a Taylor expansion of $H(p,q+x)$ in powers of q^μ:

$$H(p,q+x) = H_0(p;x) + H_I(p,q;x), \quad (3.3)$$

where

$$H_0(p;x) = A(x)\hat{g}^{\mu\nu}(x)p_\mu p_\nu, \quad (3.4)$$

and

$$\begin{aligned} H_I(p,q;x) &= \partial_\sigma[A(x)\hat{g}^{\mu\nu}(x)]p_\mu q^\sigma p_\nu \\ &\quad + \tfrac{1}{2}\partial_\sigma\partial_\tau[A(x)\hat{g}^{\mu\nu}]p_\mu q^\sigma q^\tau p_\nu \\ &\quad + 2iB^\mu(x)p_\mu + i\,\partial_\sigma B^\mu(x)[p_\mu q^\sigma + q^\sigma p_\mu] \\ &\quad + C(x) + \cdots. \end{aligned} \quad (3.5)$$

In Eqs. (3.4) and (3.5) the indices i and j have been omitted and A, B^μ, and C represent the corresponding matrices. The relevant terms of the Taylor expansion of the amplitude (3.1) in H_I are

$$\begin{aligned} &\langle 0,i|e^{-\epsilon H}|0,j\rangle \\ &= \langle 0,i|e^{-\epsilon H_0}|0,j\rangle - \epsilon\int_0^1 d\alpha \\ &\quad \times\langle 0,i|e^{-\epsilon(1-\alpha)H_0}H_I e^{-\epsilon\alpha H_0}|0,j\rangle \\ &\quad + \epsilon^2\int_0^1 \alpha\,d\alpha\int_0^1 d\beta\,\langle 0,i|e^{-\epsilon(1-\alpha)H_0}H_I \\ &\quad \times e^{-\epsilon\alpha(1-\beta)H_0}H_I e^{-\epsilon\alpha\beta H_I}|0,j\rangle + \cdots. \end{aligned} \quad (3.6)$$

Each term of the expansion (3.6) can be easily computed; the reason is that H_0 does not depend on q^μ, and the q^μ operators contained in H_I can be moved to the left (or to the right) where they vanish because of Eq. (3.2). After the q^μ operators have been eliminated, one can go to the momentum basis

$$\int d^2k\,|k\rangle\langle k| = 1, \quad (3.7)$$

with

$$p_\mu|k\rangle = k_\mu|k\rangle, \quad (3.8)$$

and

$$\langle 0|k\rangle = \langle k|0\rangle = 1/2\pi. \quad (3.9)$$

The integral in the momenta is easily performed (for a specific set of eigenvalues of the A matrix) by using

$$\begin{aligned} &\int d^2k\,e^{-F(x)\hat{g}^{\rho\sigma}(x)k_\rho k_\sigma}k_{\mu_1}k_{\nu_1}\cdots k_{\mu_n}k_{\nu_n} \\ &= \frac{\pi}{2^n F^{n+1}(x)}\,[\hat{g}_{\mu_1\nu_1}(x)\cdots\hat{g}_{\mu_n\nu_n}(x) + \text{permutations}], \end{aligned} \quad (3.10)$$

where

$$\hat{g}_{\mu\nu}(x) = [1/\sqrt{g(x)}]g_{\mu\nu}(x) \quad (3.11)$$

is the inverse of $\hat{g}^{\mu\nu}(x)$. Finally, one has to integrate over the Feynman parameters α and β. How to perform this last integration will be illustrated in the following. Now we give the details of the computation.

To zeroth order,

$$(0) = \langle 0,i|\epsilon^{-\epsilon H_0}|0,j\rangle$$
$$= \langle 0,i|\exp[-\epsilon A(x)\hat{g}^{\mu\nu}(x)p_\mu p_\nu]|0,j\rangle$$
$$= \frac{1}{(2\pi)^2}\int d^2k\{\exp[-\epsilon A(x)\hat{g}^{\mu\nu}(x)k_\mu k_\nu]\}_{ij}. \tag{3.12}$$

This integral is easily performed in the basis in which the A matrix is diagonal. From Eq. (3.10) it follows then

$$(0) = (1/4\pi\epsilon)A_{ij}^{-1}(x). \tag{3.13}$$

To first order,

$$(\mathrm{I}) = -\epsilon\int_0^1 d\alpha\langle 0,i|e^{-\epsilon(1-\alpha)H_0}H_I e^{-\epsilon\alpha H_0}|0,j\rangle$$
$$= -\epsilon\int_0^1 d\alpha\langle 0,i|e^{-\epsilon(1-\alpha)H_0}\{C(x) + i\partial_\sigma B^\mu(x) \times[p_\mu q^\sigma + q^\sigma p_\mu] + \tfrac{1}{2}\partial_\sigma\partial_\tau[A(x)\hat{g}^{\mu\nu}(x)] \times p_\mu q^\sigma q^\tau p_\nu\}e^{-\epsilon\alpha H_0}|0,j\rangle. \tag{3.14}$$

In Eq. (3.14) we have omitted the terms of H_I that give a vanishing contribution in the $\epsilon\to 0$ limit. We also have omitted the terms that, even if they are relevant in the $\epsilon\to 0$ limit by power counting, give a vanishing result by parity. In Eq. (3.14) we have three terms; let us consider each of them separately:

$$(\mathrm{I})_1 = -\frac{\epsilon}{4\pi^2}\int_0^1 d\alpha\int d^2k \times(\exp[-\epsilon(1-\alpha)A(x)\hat{g}^{\mu\nu}(x)k_\mu k_\nu])_{im}C_{mn}(x) \times(\exp[-\epsilon\alpha A(x)\hat{g}^{\rho\sigma}(x)k_\rho k_\sigma])_{nj}.$$

In the basis in which $A(x)$ is diagonal with eigenvalues A_i, one has

$$(\mathrm{I})_1 = -\frac{1}{4\pi}\int_0^1 d\alpha[(1-\alpha)A_i + \alpha A_j]^{-1}C_{ij}(x). \tag{3.15}$$

In order to perform the integration in α, we use the following identity:

$$[(1-\alpha)A_i + \alpha A_j]^{-1} = \int_0^\infty dt\,[(1-\alpha)(A_i+t) + \alpha(A_j+t)]^{-2}. \tag{3.16}$$

Equation (3.15) can then be written as

$$(\mathrm{I})_1 = -\frac{1}{4\pi}\int_0^\infty dt\left[\frac{1}{A(x)+t}C(x)\frac{1}{A(x)+t}\right]_{ij}. \tag{3.17}$$

It should be noted that in order to perform the integration on the Feynman parameters, we shall frequently use the trick illustrated in Eq. (3.16). In general one has

$$F^{-n} = n\int_0^\infty dt(t+F)^{-n-1},\ \text{for}\ n\geqslant 1. \tag{3.18}$$

Now let us consider $(\mathrm{I})_2$:

$$(\mathrm{I})_2 = -i\epsilon\int_0^1 d\alpha\langle 0,i|\epsilon^{-\epsilon(1-\alpha)H_0}\partial_\sigma B^\mu(x) \times[p_\mu q^\sigma + q^\sigma p_\mu]e^{-\epsilon\alpha H_0}|0,j\rangle$$
$$= 2\epsilon^2\int_0^1 d\alpha\langle 0,i|e^{-\epsilon(1-\alpha)H_0}[(1-\alpha)A\,\partial_\sigma B^\mu - \alpha\,\partial_\sigma B^\mu A]\hat{g}^{\sigma\nu}p_\mu p_\nu e^{-\epsilon\alpha H_0}|0,j\rangle. \tag{3.19}$$

The integral in the momenta [for diagonal $A(x)$] gives

$$(\mathrm{I})_2 = \frac{1}{4\pi}\int_0^1 d\alpha\frac{(1-\alpha)A_i(\partial_\mu B^\mu)_{ij} - \alpha(\partial_\mu B^\mu)_{ij}A_j}{[(1-\alpha)A_i + \alpha A_j]^2}. \tag{3.20}$$

By means of Eq. (3.18), one finally obtains

$$(\mathrm{I})_2 = \frac{1}{4\pi}\int_0^\infty dt\,t\left[\frac{1}{A+t}\partial_\mu B^\mu\frac{1}{(A+t)^2} - \frac{1}{(A+t)^2}\partial_\mu B^\mu\frac{1}{A+t}\right]_{ij}. \tag{3.21}$$

In deriving the results (3.13), (3.17), and (3.21) we have illustrated the basic ingredients used in the computation. From now on we simply report the final expression:

$$(\mathrm{I})_3 = -\frac{\epsilon}{2}\int_0^1 d\alpha\langle 0,i|e^{-\epsilon(1-\alpha)H_0}\partial_\sigma\partial_\tau \times[A(x)\hat{g}^{\mu\nu}(x)]p_\mu q^\sigma q^\tau p_\nu e^{-\epsilon\alpha H_0}|0,j\rangle$$
$$= -\frac{1}{24\pi}\left(\partial_\mu\partial_\nu\hat{g}^{\mu\nu} + \frac{1}{2}\hat{g}^{\mu\nu}\partial_\mu\partial_\nu\hat{g}^{\rho\sigma}\hat{g}_{\rho\sigma}\right)\delta_{ij} - \frac{1}{2\pi}\int_0^\infty dt\,t\,\partial_\mu\hat{g}^{\mu\nu}\left[\frac{1}{A+t}\partial_\nu A\frac{A}{(A+t)^3}\right]_{ij} - \frac{1}{2\pi}\int_0^\infty dt\,t\hat{g}^{\mu\nu}\left[\frac{1}{A+t}\partial_\mu\partial_\nu A\frac{A}{(A+t)^3}\right]_{ij}. \tag{3.22}$$

To second order,

$$(\mathrm{II}) = \epsilon^2\int_0^1\alpha\,d\alpha\int_0^1 d\beta\,\langle 0,i|e^{-\epsilon(1-\alpha)H_0}H_I e^{-\epsilon\alpha(1-\beta)H_0}H_I e^{-\epsilon\alpha\beta H_0}|0,j\rangle$$
$$= \epsilon^2\int_0^1\alpha\,d\alpha\int_0^1 d\beta\,\langle 0,i|e^{-\epsilon(1-\alpha)H_0}\{\partial_\sigma(A\hat{g}^{\mu\nu})p_\mu q^\sigma p_\nu e^{-\epsilon\alpha(1-\beta)H_0}\partial_\tau(A\hat{g}^{\rho\sigma})p_\rho q^\tau p_\sigma + 2i\,\partial_\sigma(A\hat{g}^{\mu\nu})p_\mu q^\sigma p_\nu e^{-\epsilon\alpha(1-\beta)H_0}B^\rho p_\rho + 2iB^\rho p_\rho e^{-\epsilon\alpha(1-\beta)H_0}\partial_\tau(A\hat{g}^{\mu\sigma})p_\mu q^\tau p_\sigma - 4B^\mu p_\mu e^{-\epsilon\alpha(1-\beta)H_0}B^\nu p_\nu\}e^{-\epsilon\alpha\beta H_0}|0,j\rangle. \tag{3.23}$$

As usual, terms that (by power counting) are irrelevant in the $\epsilon \to 0$ limit have been omitted in Eq. (3.23). The contributions of the four terms shown in Eq. (3.23) are

$$(\mathrm{II})_1 = \epsilon^2 \int_0^1 \alpha\, d\alpha \int_0^1 d\beta\, \langle 0,i| e^{-\epsilon(1-\alpha)H_0} \partial_\sigma (A\hat{g}^{\mu\nu}) p_\mu q^\sigma p_\nu e^{-\epsilon\alpha(1-\beta)H_0} \partial_\tau (A\hat{g}^{\rho\sigma}) p_\rho q^\tau p_\sigma e^{-\epsilon\alpha\beta H_0} |0,j\rangle$$

$$= \frac{1}{8\pi} \delta_{ij} \left(\frac{1}{6} \partial_\mu \hat{g}^{\nu\tau} \partial_\nu \hat{g}^{\mu\lambda} \hat{g}_{\tau\lambda} - \frac{1}{12} \hat{g}^{\mu\nu} \partial_\mu \hat{g}^{\tau\lambda} \partial_\nu \hat{g}_{\tau\lambda} \right) + \frac{1}{8\pi} \int_0^\infty dt\, \partial_\mu \hat{g}^{\mu\nu}$$

$$\times \left[\frac{1}{A+t} \partial_\nu A \frac{1}{A+t} - \frac{A}{(A+t)^2} \partial_\nu A \frac{A}{(A+t)^2} - \frac{A^2}{(A+t)^3} \partial_\nu A \frac{1}{A+t} - \frac{1}{A+t} \partial_\nu A \frac{A^2}{(A+t)^3} \right]_{ij}$$

$$+ \frac{1}{8} \int_0^\infty dt\, \hat{g}^{\mu\nu} \left[\frac{1}{A+t} \partial_\mu A \frac{1}{A+t} \partial_\nu A \frac{1}{A+t} - 2t \frac{A}{(A+t)^2} \partial_\mu A \frac{1}{A+t} \partial_\nu A \frac{1}{(A+t)^2} \right.$$

$$\left. - 2t \frac{1}{(A+t)^2} \partial_\mu A \frac{1}{A+t} \partial_\nu A \frac{A}{(A+t)^2} \right]_{ij}, \tag{3.24}$$

$$(\mathrm{II})_2 = 2i\epsilon^2 \int_0^1 \alpha\, d\alpha \int_0^1 d\beta\, \langle 0,i| e^{-\epsilon(1-\alpha)H_0} \partial_\sigma (\hat{g}^{\mu\nu} A) p_\mu q^\sigma p_\nu e^{-\epsilon\alpha(1-\beta)H_0} B^\tau p_\tau e^{-\epsilon\alpha\beta H_0} |0,j\rangle$$

$$= \frac{-1}{4\pi} \int_0^\infty dt \left[\frac{1}{A+t} \partial_\mu A \frac{1}{A+t} B^\mu \frac{1}{A+t} + t \hat{g}^{\mu\nu} \partial_\mu \hat{g}_{\nu\lambda} \frac{A}{(A+t)^3} B^\lambda \frac{1}{A+t} \right.$$

$$\left. - 2t \frac{1}{(A+t)^2} \partial_\mu A \frac{1}{A+t} B^\mu \frac{1}{A+t} \right]_{ij}, \tag{3.25}$$

$$(\mathrm{II})_3 = 2i\epsilon^2 \int_0^1 \alpha\, d\alpha \int_0^1 d\beta\, \langle 0,i| e^{-\epsilon(1-\alpha)H_0} B^\tau p_\tau e^{-\epsilon\alpha(1-\beta)H_0} \partial_\sigma (\hat{g}^{\mu\nu} A) p_\mu q^\sigma p_\nu e^{-\epsilon\alpha\beta H_0} |0,j\rangle$$

$$= \frac{1}{4\pi} \int_0^\infty dt \left[\frac{1}{A+t} B^\mu \frac{1}{A+t} \partial_\mu A \frac{1}{A+t} + t \hat{g}^{\mu\nu} \partial_\mu \hat{g}_{\nu\lambda} \frac{1}{A+t} B^\lambda \frac{A}{(A+t)^3} \right.$$

$$\left. - 2t \frac{1}{A+t} B^\mu \frac{1}{A+t} \partial_\mu A \frac{1}{(A+t)^2} \right]_{ij}, \tag{3.26}$$

$$(\mathrm{II})_4 = -4\epsilon^2 \int_0^1 \alpha\, d\alpha \int_0^1 d\beta\, \langle 0,i| e^{-\epsilon(1-\alpha)H_0} B^\mu p_\mu e^{-\epsilon\alpha(1-\beta)H_0} B^\nu p_\nu e^{-\epsilon\alpha\beta H_0} |0,j\rangle$$

$$= -\frac{1}{2\pi} \int_0^\infty dt\, \hat{g}_{\mu\nu} \left[\frac{1}{A+t} B^\mu \frac{1}{A+t} B^\nu \frac{1}{A+t} \right]_{ij}. \tag{3.27}$$

The sum of the contributions (3.13), (3.17), (3.21), (3.22), (3.24), (3.25), (3.26), and (3.27) gives the explicit form of the one loop effective measure (1.1). The full expression of $\mathscr{M}_{ij}(x)$ is quite complicated. Fortunately, if one is interested in the Diff anomaly (2.18) or in the one loop divergences (2.19), only the trace $\mathrm{tr}\,\mathscr{M}(x)$ is relevant and this quantity simplifies considerably, as shown in Sec. IV.

IV. THE Diff ANOMALY AND THE ONE LOOP DIVERGENCES

In Sec. III we computed the full expression of the one loop effective measure $\mathscr{M}_{ij}(x)$. This result is used in the present section to find the Diff anomaly (2.18) and the one loop divergences of the generalized sigma mode with action (2.1).

First of all we have to consider the trace in the internal indices i and j of $\mathscr{M}_{ij}(x)$. From the results of Sec. III, one finds

$$\mathrm{tr}\,\mathscr{M}(x) = (1/4\pi\epsilon)\mathrm{tr}(A^{-1}) + (D/24\pi)\left[-\partial_\mu \partial_\nu \hat{g}^{\mu\nu} + \tfrac{1}{2} \partial_\mu \hat{g}^{\tau\nu} \partial_\nu \hat{g}^{\mu\lambda} \hat{g}_{\tau\lambda} + \tfrac{1}{8} \hat{g}^{\mu\nu} \partial_\mu \hat{g}^{\tau\sigma} \partial_\nu \hat{g}_{\tau\sigma} \right]$$

$$- (1/4\pi)\left[\mathrm{tr}(A^{-1}C) + \tfrac{1}{4} \hat{g}^{\mu\nu} \mathrm{tr}(A^{-1} \partial_\mu A A^{-1} \partial_\nu A) + \hat{g}_{\mu\nu}\, \mathrm{tr}(A^{-1} B^\mu A^{-1} B^\nu) \right]$$

$$- (1/12\pi) \partial_\mu \left[\hat{g}^{\mu\nu}\, \mathrm{tr}(A^{-1} \partial_\nu A) \right]. \tag{4.1}$$

Now, one can substitute in Eq. (4.1) the particular values (2.8)–(2.10) of the matrices A, B^μ, and C.

Note that the expression (4.1) contains a term that is a function of the two-dimensional metric only and does not depend on A, B^μ, and C. It is proportional to D, the number of field components, and can be written as[3]

$$(D/24\pi)\sqrt{g}\left[R^{(2)} + \nabla^2 \ln\sqrt{g} \right], \tag{4.2}$$

where $R^{(2)}$ is the scalar curvature of the classical two-dimensional metric $g_{\mu\nu}$. The term (4.2) coincides precisely with the expression of the Diff anomaly for a noninteracting sigma model.[3] Its presence in the expression (4.1) was expected, of course, and provides a partial check of the computation. It should be noted that the multiplicative constant of the expression (4.2) is connected with the value of the central charge of the Virasoro algebra associated with the energy momentum tensor of the model in flat space-time.[3] The same expression (4.2) is also connected with the existence of the term $R^{(2)} D/(24\pi)$ in the trace anomaly.[3]

Finally, from Eqs. (2.18) and (4.1) one obtains

$$\Delta_V\Gamma_1 = \frac{1}{8\pi\epsilon}\int d^2x\,\partial_\lambda V^\lambda(x) G^{ii}(\phi(x)) + \frac{D}{48\pi}\int d^2x\,\partial_\lambda V^\lambda(x)\sqrt{g}[R^{(2)} + \nabla^2 \ln\sqrt{g}]$$
$$+\frac{1}{8\pi}\int d^2x\,\partial_\lambda V^\lambda(x)\{[R_{ij} - T_{ikl}T_j^{\ kl} + \nabla_i(\nabla_j \ln\sqrt{G} - G^{kl}\partial_k G_{lj})]\partial_\mu\phi^i\partial_\nu\phi^j\sqrt{g}g^{\mu\nu}$$
$$-i[-\nabla^k T_{kij} + T_{ij}^{\ \ k}(\nabla_k \ln\sqrt{G} - G^{ml}\partial_m G_{lk})]\epsilon^{\mu\nu}\partial_\mu\phi^i\partial_\nu\phi^j\}$$
$$-\frac{1}{8\pi}\int d^2x\,\sqrt{g}g^{\mu\nu}\partial_\mu\partial_\lambda V^\lambda(x) G^{ij}\partial_i G_{jk}\,\partial_\nu\phi^k + \frac{1}{12\pi}\int d^2x\,\sqrt{g}g^{\mu\nu}\partial_\mu\partial_\lambda V^\lambda(x)\partial_\nu \ln\sqrt{G}, \qquad (4.3)$$

where R_{ij} and G are the Ricci tensor and the determinant of the metric $G_{ij}(\phi)$ of the target space.

Equation (4.3) shows the behavior of the one loop effective action under a Diff transformation (2.13) and (2.14). But does the Eq. (4.3) represent an anomaly? Perhaps some of the terms shown in Eq. (4.3) can be eliminated by adding to Γ_1 some appropriate local counterterms. As discussed in Refs. 3 and 7, the answer to this question is meaningful for the combined set of Diff and Weyl transformations because it may happen that it is possible, for instance, to eliminate a Diff "anomaly" at the price of introducing a Weyl "anomaly," and vice versa. In the following I will require invariance under Weyl transformations. This means that the possible counterterms must be Weyl invariant; that is, the counterterms do not depend on the determinant of the classical metric $g_{\mu\nu}$.

So, let us consider the first term in the expression (4.3). This term *cannot* represent an anomaly because it is divergent in the $\epsilon\to 0$ limit. In fact, it can be eliminated by the counterterm

$$S_c = \frac{1}{8\pi\epsilon}\int d^2x\, G^{ii}(\phi(x)). \qquad (4.4)$$

Actually, the counterterm (4.4) is automatically present if one takes care of the quadratic divergences of the integration measure in the Feynman path integral.

As discussed in Ref. 3, the second term in the expression (4.3) cannot be eliminated by any counterterm without violating the Weyl invariance. In fact, it corresponds to the Diff anomaly of a noninteracting bosonic sigma model.

In all the remaining terms of the expression (4.3), $\partial_\lambda V^\lambda$ multiplies a scalar density. In general, it is not possible to eliminate these terms in a Weyl invariant way. There is an exception, however. A term of the form

$$\int d^2x\sqrt{g}g^{\mu\nu}\partial_\mu\partial_\lambda V^\lambda\,\partial_\nu B, \qquad (4.5)$$

where B is a scalar, can be eliminated by the Weyl-invariant counterterm

$$S'_c = \int d^2x\sqrt{g}B\,[R^{(2)} + \nabla^2 \ln\sqrt{g}]. \qquad (4.6)$$

In fact, S'_c is Weyl invariant because under a transformation (2.12) one has

$$\Delta_\sigma R^{(2)} = -2\sigma R^{(2)} - 2\nabla^2\sigma. \qquad (4.7)$$

On the other hand, $\Delta_V S'_c$ cancels the expression (4.5) because under a Diff transformation (2.13) one obtains

$$\Delta_V \ln\sqrt{g} = V^\lambda\partial_\lambda \ln\sqrt{g} + \partial_\lambda V^\lambda. \qquad (4.8)$$

So, the last term of the expression (4.3) does not represent an anomaly because it can be eliminated by an appropriate Weyl-invariant counterterm. It should be stressed that this fact is quite important in the string theory context and it is related in particular to the dilaton coupling.[5]

In conclusion, apart from the first and the last term, the expression (4.3) represents just the one loop Diff anomaly for the two-dimensional generalized sigma model (2.1). It is easy to see that the expression for the Diff anomaly satisfies the consistency conditions following from the structure of the Diff transformation group.[3] Moreover, as illustrated in Ref. 3, it is possible to eliminate the Diff anomaly by adding to Γ a local functional of the *entire* metric $g_{\mu\nu}(x)$ but at the price of introducing Weyl anomalies.

The expression (4.3) also gives rise to the correct low energy equations of motion for the graviton, dilaton, and antisymmetric tensor in the string theory context.[5] I shall not enter into details here; a complete analysis of this point necessarily involves the discussion of many other string aspects. I will consider these problems in a future publication.[5] Just two comments are in order.

First, in general the introduction of a new interaction term in the σ-model Lagrangian gives new contributions to the expression of the Diff anomaly. Thus the results of the present paper remain valid: one simply has to add the new contributions coming from the insertion of new vertex operators.

Second, the interested reader may ask how the dilaton coupling can be introduced since, at first sight, the Lagrangian (2.1) only contains the metric (or graviton) vertex operator. Now, there is not a unique answer to this question because it turns out that the dilaton coupling can be introduced in many different ways. One way is particularly significant, however. It is well-known from the early days of the string theory that, in the limit of vanishing momentum, the physical dilaton couples to the trace of the energy-momentum tensor (of the external world, of course). In fact, this is the reason why this particle is called the dilaton. But this also means that, with the proper identification of the variables, the dilaton coupling should already be present in the Lagrangian (2.1)! And, indeed, this is the case.[5]

Let us consider now the one loop divergences of the model. From Eqs. (2.19) and (4.1) it follows that

$$\epsilon \frac{d\Gamma_1}{d\epsilon} = \frac{1}{8\pi\epsilon} \int d^2x\, G^{ii}(\phi(x)) + \frac{1}{8\pi} \int d^2x \{ [R_{ij} - T_{ikl} T_j{}^{kl} + \nabla_i (\nabla_j \ln\sqrt{G} - G^{kl} \partial_k G_{lj})] \partial_\mu \phi^i \partial_\nu \phi^j \sqrt{g} g^{\mu\nu} - i [-\nabla^k T_{kij} + T_{ij}{}^k (\nabla_k \ln\sqrt{G} - G^{lm} \partial_m G_{lk})] \epsilon^{\mu\nu} \partial_\mu \phi^i \partial_\nu \phi^j \}. \tag{4.9}$$

The first term in Eq. (4.9) represents just the quadratic divergence, which is eliminated by the counterterm (4.4). The remaining part of the expression (4.9) corresponds to the logarithmic divergence of the model.

Note that in the expression (4.9) I have not forgotten to add the term

$$I = \frac{D}{24\pi} \int d^2x \sqrt{g} [R^{(2)} + \nabla^2 \ln \sqrt{g}]. \tag{4.10}$$

The point is that the integral (4.10) is vanishing! The easiest way to see this is to note that, even if the integrand in (4.10) is not a scalar denisty, the integral I is invariant under Diff transformations (2.13). Thus in computing it one can use a metric that is conformally flat. But with a conformally flat metric the integrand $R^{(2)} + \nabla^2 \ln \sqrt{g}$ is vanishing, and therefore $I = 0$. We are assuming here that the metric $g_{\mu\nu}$ has no singularities, of course. It should also be noted that we are considering the set of metrics specified at the end of Sec. II.

In Eq. (4.9) the expression of the logarithmic divergences of the model is in agreement with the one loop beta functions and wave function regularization computed in the case of flat two-dimensional space-time.[2,8,9]

Since the integral I is vanishing, one concludes that at one loop the divergences of the generalized sigma model (2.1) in a gravitational background (specified by a metric $g_{\mu\nu}$ which is asymptotically conformally flat) are just the same divergences of the model in flat space-time. This result can be formulated in a way which is more appropriate for a generalization as follows: Suppose we have a generalized sigma model in flat two-dimensional space-time without divergences; this means vanishing of the beta functions and absence of a divergent wave function regularization.[9] Then, the same model is finite even in presence of an (asymptotically conformally flat) gravitational background.

At the one loop level this is certainly true; we have just performed the explicit computation in the present paper, Eq. (4.9). I do not have a rigorous proof that this statement remains true at higher loops, but I conjecture that this is indeed the case at any order of the loop expansion.

Finally, let us consider the relation between the Diff anomaly and the logarithmic divergences ("beta" functions) of the model at one loop. Comparing the Eqs. (4.3) and (4.9), it is clear that the knowledge of the divergences of the model is not sufficient to determine the whole expression of the Diff anomaly. One could have obtained the same conclusion just by considering Eqs. (2.18) and (2.19). By the way, it is not difficult to see that Eqs. (2.18) and (2.19) can be generalized at higher loop orders. In particular, if one adopts the Schwinger proper-time method to regularize the theory, then the relation

$$\Delta_V \Gamma = \int d^2x \; \partial_\lambda V^\lambda(x) G(x), \qquad \epsilon \frac{d\Gamma}{d\epsilon} = \int d^2x \; G(x) \tag{4.11}$$

holds at every order of the loop expansion.[5]

In conclusion, the divergences of the sigma model (which are important from the field theoretical point of view) do not give complete information on the Diff anomaly which, on the other hand, is the relevant quantity to consider in the string theory context. In particular, the value of the central charge is not necessarily related to a divergence of the sigma model. This result is not surprising, after all. If it makes sense to speak about the critical dimensions of a string theory, then the value of the central charge has to be computed without the ambiguities related to a modification of some arbitrary convention in the renormalization procedure. In fact, the value of the central charge is fixed by the Diff anomaly, whose expression is obtained by computing finite Feynman diagrams.

V. SUMMARY AND CONCLUSIONS

In the present paper, I have derived the whole expression of the one loop effective measure for a two-dimensional generalized sigma model in gravitational background. By using the Schwinger proper-time regularization, it has been possible to carry out the computation without any approximation. The advantage of producing an exact result has been used to show the precise relation at one loop between the conformal anomalies and the logarithmic divergences of the model.

First of all I have shown how it is possible, and in fact more convenient, to preserve the Weyl invariance in the construction of the effective action. In this framework the conformal breakdown is controlled by a Diff anomaly, whose expression has been derived at one loop. The result gives the right value of the central charge and gives rise to the correct low energy equations of motion for the massless states of the string.[5]

Concerning the logarithmic divergences of the model, the result is in agreement with the beta function and wave function computations in flat space-time. As expected, the presence of a gravitational background does not modify the short-distance singularities of the theory. It has been shown also that the divergences of the model do not determine completely the expression of the Diff anomaly. In particular, the value of the central charge is not necessarily related to a divergence of the sigma model.

Note added in proof: After this work was completed, R. Silvotti informed me of the preprint,[10] in which some of the problems considered here have been studied in a different framework.

ACKNOWLEDGMENTS

I wish to thank P. Menotti, M. Mintchev, R. Silvotti, and E. Vesentini for useful discussions.

[1]A. M. Polyakov, Phys. Lett. B **103**, 207 (1981); C. Lovelace, *ibid.* **135**, 75 (1984); E. S. Fradkin and A. A. Tseytlin, Nucl. Phys. B **261**, 1 (1985).
[2]C. G. Callan, D. Friedan, E. J. Martinec, and M. J. Perry, Nucl. Phys. B **262**, 593 (1985); E. S. Fradkin and A. A. Tseytlin, Phys. Lett. B **158**, 316 (1985).
[3]E. Guadagnini, "Central charge, trace and Diff anomalies in two dimensions," Pisa preprint, IFUP-TH 30/87, 1987.
[4]J. Schwinger, Phys. Rev. **82**, 664 (1951).
[5]E. Guadagnini, in preparation.
[6]L. F. Abbott, Nucl. Phys. B **185**, 189 (1981).
[7]L. Bonora, P. Pasti, and M. Tonin, Phys. Lett. B **149**, 346 (1984); L. Bonora, P. Pasti, and M. Bregola, Class. Quant. Gravity **3**, 635 (1986).
[8]D. Friedan, Ann. Phys. (NY) **163**, 318 (1985); L. Alvarez-Gaumè, D. Z. Freedman, and S. Mukhi, *ibid.* **134**, 85 (1981).
[9]E. Guadagnini, Phys. Rev. D **37**, 2253 (1988).
[10]G. Keller and R. Silvotti, "Quantum measure and Weyl anomaly of two dimensional bosonic non linear σ-models," Zurich University preprint, 1987.

Gauge group cohomology in the monopole sector of Yang–Mills theories

I. Bakas
Center for Relativity, The University of Texas at Austin, Austin, Texas 78712
D. McMullan
Department of Physics and Astronomy, The University of Glasgow, Glasgow G12 8QQ, Scotland, United Kingdom

(Received 12 August 1987; accepted for publication 30 March 1988)

The theory of Chern–Simons secondary characteristic classes is used to study the gauge group cohomology associated with Yang–Mills theories in sectors with persistent boundary effects. A generalized "descent" tower of globally defined cocycles (for manifolds with boundaries) is adopted and its relevance to monopole physics is investigated by means of the Christ–Jackiw interpretation for Yang–Mills dyons. In particular, it is found that three-cocycles can arise in the monopole sector of gauge theories and explicit formulas are derived for their description.

I. INTRODUCTION

Topological methods have been shown to be of paramount importance in studying various nonperturbative aspects of field theories. In particular, the physics of nonlinear σ-models, instantons, and monopoles as well as the description of anomalous gauge theories provide the most characteristic examples (see for instance Ref. 1). Moreover, the dynamical role of persistent boundary effects in gauge theories, due to the presence of topological objects (such as monopoles) has been well appreciated and led to a series of important developments including the fermion number fractionization,[2] the Witten effect,[3] the Callan–Rubakov effect,[4] and more recently the color problem in grand unified theories (GUT).[5] It is the purpose of the present work to elaborate on the significance of the gauge group cohomology in theories with persistent boundary effects.

Traditionally, gauge group cohomology provides the appropriate framework in which to study the structure of anomalous gauge theories. In particular, one- and two-cocycles describe Wess–Zumino actions and Schwinger terms (together with their consistency conditions) that result upon quantization of Yang–Mills theories coupled to chiral fermions.[6] The basic mathematical construction used in anomalous gauge theories has been the descent tower of globally defined gauge group cocycles α_k that are obtained from the Chern–Pontryagin density

$$\Omega_{2n}^0(A) \equiv \omega_{2n}(A) = \frac{1}{n!}\left(\frac{i}{2\pi}\right)^n \operatorname{Tr} F^n, \tag{1}$$

via a standard transgressional procedure.[6] The coboundary operator Δ in group cohomology acts on cochains by

$$\begin{aligned}\Delta\alpha_k(A;g_1,g_2,\ldots,g_{k+1}) &= \alpha_k(A^{g_1};g_2,\ldots,g_{k+1}) + (-)^{k+1}\alpha_k(A;g_1,\ldots,g_k)\\ &\quad + \sum_{i=1}^{k}(-)^i\alpha_k(A;g_1,\ldots,g_ig_{i+1},\ldots,g_{k+1}),\end{aligned} \tag{2}$$

where $A^g = g^{-1}Ag + g^{-1}\,dg$. The gauge group cocycles $\alpha_k(A;g_1,\ldots,g_k)$ are real valued functions of the vector potential A and the gauge group elements $g_1,\ldots,g_k$, satisfying the condition $\Delta\alpha_k = 0$. For all such k-cocycles α_k the defining k-simplex C_k belongs entirely in the identity component of the gauge group $\mathcal{G}$ with one vertex identified with the identity element and the rest being $g_1, g_1g_2,\ldots,g_1g_2\cdots g_k$.

Gauge anomalies in $2n$ dimensions are characterized by cocycles obtained transgressionally from ω_{2n+2}, i.e., the Chern–Pontryagin density in two dimensions higher. Let $t_1,t_2,\ldots,t_k$ be the parameters defining the C_k simplex in the gauge group $\mathcal{G}$ and let δ be the corresponding differentiation operator, i.e., $\delta = \delta t_i\,\partial/\partial t_i$ (conventionally, $\delta d + d\delta = 0$, where d is the de Rham operator for the space M on which the theory is defined). We take

$$\alpha_k(A;g_1,\ldots,g_k) = 2\pi\int_{M^{2n+1-k}}\int_{C_k}\omega_{2n+1-k}^k(v(x),A^g(x)), \tag{3}$$

where $x\in M^{2n+1-k}$, the $(2n+1-k)$-dimensional space M and $v = g^{-1}\delta g$; ω_{2n+1-k}^k are the globally defined members of the tower

$$\begin{aligned}&\omega_{2n+2} = d\omega_{2n+1}^0, \quad \delta\omega_{2n+1}^0 = -d\omega_{2n}^1,\\ &\delta\omega_{2n}^1 = -d\omega_{2n-1}^2,\ldots,\end{aligned} \tag{4}$$

up to ω_0^{2n+1}. Explicit expressions for the ω's can be found, for instance, in Ref. 6. The overall normalization factor in (1) and hence in the rest of the tower is fixed mod Z, an integer counting the overall (net) chirality of the interacting fermions that the theory contains.

Within this framework, anomalies are made manifest in the representation theory of the gauge group on the space of wave functionals $\Psi(A)$. In particular, the Wess–Zumino action is precisely the one-cocycle α_1 entering in the representation of the space-time gauge group,

$$U(g)\Psi(A) = e^{-i\alpha_1(A;g)}\Psi(A^g), \tag{5}$$

while the projective representation of the space gauge group

$$U(g_1)U(g_2) = e^{-i\alpha_2(A;g_1,g_2)}U(g_1g_2), \tag{6}$$

is associated with the anomalous Schwinger term in the equal-time commutation relations of the gauge group generators.[6] The cocycle conditions $\Delta\alpha_1 = 0$, $\Delta\alpha_2 = 0$ are the consistency conditions that gauge anomalies satisfy in the covariant and canonical formalism, respectively. In four-dimensional theories for instance, the consistency conditions emerge from the compactification of space-time M^4 (resp.

 0022-2488/88/081909-07$02.50

space M^3) to S^4 (resp. S^3) due to the falloff rates

$$A_\mu \underset{r\to\infty}{\to} O(r^{-2}), \quad F_{\mu\nu} \underset{r\to\infty}{\to} O(r^{-3}), \tag{7}$$

that are imposed on the Yang–Mills fields at infinity. Indeed, we have that

$$\Delta\alpha_1 = 2\pi \int_{M^4} \int_{C_2} \delta\omega_4^1 = -2\pi \int_{\partial M^4} \int_{C_2} \omega_3^2 = 0, \tag{8a}$$

$$\Delta\alpha_2 = 2\pi \int_{M^3} \int_{C_3} \delta\omega_3^2 = -2\pi \int_{\partial M^3} \int_{C_3} \omega_2^3 = 0, \tag{8b}$$

by means of $\partial S^4 = \emptyset = \partial S^3$.

However, if there are persistent boundary effects in the theory (such as those in the monopole sector of gauge theories), where the falloff rates of the fields at infinity are[3]

$$A_\mu \underset{r\to\infty}{\to} O(r^{-1}), \quad F_{\mu\nu} \underset{r\to\infty}{\to} O(r^{-2}), \tag{9}$$

the compactification used above is not valid anymore. Moreover, the tower of the gauge group cocycles α_k is not globally defined in (nontrivial) monopole sectors because $\omega_{2n+2}(A) = d\omega_{2n+1}^0(A)$ is now inappropriate. Chern–Simons secondary classes have to be used in order to compensate for boundary effects.

This paper is an elaboration on some ideas described briefly in Ref. 7 that relate the gauge group cohomology to the monopole sector of Yang–Mills theories and the associated surface effects.

In Sec. II we present a generalized tower of globally defined gauge-invariant forms obtained from a transgression of the Chern–Simons secondary characteristic classes. By adopting the Chirst–Jackiw interpretation of Yang–Mills dyons[3] we are able (Sec. III) to introduce a new construction of the group cocycles which we believe to be physically relevant when surface effects are important. As an application we show how the magnetic charge of the dyon emerges from this generalized tower of cocycles.

In Sec. IV we examine the possibility for three-cocycles to occur and find that they are mathematically allowed when the defining three-simplex C_3 interpolates (via dyonic configurations) between disconnected components of the (space)-gauge group. Finally we draw our conclusions and compare this construction with other related work.

II. COCYCLES AND SURFACE TERMS

In nontrivial sectors of Yang–Mills theories (e.g., instanton or monopole sectors), the Chern–Pontryagin density ω_{2n} cannot be written globally as a total divergence $\omega_{2n}(A) = d\omega_{2n-1}^0(A)$. In fact, the Chern–Simons secondary characteristic class has to be taken into account[8]:

$$\omega_{2n}(A) - \omega_{2n}(A') = -d\Omega_{2n-1}^1(A',A), \tag{10}$$

where the gauge potentials A, A' belong in the same topological sector of the theory. If the underlying $2n$-dimensional space-time M^{2n} has no boundary, the integral of $d\Omega_{2n-1}^1(A',A)$ vanishes. This situation arises in the instanton sector of gauge theories where the space-time is compactified to a sphere; however, only in the zero instanton sector may we choose $A'=0$ consistently and obtain $\omega_{2n}(A) = d\omega_{2n-1}^0(A)$, where $\omega_{2n-1}^0(A) = \Omega_{2n-1}^1(A,0)$. On the other hand, if $\partial M \neq \emptyset$ we get by Stoke's theorem a surface term that is not necessarily zero. Such a situation arises in the monopole sector of gauge theories due to persistent boundary effects resulting from the falloff rates of the fields at infinity [cf. (9)].

Descent towers of gauge group cocycles, built from (10), have been introduced to construct Wess–Zumino actions in instanton sectors[9] as well as for nonlinear σ-models over homogeneous spaces G/H.[10] We shall adopt similar techniques to obtain cocycles of the gauge group cohomology appropriate for the monopole sector of Yang–Mills theories. In particular, for a structure group G we consider G-principal fiber bundles over the space and/or space-time M and define cochains (Ω^k) in $\mathscr{A}$, the space of all gauge connections with a fixed topological number; these cochains will be taking values in the set of differential forms in M. The coboundary operator $\bar{\Delta}$ is defined to be[10]

$$\begin{aligned}\bar{\Delta}\Omega^k&(A_0,A_1,\ldots,A_{k+1})\\ &= \Omega^k(A_1,A_2,\ldots,A_{k+1}) - \Omega^k(A_0,A_2,\ldots,A_{k+1})\\ &\quad + \cdots + (-)^{k+1}\Omega^k(A_0,A_1,\ldots,A_k).\end{aligned} \tag{11}$$

It satisfies the condition $\bar{\Delta}^2 = 0$ and anticommutes with the space-M differentiation operator d, i.e., $\bar{\Delta}d + d\bar{\Delta} = 0$. Cohomology theory over $\mathscr{A}$ is more general than gauge group cohomology (in fact, it reduces to it when the vertices of the defining simplices belong in the same gauge group orbit in $\mathscr{A}$). However, it is best suited for accommodating properties of monopole physics, as it will be demonstrated later.

Let $\Omega_{2n}^0(A) = \omega_{2n}(A)$ be the Chern–Pontryagin density given by (1). Application of the coboundary operator $\bar{\Delta}$ yields

$$\bar{\Delta}\Omega_{2n}^0(A_0,A_1) = \int_{C_1} \delta\Omega_{2n}^0(A_t),$$

where C_1 is a one-simplex in $\mathscr{A}$-space with vertices A_0, A_1 [e.g., $A_t = tA_1 + (1-t)A_0$; $0 \leqslant t \leqslant 1$] and δ is the differentiation operator with respect to t. Obviously, it reproduces (10) for the choice

$$\Omega_{2n-1}^1(A_0,A_1) = n\frac{1}{n!}\left(\frac{i}{2\pi}\right)^n \int_{C_1} \mathrm{Str}(\delta A_t, F_t^{n-1}). \tag{12}$$

Here Str denotes the symmetrized trace as in Ref. 10. Further application of the coboundary operator $\bar{\Delta}$ leads to the following tower of cochains Ω_{2n-k}^k $(0 \leqslant k \leqslant n)$:

$$\begin{aligned}\bar{\Delta}\Omega_{2n-k}^k&(A_0,A_1,\ldots,A_{k+1})\\ &= -d\Omega_{2n-k-1}^{k+1}(A_0,A_1,\ldots,A_{k+1}),\end{aligned} \tag{13}$$

where Ω_{2n-k}^k is a k-cochain in $\mathscr{A}$-space and a $(2n-k)$-form in $x \in M$-space, given by[10]

$$\begin{aligned}\Omega_{2n-k}^k&(A_0,A_1,\ldots,A_k)\\ &= \frac{1}{n!}\left(\frac{i}{2\pi}\right)^n \binom{n}{k} \int_{C_k} \mathrm{Str}((\delta A_{t_1\cdots t_k})^k, (F_{t_1\cdots t_k})^{n-k}).\end{aligned} \tag{14}$$

C_k denotes the k-simplex in $\mathscr{A}$-space that is taken to be

$$A_{t_1\cdots t_k} = (1 - t_1 - t_2 - \cdots - t_k)A_0 + t_1A_1 + \cdots + t_kA_k;$$

$$0 \leqslant t_1,\ldots,t_k \leqslant 1.$$

In analogy with Eq. (3) we may define k-cochains in $\mathscr{A}$-

space, $\alpha_k(A_0,A_1,\ldots,A_k)$, by integrating Ω^k_{2n-k} over x-space,

$$\alpha_k(A_0,A_1,\ldots,A_k) = 2\pi \int_{M^{2n-k}} \Omega^k_{2n-k}(A_0,A_1,\ldots,A_k). \quad (15)$$

Then,

$$\bar{\Delta}\alpha_k = -2\pi \int_{\partial M^{2n-k}} \Omega^{k+1}_{2n-k-1} \quad (0<k<n)$$

and therefore when there are persistent boundary effects (i.e., $\partial M \neq \emptyset$), α_k does not automatically satisfy the cocycle condition $\bar{\Delta}\alpha_k = 0$ for all k. However, $\bar{\Delta}\alpha_n = 0$ holds irrespective of the topology of the boundary ∂M: This follows from the fact that the tower of cochains (14) terminates at $k=n$, i.e., $\bar{\Delta}\Omega^n_n = 0$. We shall return to it later.

Two remarks are in order. First, notice that Ω^k_{2n-k} [as given by (14)] are all globally defined and gauge invariant under $A_i \to A_i^g$ $(0<i<k)$, i.e.,

$$\Omega^k_{2n-k}(A_0,A_1,\ldots,A_k) = \Omega^k_{2n-k}(A_0^g,A_1^g,\ldots,A_k^g). \quad (16)$$

Also, the space of gauge potentials $\mathscr{A}$ is an affine space (and hence topologically trivial) and so one expects the cohomology theory defined on it to be trivial, i.e., if $\bar{\Delta}\alpha_k = 0$ then there exists β_{k-1} such that $\alpha_k = \bar{\Delta}\beta_{k-1}$. This is definitely the case if no restrictions are placed on the cochains considered. However, in gauge invariant cohomology, where cochains with the property $\alpha_k(A_0^g,A_1^g,\ldots,A_k^g) = \alpha_k(A_0,A_1,\ldots,A_k)$ are only considered, nontrivial cohomology classes can arise: although every cocycle α_k can be written in the form $\bar{\Delta}\beta_{k-1}$, β_{k-1} might not be gauge invariant.

Second, if the vertices of the k-simplices C_k all belong in the same gauge group orbit in $\mathscr{A}$-space, i.e.,

$$A_0 = A\colon\ A_1 = A^{g_1}; A_2 = A^{g_1g_2};\ldots;A_k = A^{g_1g_2\cdots g_k},$$

the cochains $\alpha_k(A_0,A_1,\ldots,A_k) \equiv \alpha_k(A;g_1,g_2,\ldots,g_k)$ satisfy Eq. (2). This means that we can take $\bar{\Delta} = \Delta$ along a gauge group orbit and when $g_1,g_2,\ldots,g_k$ all belong in the identity component of the gauge group[11] the whole construction reduces to the conventional tower of cochains reviewed in the Introduction (for details on the reduction see Ref. 10).

Let us now consider the k-simplices C_k defined by

$$A_{t_1\cdots t_k} \equiv A^{g(t_1,\ldots,t_k)} := g^{-1}(t_1,\ldots,t_k)Ag(t_1,\ldots,t_k) + g^{-1}(t_1,\ldots,t_k)dg(t_1,\ldots,t_k), \quad (17)$$

where $g(1,0,\ldots,0) = g_1$, $g(0,1,\ldots,0) = g_1g_2,\ldots,g(0,\ldots,0,1) = g_1g_2\cdots g_k$ all belong in the same gauge group orbit. In this case we obtain by explicit calculation the following expressions for the tower of cochains (14), corresponding to the values $n=2$ and 3:

$\underline{n=2:}$

$$\Omega^0_4(A) = -(1/8\pi^2)\operatorname{Tr} F^2, \quad (18a)$$

$$\Omega^1_3(A;g) = \frac{1}{4\pi^2} d \int_{C_1} \operatorname{Tr}(\omega(t)F), \quad (18b)$$

$$\Omega^2_2(A;g_1,g_2) = \frac{1}{4\pi^2} \int_{C_2} \delta \operatorname{Tr}(\omega(t)F) - \frac{1}{8\pi^2} d \int_{C_2} \operatorname{Tr}(\omega(t)D_A\omega(t)). \quad (18c)$$

$\underline{n=3:}$

$$\Omega^0_6(A) = -(i/48\pi^3)\operatorname{Tr} F^3, \quad (19a)$$

$$\Omega^1_5(A;g) = \frac{i}{16\pi^3} d \int_{C_1} \operatorname{Tr}(\omega(t)F^2), \quad (19b)$$

$$\Omega^2_4(A;g_1,g_2) = -\frac{i}{16\pi^3} d \int_{C_2} \operatorname{Tr}(\omega(t)D_A\omega(t)F) + \frac{i}{16\pi^3} \int_{C_2} \delta \operatorname{Tr}(\omega(t)F^2), \quad (19c)$$

$$\Omega^3_3(A;g_1,g_2,g_3) = -\frac{i}{48\pi^3} d \int_{C_3} \operatorname{Tr}(\omega(t)^3F - \omega(t)D_A\omega(t)D_A\omega(t)) - \frac{i}{16\pi^3} \int_{C_3} \delta \operatorname{Tr}(\omega(t)D_A\omega(t)F). \quad (19d)$$

Here the notation $\omega(t) = \delta g(t)g^{-1}(t)$, $D_A = d + [A,\ ]$, $A = A_{t=0}$, $F = F_{t=0}$ has been used together with the abbreviation $t = (t_1,t_2,\ldots,t_k)$ for the set of parameters of C_k. These expressions were not obtained in Ref. 10 in any detail and will be useful in our application to monopole physics. (For the purposes of four-dimensional theories we do not need towers with $n \geq 4$ —for them similar expressions are available.[12])

A remarkable property of the monopole sector of Yang–Mills theories is that group elements that do not belong in the identity component of the gauge group can be reached continuously from the identity by an "allowed" gauge transformation in the sense that it does not change the values of the fields at infinity. To put it differently, in the presence of monopoles there are classically allowed motions with nonzero Pontryagin number (dyons) interpolating between (otherwise) topologically distinct gauge transformations.[3] This property is responsible for the Witten effect that permits dyons with fractional electric charge to occur and the Callan–Rubakov effect arising in fermion–dyon dynamics. Notice that the generalized cohomology we have considered in this section is defined over the entire space of gauge potentials and hence is appropriate for incorporating such features of monopole physics. In particular, Eqs. (18) and (19) are valid even if the defining simplices are not entirely contained in the identity component of a gauge group orbit but rather are taken to interpolate between disconnected components: in this case $g(t)$ is an allowed gauge transformation in the sense mentioned above.

III. INTERPOLATING DYON CONFIGURATIONS

We shall now describe in more detail the way that dyons can interpolate between topologically disconnected components of the gauge group $\mathscr{G}$ and apply the results of the previous section to obtain their magnetic charge. For simplicity, this will be demonstrated for the Julia–Zee dyon[13] arising from the symmetry breaking $SU(2) \to U(1)$ in four dimensions; however, the treatment generalizes to more realistic GUT models with breaking $G \to H$, whose monopoles are obtained by embedding $SU(2)$ in G.[14]

The Julia–Zee dyon solution is a static solution A^s_μ of the SU(2) Yang–Mills–Higgs system whose gauge potential has the form

$$(A^i_a)^s(\mathbf{r}) = \epsilon^{aij}\hat{r}^j[(K(r)-1)/r];\quad (A^0_a)^s(\mathbf{r}) = \hat{r}^a[J(r)/r], \tag{20}$$

while the Higgs field $\phi^a(\mathbf{r}) = \hat{r}^a[H(r)/r]$. Here $i,j = 1,2,3$ are spatial indices, $a = 1,2,3$ are SU(2) group indices and $\hat{r}$ denotes the unit vector in the radial direction. The radial scalar functions $K(r), J(r), H(r)$ satisfy the system of differential equations given in Ref. 13 together with their boundary conditions. In particular, $\lim_{r\to\infty} K(r) = 0$, $\lim_{r\to\infty}[J(r)/r] = M$ (constant) and similarly for $H(r)$. Then, for the static solution we have $(A^i_a)^s \underset{r\to\infty}{\to} O(r^{-1})$ and $(A^0_a)^s \underset{r\to\infty}{\to} M \neq 0$. Christ and Jackiw gave an alternative description of this solution by introducing a time-dependent gauge field configuration that is periodic in time up to a gauge transformation.[3] Let us recall the salient features of their construction.

Consider the Yang–Mills configurations $A_\mu(\mathbf{r},t)$ that are regular everywhere and periodic in time, with period T, up to a gauge transformation

$$A^\mu(\mathbf{r},t+T) = \text{gauge transformed } A^\mu(\mathbf{r},t),\quad \mu = 0,1,2,3. \tag{21}$$

Moreover, we consider physically static field configurations with t dependence described infinitesimally by

$$\frac{\partial A^a_\mu(\mathbf{r},t)}{\partial t} = -D_\mu\omega^a(\mathbf{r},t). \tag{22}$$

Here D_μ is the covariant derivative and the behavior of the fields at infinity is taken to be

$$A_\mu \underset{r\to\infty}{\to} O(r^{-1}),\quad F_{\mu\nu} \underset{r\to\infty}{\to} O(r^{-2}),\quad \omega \underset{r\to\infty}{\to} (2\pi/T)\tilde{\omega}(\hat{r}), \tag{23}$$

where $\tilde{\omega}(\hat{r})$ is covariantly constant (and hence induces the same topological behavior at infinity as the Higgs field). Upon integration of (22) we obtain (here T denotes time ordering)

$$G(\mathbf{r},t) = \mathrm{T}\exp\left(\int_0^t \omega(\mathbf{r},t)dt\right), \tag{24}$$

and the Julia–Zee static solution $(A^\mu)^s(\mathbf{r})$ is then described in terms of everywhere[15] regular potentials $A_\mu(\mathbf{r},t)$ (Ref. 3):

$$A^s_\mu(\mathbf{r}) = G^{-1}(\mathbf{r},t)A_\mu(\mathbf{r},t)G(\mathbf{r},t) + G^{-1}(\mathbf{r},t)\partial_\mu G(\mathbf{r},t). \tag{25}$$

We first note that the falloff rates of the fields at infinity (23) are of the same type as given by (9) and moreover the topological twistings of the scalar field $\omega^a(\mathbf{r},t)$ (introduced in developing the formalism) at spatial infinity (S^2_∞) are the same as that of the Higgs fields ϕ^a; this gives rise to noncontractible two-spheres in the Higgs vacuum, $SU(2)/U(1) = S^2$ of the theory and hence describes (equivalently) the monopole solution (20). It is within this spirit that Yang–Mills dyons are solely described in terms of gauge fields in Ref. 3 with no reference to Higgs scalars. Second, notice that although $(A^0)^s(\mathbf{r}) \underset{r\to\infty}{\to} M\hat{r}$, for the time-dependent configuration we have $A^0(\mathbf{r},t) \underset{r\to\infty}{\to} O(r^{-1})$ and in fact the period of the gauge transformation in (21) is related with M by $M\cdot T = 2\pi$. The (t-independent) choice for ω, $\omega^a(\mathbf{r},t) = (A^0_a)^s(\mathbf{r})$ amounts to the $A^0(\mathbf{r},t) = 0$ temporal gauge choice, as follows immediately from (25):

$$(A^\mu_a)^s(\mathbf{r}) = A^\mu_a(\mathbf{r},0) + \delta_{\mu,0}\omega^a(\mathbf{r},0). \tag{26}$$

[In the limit $T\to\infty$, the 't Hooft–Polyakov (electrically neutral) monopole solution is obtained[3].]

Gauging A^0 to zero, we can describe the dyon solution in $\mathscr{A}^{(3)}$, the space of all spatial gauge connections, not by one single configuration but rather with a whole one-parameter family of gauge potentials:

$$A_i(\mathbf{r},t) = G(\mathbf{r},t)A^s_i(\mathbf{r})G^{-1}(\mathbf{r},t) + G(\mathbf{r},t)\partial_i G^{-1}(\mathbf{r},t), \tag{27}$$

where $A^s_i(\mathbf{r}) = A_i(\mathbf{r},0)$ [cf. (26)]. This string of gauge potentials in $\mathscr{A}^{(3)}$ is periodic in t up to a gauge transformation:

$$A_i(\mathbf{r},t+T) = \mathscr{G}(\mathbf{r},t)A_i(\mathbf{r},t)\mathscr{G}^{-1}(\mathbf{r},t) + \mathscr{G}(\mathbf{r},t)\partial_i\mathscr{G}^{-1}(\mathbf{r},t), \tag{28}$$

where the identification $\mathscr{G}(\mathbf{r},t) = G(\mathbf{r},t+T)G^{-1}(\mathbf{r},t)$ has been used. The advantage of the picture presented here is that it provides the description of the Julia–Zee dyon solution within the canonical formalism of gauge theories (where the temporal gauge choice has been made). Due to the falloff rates (23) for the gauge fields at infinity, the topology of the three-dimensional space on which the theory is defined is that of a disk D_3 with $\partial D_3 = S^2_\infty$. Apparently, there is only one $G = SU(2)$-principal fiber bundle over D_3 whose space of connections is $\mathscr{A}^{(3)}$; however, the group of spatial gauge transformations $\mathscr{G}^{(3)}$ has more than one connected component—in fact, $\pi_0(\mathscr{G}^{(3)}) = Z$. Hence the gauge group orbits in $\mathscr{A}^{(3)}$ have Z-disconnected components characterized by the winding numbers of the gauge group elements.

In describing the dyon solution within the canonical formalism, one is led to examine the winding number of $G(\mathbf{r},t) \equiv G_t(\mathbf{r})$, the one-parameter family of elements entering in (27). For $t = 0$ the identity element of $\mathscr{G}^{(3)}$ results (which obviously has winding number 0), while for $t = T$ the winding number of $G_{t=T}(\mathbf{r}) \in \mathscr{G}^{(3)}$ can be calculated and found to be 1.[3] However, for all the intermediate values of the parameter $0 < t < T$, $G_t(\mathbf{r})$ approaches a direction dependent constant at spatial infinity given by

$$G_t(\mathbf{r}) \underset{r\to\infty}{\to} \exp[(2\pi/T)t\tilde{\omega}(\hat{r})]. \tag{29}$$

At this point recall that the group of spatial gauge transformations $\mathscr{G}^{(3)}$ is conventionally defined to be the set of smooth maps $D_3 \to G$ that have a direction-independent value on the boundary $\partial D_3 = S^2_\infty$. Thus $G_{0<t<T}(\mathbf{r}) \notin \mathscr{G}^{(3)}$ and so they have no definite winding number: the existence of such transformations in the monopole sector of gauge theories is a consequence of the boundary conditions (23). In the presence of monopoles these allowed gauge transformations provide a homotopy between (otherwise) topologically distinct elements of $\mathscr{G}^{(3)}$ characterized by different winding numbers.

It is by now clear that the one-parameter family of periodic (up to a gauge transformation) potentials (27) describing the Julia–Zee dyon in the Hamiltonian framework interpolates continuously between disconnected components of a gauge group orbit in $\mathscr{A}^{(3)}$ (see Fig. 1). In physical terms this means that when there are persistent boundary effects in the theory, the vacuum angle θ dependence does not arise as an instanton (quantum) tunnelling effect but is rather of leading order semiclassically. This effect was first noticed by Witten[3] who showed that dyons can have fractional electric charge proportional to the strength of *CP* violation and subsequently generalized to chromodyons (see Ref. 16 for a review).

In developing the cohomology theory appropriate for incorporating persistent boundary effects in gauge theories, we defined (in Sec. II) generalized towers of cochains over the entire space of gauge connections and have shown that upon restriction of the vertices of C_k's along a gauge group orbit the gauge group coboundary operator emerges. However, there is no need for the simplices to be confined within one connected component of the gauge group orbit, and they can also be taken to interpolate between different disconnected components (as far as the whole construction takes place in the space of gauge potentials). Consequently, one may naturally accommodate interpolating dyon configurations in such generalized towers of cochains and ask for the physical relevance of the resulting cocycles.

We shall show that the magnetic charge of dyons can be obtained by using the results of the previous section. Since we have been describing the Hamiltonian formulation of the Julia–Zee dyon it is appropriate to consider [cf. (18b)],

$$\alpha_1(A;g) = 2\pi \int_{D_3} \Omega_3^1 = \frac{1}{2\pi} \int_{D_3} \int_{C_1} d\,\mathrm{Tr}(\omega(t)F), \quad (30)$$

where $\omega(t) = \delta g(t) g^{-1}(t)$. Let us further consider the one-simplex C_1 to be defined by the dyon configuration (27) in $\mathscr{A}^{(3)}$, i.e., $A = A^s(\mathbf{r}) = A(\mathbf{r},0)$ and $g(t) = G^{-1}(t)$ [for the vertex g in (30) we have $g = G^{-1}(T)$]. Then

$$\alpha_1(A;g) = -\frac{1}{2\pi} \int_{D_3} \int_{C_1} d\,\mathrm{Tr}(\delta G(t) G^{-1}(t) F(t)),$$

where $F(t) = G(t) F G^{-1}(t)$. But $\delta G(t) G^{-1}(t) = \omega(\mathbf{r},t)$

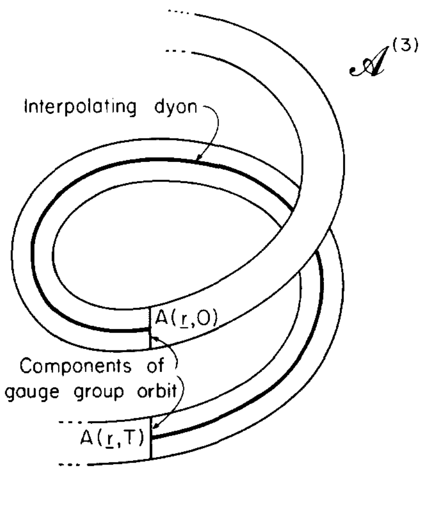


FIG. 1. Schematic picture for dyon interpolating configuration in $\mathscr{A}^{(3)}$.

and so we obtain

$$\alpha_1(A;g) = -\frac{1}{2\pi} \int_{S^2_\infty} \int_{C_1} \mathrm{Tr}(\omega(\mathbf{r},t)F(t)).$$

At spatial infinity $\omega(\mathbf{r},t) = (2\pi/T)\tilde{\omega}(\hat{\mathbf{r}})$ and moreover the integrand is time independent (as straightforward calculations show). We can then perform the t integration along C_1 $(0 \leqslant t \leqslant T)$ and arrive at the T-independent result:

$$\alpha_1(A;g) = -\int_{S^2_\infty} d\mathbf{S}\cdot\mathrm{Tr}(\mathbf{B}(\hat{r},t)\tilde{\omega}(\hat{r})), \quad (31)$$

where $\mathbf{B}(\hat{r},t)$ is the magnetic field at infinity. Let p denote the magnetic charge of the dyon; then $\alpha_1(A;g) = -2\pi p$ for the dyon configuration. This is a gauge invariant result that is consistent with the property (16) of the cochains. Moreover, $\alpha_1(A;g)$ satisfies the one-cocycle condition $\Delta\alpha_1 = 0$ in gauge group cohomology. Indeed, for any $g' \in \mathscr{G}^{(3)}$ we have

$$\Delta\alpha_1(A;g,g') = 2\pi \int_{D_3} (-d\Omega_2^2) = -2\pi \int_{S^2_\infty} \Omega_2^2$$

which is zero because at spatial infinity the two-simplex C_2 that defines Ω_2^2 degenerates to a one-simplex. This provides an example of a global gauge group cocycle in the sense that group elements from different disconnected components have beeen considered. Clearly such cocycles possess no infinitesimal (i.e., Lie algebra) analog and yet they have physical relevance in spite of the fact that they are defined over the space of all gauge connections (which is topologically trivial).

IV. THREE-COCYCLES

Physical gauge potentials are the arguments of quantum wave functionals $\Psi(A)$ and the boundary conditions they satisfy at infinity determine to which sector they belong. In particular, potentials satisfying (9) are separated by an infinite energy barrier from those satisfying (7) and hence reside in a topologically distinct sector of the Hilbert space—in our case the monopole sector. For this reason, towers of cochains descending from the Chern–Simons secondary characteristic class have to be taken into account since the choice $A' = 0$ in (10) [that reduces the generalized tower (14) to the one described in Sec. I—appropriate for the vacuum sector] is not allowed physically in the presence of monopoles.

In describing dyons within the (canonical) Hamiltonian framework, the existence of the transformations (28) led to the interpolating configuration picture between disconnected components of the orbits of $\mathscr{G}^{(3)}$ in $\mathscr{A}^{(3)}$. It has been shown in Ref. 17 that the constraints of classical Yang–Mills theories form a $\mathscr{G}^{(3)}$-current algebra even in the presence of monopoles, provided that the appropriate boundary terms are included in their expressions. It is then natural to inquire whether other members of the generalized towers considered in Sec. II play any role in the Hamiltonian description of the monopole sector of gauge theories. Apart from Ω_3^1, which has already been investigated, the next cochain defined over a three-dimensional space is Ω_3^3, which arises in the $n = 3$ tower:

$$\Omega_3^3 = -\frac{i}{48\pi^3} \int_{C_3} \mathrm{Tr}(\delta A(t_1,t_2,t_3))^3. \quad (32)$$

When the vertices of the three-simplex C_3 belong in a gauge group orbit in $\mathscr{A}^{(3)}$ the cochain $\alpha_3(A;g_1,g_2,g_3) = 2\pi \int_{D_3} \Omega_3^3$ satisfies the three-cocycle condition $\Delta\alpha_3 = 0$ in gauge group cohomology. As such, it may occur mathematically in

$$(U(g_1)U(g_2))U(g_3) = e^{-i\alpha_3(A;g_1,g_2,g_3)}U(g_1)(U(g_2)U(g_3)), \quad (33)$$

where $g_1, g_2, g_3 \in \mathscr{G}^{(3)}$ and thus determines when associativity is abandoned and/or imposes consistency conditions to the theory when $\alpha_3(A;g_1,g_2,g_3) = 2\pi n$, $n \in Z$. If a three-cocycle arises in a theory it obstructs severely its quantum mechanical description since the notion of operators and Hilbert space is not compatible with (33). In this case exotic effects like pure quantum states evolving to mixed states (and others) can occur and so an alternative description formulated in terms of nonassociative algebras is needed—though lacking at the moment (see Ref. 18 for a recent review on nonassociative algebras and representations). It was along the lines of (nonassociative) Jordan algebras that similar effects arising in the quantum theory of black holes were considered in Ref. 19, though no cohomological characterization was presented.

In the monopole sector of gauge theories three-cocycles in the $\mathscr{G}^{(3)}$-gauge group cohomology may arise and when the vertices of C_3 are taken to be in a gauge group orbit, they assume the form (19d). At this point notice that for $G = \mathrm{SU}(2)$, the three-cocycle vanishes identically (and so does the entire tower for $n = 3$). This follows from the fact that for any representation of SU(2) its generators T_a ($a = 1,2,3$) satisfy

$$d_{abc} := \mathrm{Tr}(T_a(T_bT_c + T_cT_b)) = 0, \quad (34)$$

i.e., SU(2) is a safe group.[20] Consequently, when examining the occurrence of three-cocycles in the presence of monopoles one has to consider (more general) GUT monopoles arising in the symmetry breaking $G \to H$ of an unsafe group G. GUT monopoles and/or dyons have been discussed in Ref. 14 by taking appropriate embeddings of SU(2) in G.

Let us first derive a simplified expression for the three-cocycle $\alpha_3(A; g_1,g_2,g_3)$ entering in (33). Upon integration of Ω_3^3 over D_3 the first term in (19d) is equal to a suface integral that vanishes for the following reason: for any elements g_1, g_2, g_3 of $\mathscr{G}^{(3)}$ the defining three-simplex C_3 degenerates to a single path since at spatial infinity S^2_∞ (over which the first integral is defined) the direction independent (constant) limit for the group elements is attained.[21] However, the second term in (19d) results in a (generally) nonvanishing expression when unsafe groups G are used (or at least an unsafe representation is taken for them). So

$$\alpha_3(A;g_1,g_2,g_3) = -\frac{i}{8\pi^2}\int_{D_3}\int_{C_3} \delta\,\mathrm{Tr}(\omega(t)D_A\omega(t)F), \quad (35)$$

in the notation of Sec. II. Further manipulations yield the following expression for the three-cocycle:

$$\alpha_3(A;g_1,g_2,g_3) = -\frac{i}{8\pi^2}\int_{D_3}\int_{\partial C_3} \mathrm{Tr}(\nu(t)A_t\nu(t)F_t), \quad (36)$$

where $t = (t_1,t_2,t_3)$ parametrizes C_3 and $\nu(t) = g^{-1}(t) \times \delta g(t)$, while $A_t = g^{-1}(t)Ag(t) + g^{-1}(t)dg(t)$. In writing (36) we have made use of the (Stoke) identity $\int_{C_3}\delta = \int_{\partial C_3}$, which in turn implies that $\alpha_3(A;g_1,g_2,g_3)$ is a coboundary in gauge group cohomology. In the Hamiltonian framework, magnetic source effects are taken into account by considering dyon interpolating configurations between disconnected components of a gauge group orbit in $\mathscr{A}^{(3)}$; so the three-cocycle $\alpha_3(A;g_1,g_2,g_3)$ can successfully account for the presence of monopoles only when C_3 is an interpolating simplex as well. For this we may choose g_2, g_3 to be elements of the identity component of $\mathscr{G}^{(3)}$ and g_1 to be a large gauge transformation, where $g(T,0,0) = g_1$ and $g(0 \leqslant t \leqslant T, 0, 0)$ is the one-parameter family of allowed gauge transformations that decribe the dyon. Hence α_3 is a global cocycle in gauge group cohomology possessing no infinitesimal analog; it is also gauge invariant and so is a physically relevant quantity when it is not zero. It is appropriate to mention at this point that the global cocycle (36) is only formally trivial, as a group cocycle, since C_3 is not entirely confined within a single connected component of the gauge group orbits in $\mathscr{A}^{(3)}$.

It is instructive to compare this construction of a field theoretic cocycle with the three-cocycle that arises in finite-dimensional systems when Dirac monopoles are present. It has been shown by various methods[22] that the translation group $T(n)$ in R^n picks up a three-cocycle that measures the magnetic flux through the faces of a suitably defined three-simplex enclosing the monopole. This cocycle is also trivial in the $T(n)$-group cohomology. However, it is physically meaningful since it is U(1)-gauge invariant and, in order to restore associativity, the Dirac quantization condition on the magnetic charge is enforced.

The three-cocycle we have constructed occurs in the Hamiltonian formulation of gauge theories when monopoles (associated with the symmetry breaking of unsafe structure groups G) are present. Of course, when chiral fermions are coupled to gauge fields, it is known that for unsafe groups gauge anomalies occur that obstruct the quantum mechanical implementation of Gauss's law.[6] Three-cocycles could provide further obstructions to the consistent quantum mechanical treatment of the theory in its monopole sector. The existence of the transformations (28), which have played a crucial role in our investigation, is also the origin of the Witten effect for SU(2) dyons. It persists for more general GUT monopoles and in fact leads to the color problem, as was examined in Ref. 5. It would be interesting to know whether the three-cocycle we have obtained has any relation with this problem.

V. CONCLUSIONS AND DISCUSSION

We have presented a general framework for incorporating boundary effects into the gauge group cohomology appropriate for studying Yang–Mills theories in (nontrivial) monopole sectors. The members of the generalized towers of cochains considered are all globally defined and gauge-invariant quantities over the entire space of gauge connections. We focus our attention to one- and three-cocycles in discussing the Hamiltonian formulation of Yang–Mills dyons. In this case, the existence of allowed gauge transformations that interpolate between (otherwise) disconnected components of a gauge group orbit in $\mathscr{A}$ provided a way of taking into account the presence of monopoles in gauge group cohomology and led to the notion of global cocycles. The Witten effect (and its generalizations to bigger structure groups) constitutes the core of the present work as it implies the existence of interpolating (dyon) configurations in the Hamiltonian formalism of gauge theories.

This has also given rise to a series of interesting developments in monopole physics; in particular, the inclusion of fermions has led to the Callan–Rubakov effect, while its extension to chromodyons gave rise to the color problem in GUT models. It is remarkable that in the last case the presence of a fundamental monopole obstructs the integrability of the unbroken local symmetries (in $G \to H$) into well-defined global symmetries. The simplest example is provided by the following model: $SU(3) \to U(2) \approx [SU(2) \times U(1)]/Z_2$; it is only for monopoles with magnetic charge k mod 2, $k \in Z$, that "color" (in this case the isospin) is well-defined globally.[5] This problem manifests only in global considerations and is absent if and only if the quantization condition k mod N is assumed for the magnetic charge of the monopoles (the integer N depends on the model considered). It is plausible that our global three-cocycle provides a cohomological derivation of this quantization condition, thus restoring associativity. Work towards this direction is in progress and will be reported elsewhere.[23]

Finally, we would like to conclude this work by comparing our results with the ones that have recently appeared in the literature. In Ref. 24 it was argued that three-cocycles can be described by using the conventional towers of cochains (appropriate for the vacuum sector of the theory) and the expression obtained has the form of an integral over spatial infinity. In our exposition it was shown that generalized towers of cochains have to be taken into account (thus compensating for boundary effects) and moreover all surface terms for $(k \geqslant 2)$-cocycles vanish identically as at spatial infinity the defining simplices C_k become degenerate.

ACKNOWLEDGMENTS

We wish to thank Professor Y. -S. Wu for many useful conversations on gauge group cohomology and his patience to listen to our ideas. Also we have both benefited from conversations and correspondence with Professor J. Stasheff. This work was initiated while both authors were at the University of Utah and completed while one of us (I.B.) was visiting the University of Glasgow.

Partial financial support was provided by the National Science Foundation Grants No. PHY85-03653 as well as No. PHY84-04931 (to I.B.).

[1] R. Jackiw, in *Relativity, Groups and Topology II*, Les Houches, edited by B. De Witt and R. Stora (North-Holland, Amsterdam, 1984).
[2] A. Niemi and G. Semenoff, Phys. Rep. **135**, 99 (1986), and references therein.
[3] E. Witten, Phys. Lett. B **86**, 283 (1979); N. Christ and R. Jackiw, *ibid.* **91**, 228 (1980); R. Jackiw, Acta Phys. Austriaca, Suppl. **XXII**, 383 (1980).
[4] C. Callan, Phys. Rev. D **26**, 2058 (1982), and references therein.
[5] P. Nelson and A. Manohar, Phys. Rev. Lett. **50**, 943 (1983); A. Balachandran, G. Marmo, N. Mukunda, J. S. Nilsson, E. C. G. Sudarshan, and F. Zaccaria, *ibid.* **50**, 1553 (1983).
[6] B. Zumino, Nucl. Phys. B **253**, 477 (1985), and references therein, E. Witten, *ibid.* **223**, 422 (1983); L. Faddeev, Phys. Lett. B **145**, 81 (1984); R. Stora, in "Progress in gauge field theory," Cargese Summer School (1983).
[7] I. Bakas and D. McMullan, in the *Proceedings of the XVth International Colloquium on Group Theoretical Methods in Physics*, edited by R. Gilmore (World Scientific, Singapore, 1987); I. Bakas, Invited talk at the 6th Annual Western States Mathematical Physics Meeting (Caltech, 1987) (unpublished).
[8] S. Chern and J. Simons, Ann. Math. **99**, 48 (1974).
[9] J. Manes, R. Stora, and B. Zumino, LBL-19138 preprint, 1985.
[10] Y. -S. Wu, Phys. Lett. B **153**, 70 (1985); H. -Y. Guo, in *Anomalies, Geometry and Topology*, edited by W. Bardeen and A. White (World Scientific, Singapore, 1985).
[11] In general the gauge group might have many disconnected components, e.g., for $G = SU(N)$ and $M =$ three-space, we have $\pi_0(\mathscr{G}) = \pi_3(SU(N)) = Z$.
[12] I. Bakas and D. McMullan (unpublished).
[13] B. Julia and A. Zee, Phys. Rev. D **11**, 2227 (1975).
[14] C. Dokos and T. Tomaras, Phys. Rev. D **21**, 2940 (1980).
[15] The reason for introducing time periodicity up to a gauge transformation (rather than just periodic potentials) is to allow for everywhere regular potentials including the origin $r = 0$.
[16] S. Coleman, in *3rd Annual ICTP Summer Workshop on Particle Physics*, Trieste, 1983.
[17] S. Wadia, Phys. Rev. D **15**, 3615 (1977); R. Benguria, P. Cordero, and C. Teitelboim, Nucl. Phys. B **122**, 61 (1977).
[18] Yu. Bakhturin, A. M. Slinko, and I. P. Shestakov, J. Sov. Math. **18**, 169 (1982).
[19] P. Townsend, "The Jordan formulation of quantum mechanics: A review," DAMTP preprint.
[20] H. Georgi and S. Glashow, Phys. Rev. D **6**, 429 (1972).
[21] A similar argument shows that all $\alpha_k(A; g_1, \dots, g_k)$ obtained by integrating Ω^k_{2n-k} over space M^{2n-k} obey the cocycle condition $\Delta\alpha_k = 0$ even if $\partial M^{2n-k} \neq 0$.
[22] R. Jackiw, Phys. Rev. Lett. **54**, 159 (1985); B. Grossman, Phys. Lett. B **152**, 93 (1985); Y. -S. Wu and A. Zee, *ibid.* **152**, 98 (1985); M. Gunaydin and B. Zumino, LBL-19200 preprint, 1985.
[23] I. Bakas and D. McMullan, in preparation.
[24] Y. -Z. Zhang, Phys. Lett. B **189**, 149 (1987).

Free bosonic string field theory without supplementary fields

Franz Embacher
Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria

(Received 2 February 1988; accepted for publication 23 March 1988)

A covariant local action for free bosonic string fields is constructed without the use of supplementary fields. The open string case is treated in detail. Up to a mathematical conjecture that is likely to hold it is shown that the Virasoro constraints arise as a special choice of gauge. The kinetic operator turns out to be extremely simple, the gauge transformation law arising rather implicitly. The case of closed strings is briefly discussed.

I. INTRODUCTION

Covariant bosonic string field theory interprets the first-quantized wave functionals of the free string dynamics as classical fields that have to be second-quantized along the lines of quantum field theory. Of course, the wave functionals necessary to describe the first-quantized string are not unique (for reviews of bosonic string theory see Refs. 1 and 2). BRS(T) and related techniques give rise to a field spectrum including ghost fields in addition to the pure string functional generated by the oscillator modes and the center of mass motion. Several formulations of bosonic string field theory have been given within this framework[3–9] or using pure geometric reasoning.[10]

Here we approach the subject along the lines of the old covariant quantization, starting from the Virasoro conditions, which select physical states and are usually formulated in the open or closed string Fock space. In the Appendix, the relevant notation is summarized for the case of free open strings, and some mathematical preliminaries are given.

The situation we encounter is at first sight analogous to the covariant quantization of the free massive point particle whose first-quantized wave equation

$$(\Box - m^2)\phi(x) = 0 \tag{1.1}$$

may be interpreted as selecting the physical states contained in the total space of functions $\phi(x)$. The first step in second-quantizing the Klein–Gordon field consists of finding a classical action functional that reproduces (1.1) as its field equation. Interactions are then incorporated by adding suitable nonquadratic terms to the free Lagrangian.

In the theory of open bosonic strings, the analog of (1.1) is the Virasoro conditions

$$L_n|\psi\rangle = 0 \tag{1.2}$$

($n \geqslant 1$), together with the mass shell equation

$$(L_0 - 1)|\psi\rangle = 0\,, \tag{1.3}$$

for elements $|\psi\rangle$ of the total Fock space. Once having interpreted $|\psi\rangle$ as a classical field, subject to the above equations, the next step is to look for an action functional reproducing these.

Usually, one is interested in a gauge-covariant formulation of string field theory. In most of the approaches this means that the conformal generators L_{-n}—which give rise to reparametrizations of the string world sheet—should also provide the gauge symmetries

$$\delta|\psi\rangle = L_{-n}|\Lambda\rangle \tag{1.4}$$

($|\Lambda\rangle$ unconstrained) of the action. However, this is only possible if supplementary fields are introduced and subjected to appropriate gauge transformation laws as well. The requirement then is that in a special gauge the field equations should be identical to (1.2) and (1.3) with all supplementary fields vanishing. Moreover, the action shall be local in the space-time fields contained in $|\psi\rangle$, which, in other words, means that the kinetic operator must be polynomial in the derivatives ∂_μ at each mass level.

This problem has been solved by several authors in different ways,[11–19] including the interaction of string fields[20,21] and the superstring case.[22] There are a lot of arguments in favor of the symmetries (1.4) and the introduction of supplementary fields. The structure of these was also related to the ghosts in the BRST approach.[4]

However, there exists an alternative approach. Giving up the full gauge symmetry and the supplementary fields, we will construct a local action functional for free bosonic strings, formulated in terms of the Fock-space state $|\psi\rangle$ alone. It admits a gauge symmetry considerably *smaller* than (1.4). Upon a conjecture that is very likely to hold, it is argued in the open string case that the correct physical state conditions (1.2) and (1.3) may be obtained as a gauge fixing by means of this symmetry. For the closed strings, the existence of an appropriate action functional without supplementary fields is not as firmly established as for the open strings, but we suggest a possible candidate that does the job up to the second mass level. While in the open string case the gauge transformation yielding the correct gauge fixing exists as a consequence of pure algebraic and differential manipulations, one is forced to impose appropriate boundary conditions upon the space-time fields of the closed strings.

The physical significance of the actions suggested is not yet clarified. It is likely that they may be obtained by a partial gauge fixing and elimination of supplementary fields from the fully gauge-covariant actions. However, *what* has been fixed remains unclear.

Also, the graviton is described here—in the linearized case—by the symmetric traceless tensor contained in the closed string field at the first mass level. Thus, upon including interactions, one expects that the gravitational sector will be described either in a not fully general covariant way or by some nonlocal action of the Fradkin–Vilkovisky type.[23]

 0022-2488/88/081916-14$02.50

Nevertheless, the approach presented here has the advantage of an extremely simple kinetic operator. The price paid for this is—in addition to the remarks made above—a quite inexplicit form of the gauge transformation law. In their final form, the action functionals suggested are identical to those given in Ref. 15 for the free open and closed string fields when all supplementary fields are set equal to zero. This observation might provide a hint how to treat the superstring fields analogously.

The paper is organized as follows. Sections II–VIII are concerned with the free field theory of bosonic open strings. In Sec. II, the well-known action functionals for the mass levels 0 and 1 are reviewed. Section III is devoted to the construction of a Lagrangian for the fields occurring at the mass level 2. The computation is first carried out explicitly in terms of the space-time fields and then condensed in the Fock-space notation. The necessary gauge symmetry is exhibited. In Sec. IV, the kinetic operator, which should apply for all mass levels, is presented. The existence of an infinite series of gauge transformations is proved in Sec. V. In this formulation, the gauge parameters are constrained to lie in the space defined by Eq. (1.2). In Sec. VI, a mathematical conjecture concerning the kinetic operator is made. Under the assumption that it holds, it is shown that the gauge symmetries are in fact enough to admit the gauge fixing (1.2) and (1.3). In Sec. VII, it is outlined how the infinite series of transformation laws may be unified into a single one, involving only one unconstrained element of the Fock space as gauge parameter. The residual gauge freedom that still exists after the Virasoro gauge has been fixed is mentioned in Sec. VIII. Section IX presents a candidate for an appropriate action functional in the closed string case, although the matter is not so clear here as compared to the open strings. Some concluding remarks are made in Sec. X, discussing the properties of the kinetic operator responsible for the existence of a free covariant string field theory without supplementary fields, and touching the question of whether there may be other formulations working equally well.

The most interesting questions of how the shortcomings of the formulation presented here might be overcome in the interacting case and what it tells us for the theory of superstrings are presently under consideration.

II. ACTION FUNCTIONALS FOR THE MASS LEVELS 0 AND 1

We first consider the states of zero mass level, i.e., the tachyonic field

$$|\psi\rangle \equiv |\psi\rangle_0 = \phi(x)|0\rangle . \tag{2.1}$$

Its only field equation is the mass-shell condition (A17)

$$(\Box + 2)\phi = 0 . \tag{2.2}$$

Clearly, this field equation follows from the action

$$S = \frac{1}{2}\int d^{D}x\, \phi(\Box + 2)\phi , \tag{2.3}$$

or, when written in the language of the Fock space,

$$S = -\tfrac{1}{2}\langle\psi|2(L_0 - 1)|\psi\rangle . \tag{2.4}$$

As well-known, the first mass level may be treated in an analogous way without complications. The states at the first mass level have the form

$$|\psi\rangle \equiv |\psi\rangle_1 = -iA_\mu(x)\alpha_{-1}{}^{\mu}|0\rangle , \tag{2.5}$$

the field equations being the mass-shell equation (A17)

$$\Box A^{\mu} = 0 \tag{2.6}$$

supplemented by the Virasoro conditions (A16)

$$\partial_\mu A^{\mu} = 0 , \tag{2.7}$$

which requires $|\psi\rangle$ to lie in $\mathbf{V}^{(0)}$ [cf. Eq. (A18)]. Considering the free Maxwell action

$$S = \frac{1}{2}\int d^{D}x\, A^{\mu}(\Box A_\mu - \partial_{\mu\nu}A^{\nu}) , \tag{2.8}$$

one obtains the field equations

$$\Box A_\mu - \partial_{\mu\nu}A^{\nu} = 0 . \tag{2.9}$$

The action S possesses the local gauge symmetry

$$\delta A_\mu = \partial_\mu \Lambda \tag{2.10}$$

for arbitrary $\Lambda(x)$, which implies that one can always choose (2.7) as a gauge condition and hence obtains (2.6) as the field equations. In the notation of the Fock space, S reads (cf. Ref. 13)

$$S = -\tfrac{1}{2}\langle\psi|2(L_0 - 1) - L_{-1}L_1|\psi\rangle . \tag{2.11}$$

The gauge transformation law (2.10) has the form

$$\delta|\psi\rangle = L_{-1}|\chi\rangle , \tag{2.12}$$

where

$$|\chi\rangle = \Lambda(x)|0\rangle . \tag{2.13}$$

It is important to note that checking the invariance of (2.11) under (2.12) one only has to use the fact that $|\chi\rangle$ is annihilated by L_1. No reference to the mass levels involved is needed in this computation. Moreover, the action (2.11) reduces to (2.4) if $|\psi\rangle = |\psi\rangle_0$ is inserted. Thus S provides an action for the first two mass levels simultaneously.

III. ACTION FUNCTIONALS FOR THE MASS LEVEL 2

Trying to extend the action (2.11) to include also the fields at the second mass level, one faces a well-known problem. Generalizing (2.12), one usually requires the action to be invariant under the gauge symmetry [cf. Eq. (1.4)]

$$\delta|\psi\rangle_2 = L_{-1}|\chi\rangle_1 + L_{-2}|\chi\rangle_0 , \tag{3.1}$$

where the $|\chi\rangle_n$ are arbitrary states at the mass levels n. However, it is well-known that the field content of the second mass level may not be described by a local action admitting the gauge symmetry (3.1). The action considered by Banks and Peskin,[13]

$$S = -\tfrac{1}{2}\langle\psi|2(L_0 - 1)P|\psi\rangle , \tag{3.2}$$

where P is the projector onto $\mathbf{V}^{(0)}$, turns out to become nonlocal at the second mass level as a result of the existence of zeros in the Kac determinant. (In other words, P contains rational functions of L_0.) Usually, this problem is overcome by the introduction of supplementary fields that merely serve to maintain the symmetry (3.1) *and* locality.

However, there is another possibility that apparently has not been pursued before. Insisting on an action functional that involves only $|\psi\rangle$ and is local, containing only first and second space-time derivatives, it is clear that one must sacrifice at least part of the full symmetry (3.1). In the following we will show how this works, beginning with the second mass level.

The Fock states in question are given by

$$|\psi\rangle \equiv |\psi\rangle_2 = (-iv_\mu(x)\alpha_{-2}{}^\mu - \tfrac{1}{2}h_{\mu\nu}(x)\alpha_{-1}{}^\mu\alpha_{-1}{}^\nu)|0\rangle , \tag{3.3}$$

the field equations again being the mass-shell condition (A17)

$$(\Box - 2)v_\mu = (\Box - 2)h_{\mu\nu} = 0 \tag{3.4}$$

together with the constraints (A16) [cf. Eqs. (A19)]

$$C_\mu \equiv v_\mu - \tfrac{1}{2}\partial^\rho h_{\rho\mu} = 0 , \tag{3.5}$$

$$B \equiv \partial_\mu v^\mu + \tfrac{1}{4}h^\mu_\mu = 0 . \tag{3.6}$$

The question now is whether there is an action for the fields $v_\mu, h_{\mu\nu}$ admitting enough gauge freedom to choose (3.5) and (3.6) as gauge conditions.

Let us make an ansatz for the Lagrangian

$$\begin{aligned} c\mathscr{L} = {} & \tfrac{1}{2}v^\mu(\Box - 2)v_\mu + (\lambda/2)h^{\mu\nu}(\Box - 2)h_{\mu\nu} \\ & + 2a(v^\mu v_\mu - v^\mu\partial^\rho h_{\rho\mu} - \tfrac{1}{4}h^{\mu\nu}\partial^\rho_\mu h_{\rho\nu}) \\ & + 2b(v^\mu\partial_{\mu\rho}v^\rho + \tfrac{1}{2}v^\mu\partial_\mu h^\rho_\rho - \tfrac{1}{16}h^\mu_\mu h^\rho_\rho) . \end{aligned} \tag{3.7}$$

The numerical factors within the brackets are chosen such that the field equations arising by variation are of the form

$$\mathscr{V}_\mu \equiv (\Box - 2)v_\mu + 4aC_\mu + 4b\,\partial_\mu B = 0 , \tag{3.8}$$

$$\mathscr{H}_{\mu\nu} \equiv \lambda(\Box - 2)h_{\mu\nu} + a(\partial_\mu C_\nu + \partial_\nu C_\mu) - b\eta_{\mu\nu}B = 0 . \tag{3.9}$$

The game consists now of playing around with the expressions $\mathscr{H}^\mu_\mu$, $\partial^\mu\mathscr{V}_\mu$, $\partial^\nu\mathscr{H}_{\mu\nu}$, and $\partial^{\mu\nu}\mathscr{H}_{\mu\nu}$ in order to guess the required gauge transformation law. In the case $D = 26$ [D appears as δ^μ_μ when taking the trace of (3.9)], one finds the solution

$$a = \tfrac{1}{2}, \quad b = -\tfrac{1}{4}, \quad \lambda = \tfrac{1}{4} , \tag{3.10}$$

and the two gauge symmetries

$$\delta_1(v_\mu, h_{\mu\nu}) = (\chi_\mu, \partial_\mu\chi_\nu + \partial_\nu\chi_\mu) , \tag{3.11}$$

$$\delta_2(v_\mu, h_{\mu\nu}) = (-\tfrac{1}{2}(\Box + 3)\partial_\mu\Lambda, -3\,\partial_{\mu\nu}\Lambda + \tfrac{1}{2}\eta_{\mu\nu}\Lambda) , \tag{3.12}$$

where the gauge parameter $\chi_\mu(x)$ has to satisfy

$$\partial_\mu\chi^\mu = 0 . \tag{3.13}$$

Instead of χ^μ, one could also use an unconstrained gauge parameter $\xi^\mu(x)$ and insert

$$\chi^\mu = \partial^{\mu\nu}\xi_\nu - \Box\xi^\mu \tag{3.14}$$

into (3.11). The constraint (3.13) is then automatically satisfied.

In order to see how a choice of gauge works, we consider the following combination of the field equations:

$$\partial^\mu\mathscr{V}_\mu + \mathscr{H}^\mu_\mu \equiv 3\,\partial_\mu C^\mu + \tfrac{9}{2}B = 0 . \tag{3.15}$$

This situation is reminiscent of the case of a massive vector field, where the Lorentz gauge condition follows from the Proca equation without the use of any gauge symmetry. Next we use (3.12), which implies

$$\delta_2 B = -\tfrac{1}{2}\Box(\Box - 2)\Lambda \tag{3.16}$$

to attain the gauge condition

$$B = 0 \tag{3.17}$$

from which also follows

$$\partial_\mu C^\mu = 0 . \tag{3.18}$$

The remaining gauge freedom (3.11),

$$\delta_1 C_\mu = -\tfrac{1}{2}(\Box - 2)\chi_\mu , \quad \delta_1 B = 0 , \tag{3.19}$$

does not affect (3.17). By virtue of (3.13) and (3.18) one may choose χ_μ such that $\delta_1 C_\mu = -C_\mu$, thus arriving at the final gauge condition

$$C_\mu = B = 0 \tag{3.20}$$

and the field equations (3.4).

Thus we have shown that the Lagrangian

$$c\mathscr{L} = \tfrac{1}{2}v^\mu(\Box - 2)v_\mu + \tfrac{1}{8}h^{\mu\nu}(\Box - 2)h_{\mu\nu} + C^\mu C_\mu + \tfrac{1}{2}B^2 \tag{3.21}$$

satisfies all requirements. Using (A19) and setting $c = \tfrac{1}{2}$, the action reads in the Fock notation

$$\begin{aligned} S &= \int d^Dx\,\mathscr{L} \\ &= -\tfrac{1}{2}\langle\psi|2(L_0 - 1) - L_{-1}L_1 - \tfrac{1}{2}L_{-2}L_2|\psi\rangle . \end{aligned} \tag{3.22}$$

Clearly, this form of the action also applies for the mass levels 0 and 1 [cf. Eqs. (2.4) and (2.11)]. The gauge transformation laws (3.11) and (3.12) read

$$\delta_1|\psi\rangle = L_{-1}|\chi\rangle , \tag{3.23}$$

$$\delta_2|\psi\rangle = (L_{-2}L_0 - \tfrac{3}{2}L^2_{-1})|\Lambda\rangle , \tag{3.24}$$

where

$$|\chi\rangle = -i\chi_\mu(x)\alpha_{-1}{}^\mu|0\rangle \tag{3.25}$$

and

$$|\Lambda\rangle = \Lambda(x)|0\rangle , \tag{3.26}$$

the constraint (3.13) just expressing the requirement that $|\chi\rangle$ is an element of $\mathbf{V}^{(0)}$,

$$L_1|\chi\rangle = 0 . \tag{3.27}$$

Again, the invariance of (3.22) under (3.23) and (3.24) is checked without reference to any mass level number but merely follows from the assumption that $|\chi\rangle$ and $|\Lambda\rangle$ lie in $\mathbf{V}^{(0)}$ (see the Appendix for the definitions of $\mathbf{V}^{(n)}$ and $\mathbf{W}^{(n)}$). The verification of (3.24) makes use of $D = 26$. Off the critical dimension, there is no gauge symmetry of the type (3.24).

The kinetic operator appears in the literature as a *part* of string field actions but never in the form presented here (cf. Refs. 13 and 15).

The question that will concern us in Secs. IV–VI is whether this scheme carries over to all higher mass levels. In the following section, the problem is posed in a more convenient language.

IV. ACTION FUNCTIONALS FOR THE HIGHER MASS LEVELS

From now on we will abandon the notation using component fields and rather work instead with the elements of the Fock space directly. The major ingredient for the following computations is the Virasoro algebra (A22). Following the scheme of the last section, the kinetic operator should have the form

$$\mathcal{K} = 2(L_0 - 1) - \sum_{n=1}^{\infty} a_n L_{-n} L_n , \tag{4.1}$$

the action functional

$$S = -\tfrac{1}{2}(\psi|\mathcal{K}|\psi) \tag{4.2}$$

being local and containing only first and second space-time derivatives. We know already $a_1 = 1$, $a_2 = \frac{1}{2}$. If it is possible to determine the a_n such that the Virasoro constraints (A16) arise as possible gauge-fixing conditions, (4.2) constitutes an action for *all* field components contained in a general element $|\psi\rangle$ of the Fock space.

The field equations read

$$\mathcal{K}|\psi\rangle = 0 . \tag{4.3}$$

Since

$$\mathcal{K}^\dagger = \mathcal{K} , \tag{4.4}$$

the action is invariant under some gauge symmetry if and only if the field equations (4.3) are, i.e.,

$$\mathcal{K}\delta|\psi\rangle = 0 . \tag{4.5}$$

Extending the transformation laws (3.23) and (3.24), we ask whether there exist operators

$$\mathcal{S}_{-n} = \mathcal{L}^{(n)}_{-i} a^i_{(n)}(L_0) , \tag{4.6}$$

for every $n \geqslant 1$, such that

$$\delta|\psi\rangle = \mathcal{S}_{-n}|\chi\rangle \tag{4.7}$$

satisfies (4.5) for *all* $|\chi\rangle \in \mathbf{V}^{(0)}$ (see the Appendix for the definition of $\mathcal{L}^{(n)}_{-i}$). Thereby the $a^i_{(n)}$ shall be polynomials in L_0. We already know that such operators exist for $n = 1$ and 2 if $D = 26$, namely,

$$\mathcal{S}_{-1} = L_{-1} , \tag{4.8}$$

$$\mathcal{S}_{-2} = L_{-2}L_0 - \tfrac{3}{2}L^2_{-1} . \tag{4.9}$$

In order to define operators of the type (4.6) unambiguously, we always place the L_0's to the right of the $\mathcal{L}^{(n)}_{-i}$.

Inserting (4.7) into the invariance condition (4.5), one obtains

$$\mathcal{K}\mathcal{S}_{-n}|\chi\rangle = 0 , \tag{4.10}$$

for all $|\chi\rangle \in \mathbf{V}^{(0)}$, which immediately turns out to be a system of ℓ_n linear equations for the ℓ_n unknown $a^i_{(n)}$ (ℓ_n being the number of different $\mathcal{L}^{(n)}_i$'s). Using the ansatz (4.6) and commuting the operators L_p occurring in $\mathcal{K}$ to the right, one finds that

$$\mathcal{K}\mathcal{L}^{(n)}_{-i}|\xi\rangle \equiv \mathcal{L}^{(n)}_{-j}\mathcal{K}^{(n)j}{}_i(L_0)|\xi\rangle , \tag{4.11}$$

for any $|\xi\rangle \in \mathbf{V}^{(0)}$. The $\mathcal{K}^{(n)j}{}_i$ are polynomials (of degree 1) in L_0. Inserting $a^i_{(n)}(L_0)|\chi\rangle$ for $|\xi\rangle$ and using the requirement that the resulting expression vanishes for all $|\chi\rangle \in \mathbf{V}^{(0)}$, one obtains the system of equations

$$\mathcal{K}^{(n)j}{}_i(h)a^i_{(n)}(h) = 0 , \tag{4.12}$$

which has to be valid for all real numbers h. To be more precise, one could insert a state of type (A26) into (4.10) and deduce (4.12) for almost all values of h (and hence for all since we are dealing with polynomials) as indicated in the Appendix. Clearly, if (4.12) holds, then also (4.10) is valid for any $|\chi\rangle \in \mathbf{V}^{(0)}$, irrespectively of $|\chi\rangle$ being an eigenvector of L_0 or not.

The first 1×1 matrix $\mathcal{K}^{(1)}$ turns out to be identically zero. This agrees with the fact that—because of (4.8)—$\mathcal{K}$ annihilates every element of $\mathbf{V}^{(1)}$, for any value of the space-time dimension D. Counting the $\mathcal{L}^{(2)}_i$ as

$$\mathcal{L}^{(2)}_1 = L_2, \quad \mathcal{L}^{(2)}_2 = L^2_1 , \tag{4.13}$$

one finds

$$\mathcal{K}^{(2)i}{}_j(h) = \begin{pmatrix} 2-\frac{1}{4}D & -3h \\ -3 & -2h \end{pmatrix} \tag{4.14}$$

for general D. Its determinant is given by

$$\det \mathcal{K}^{(2)}(h) = \tfrac{1}{2}h(D-26) . \tag{4.15}$$

This illustrates that off the critical dimension, there is no solution $a^i_{(2)}(h)$. If $D = 26$, one reads off from (4.9)

$$a^i_{(2)}(h) = \begin{pmatrix} h \\ -\frac{3}{2} \end{pmatrix} . \tag{4.16}$$

We adopt the convention that the $a^i_{(n)}(h)$ have no common polynomial factor, i.e., they never vanish simultaneously. The rank of the matrix $\mathcal{K}^{(2)}$ turns out to be 1 for *all* values of h. If $h = -\frac{9}{4}$, the eigenvalue 0 appears twice in $\mathcal{K}^{(2)}$, but

$$\mathcal{K}^{(2)}\left(-\frac{9}{4}\right) = \begin{pmatrix} -\frac{9}{2} & \frac{27}{4} \\ -3 & \frac{9}{2} \end{pmatrix} \tag{4.17}$$

is nonzero and hence nondiagonalizable (i.e., there is no basis of eigenvectors). As a consequence, the polynomials (4.16), found as the *generic* solution of (4.12), in fact constitute the general solution for any h. This is responsible for the fact that the symmetries (3.23) and (3.24) admit the required gauge condition for *any* $|\psi\rangle_2$ subject to the field equation (4.3). We will return to this point later.

For the moment, we consider the system (4.12) for higher n. Counting the $\mathcal{L}^{(3)}_i$ as given in (A25), one may compute $\mathcal{K}^{(3)}$ for arbitrary a_3 and D. One finds that Eq. (4.12) admits nontrivial polynomials $a^i_{(3)}(h)$ if and only if $a_3 = \frac{1}{3}$ and $D = 26$. Leaving D arbitrary, one again arrives at the result that $\det \mathcal{K}^{(3)}$ is a polynomial in h and D, multiplied by $D - 26$. The $a^i_{(3)}(h)$ are unique up to common numerical factors, i.e., if $D = 26$, the generic rank of $\mathcal{K}^{(3)}$ is 2.

It has been checked on the computer—using the algebraic program REDUCE—that this scheme holds up to $n = 7$. Nontrivial solutions of (4.12) arise if and only if $a_n = 1/n$ and $D = 26$. For illustration, we quote here the results for the cases $n = 3$ and 4:

$$\mathcal{S}_{-3} = L_{-3}a^1_{(3)}(L_0) + L_{-2}L_{-1}a^2_{(3)}(L_0) + L^3_{-1}a^3_{(3)}(L_0) , \tag{4.18}$$

where

$$a^1_{(3)}(h) = \tfrac{1}{13}(5-8h)h\,,$$
$$a^2_{(3)}(h) = 2h+1\,, \tag{4.19}$$
$$a^3_{(3)}(h) = -\tfrac{3}{2}\,,$$

and

$$\mathcal{S}_{-4} = L_{-4}a^1_{(4)}(L_0) + L_{-3}L_{-1}a^2_{(4)}(L_0) + L^2_{-2}a^3_{(4)}(L_0) + L_{-2}L^2_{-1}a^4_{(4)}(L_0) + L^4_{-1}a^5_{(4)}(L_0)\,, \tag{4.20}$$

with

$$a^1_{(4)}(h) = -48h^4 - 174h^3 - 36h^2 - 774h\,,$$
$$a^2_{(4)}(h) = 176h^3 + 710h^2 - 200h + 130\,,$$
$$a^3_{(4)}(h) = 72h^3 + 456h^2 + 672h\,, \tag{4.21}$$
$$a^4_{(4)}(h) = -394h^2 - 2084h - 1690\,,$$
$$a^5_{(4)}(h) = 197h + 845\,.$$

Motivated by these results, we fix the kinetic operator as

$$\mathcal{K} = 2(L_0 - 1) - \sum_{n=1}^{\infty} \frac{1}{n} L_{-n}L_n \tag{4.22}$$

and conjecture that for any $n \geqslant 1$ there is a set of polynomials $a^i_{(n)}(h)$ such that (4.10) [or equivalently (4.12)] holds, provided $D=26$. The $a^i_{(n)}$ should be unique up to common factors. Moreover, we expect the sequence of gauge symmetries obtained in this way being large enough to bring any solution of the field equation (4.3) into the Virasoro gauge (A16). This last statement may equivalently be expressed as follows. Let $|\psi\rangle$ be any element of the Fock space satisfying (4.3). Then $|\psi\rangle$ is of the form

$$|\psi\rangle = |\chi_0\rangle + \sum_{p=1}^{\infty} \mathcal{S}_{-p}|\chi_p\rangle\,, \tag{4.23}$$

where all $|\chi\rangle$'s lie in $\mathbf{V}^{(0)}$. Now $|\chi_0\rangle$ is the "physical" part of $|\psi\rangle$, satisfying the mass-shell equation (A17). According to the considerations given in the Appendix, we expect the decomposition (4.23) not to be unique. The arbitrariness we are left with constitutes a residual gauge invariance of the type (4.7), which leaves the Virasoro conditions invariant. For illustration, consider a state of the form

$$|\psi\rangle = \mathcal{S}_{-1}|\chi\rangle \equiv L_{-1}|\chi\rangle\,, \tag{4.24}$$

with $|\chi\rangle \in \mathbf{V}^{(0)}$. This state obeys the field equation (4.3) and is an element of $\mathbf{V}^{(1)}$. According to (4.23), it is a pure gauge. If, however, $|\chi\rangle$ additionally satisfies

$$L_0|\chi\rangle = 0\,, \tag{4.25}$$

it follows [multiplying (4.24) by L_1] that $|\psi\rangle$ is also in $\mathbf{V}^{(0)}$. Thus $|\psi\rangle$ may equally well be interpreted as the physical contribution $|\chi_0\rangle$ in (4.23). At the first mass level this just means that the Lorentz condition (2.7) still admits the gauge freedom (2.10) with

$$\Box\Lambda = 0\,. \tag{4.26}$$

As indicated in the Appendix, the nonuniqueness of the decomposition (4.23) is connected with the zeros in the Kac determinants [$h=0$ being the unique zero of $\mathcal{M}^{(1)}(h) = 2h$, in agreement with (4.25)]. However, trying to fix the gauge completely means to give up the covariant formalism. In quantized string field theory, this is not even necessary, because the covariant gauge—together with appropriate Feynman boundary conditions—suffices to guarantee the existence of the propagator.

The fact that the gauge parameters $|\chi\rangle$ are constrained to be elements of $\mathbf{V}^{(0)}$ is not really a shortcoming of the present formalism. One can always express them in terms of unconstrained parameters along the lines of Eq. (3.14). The complicated form of the gauge transformation laws is the price we have to pay when using the simple action (4.2) and (4.22). We will return to the residual gauge freedom and the question of unconstrained gauge parameters in Secs. VII and VIII.

In the Secs. V and VI we will partly prove the conjectures made here. During the course of the investigations, the problem has been attacked by several different approaches, and it seems extremely unlikely that the assumptions that remain unproved are incorrect.

V. EXISTENCE PROOF

In this section we prove the existence of the operators $\mathcal{S}_{-n}$ for all $n \geqslant 1$. The kinetic operator is taken over from (4.22); the space-time dimension is left unspecified for the moment. Since the general form of the matrices $\mathcal{K}^{(n)i}{}_j$ arises rather implicitly, we proceed along another line of reasoning.

Making extensive use of the Virasoro algebra (A22), one may show

$$L_1\mathcal{K} = -\sum_{n=1}^{\infty} \frac{1}{n} L_{-n}\,\mathcal{A}_{n+1}\,, \tag{5.1}$$

$$L_2\mathcal{K} = \frac{1}{4}(26-D)L_2 - 3\,\mathcal{A}_2 - \frac{2}{3}\sum_{n=1}^{\infty}\frac{1}{n} \times L_{-n}\left(L_n\,\mathcal{A}_2 - L_1\mathcal{A}_{n+1} - \frac{n^2-2n-2}{n+1}\,\mathcal{A}_{n+2}\right), \tag{5.2}$$

where

$$\mathcal{A}_n = L_1L_{n-1} + [(n-1)(n+1)/n]L_n \tag{5.3}$$

is defined for $n \geqslant 2$. The computation is partly identical to that given in Ref. 15.

Next we note that if $D=26$ and a state $|\psi\rangle$ satisfies

$$\mathcal{A}_p|\psi\rangle = 0\,, \tag{5.4}$$

for all $p \geqslant 2$, it follows from (5.1) and (5.2) that $\mathcal{K}|\psi\rangle$ is annihilated by all L_n with positive n (remember that L_1 and L_2 generate all L_n). As a consequence, $\mathcal{K}|\psi\rangle$ is an element of $\mathbf{V}^{(0)}$.

For the rest of this section, we set

$$|\psi\rangle = \mathcal{S}_{-n}|\chi\rangle \in \mathbf{V}^{(n)}\,, \tag{5.5}$$

for a given n, where $\mathcal{S}_{-n}$ represents the ansatz (4.6). The equation for the polynomials $a^i_{(n)}$ thus reads

$$\mathcal{K}|\psi\rangle = 0 \tag{5.6}$$

and has to be valid for all $|\chi\rangle \in \mathbf{V}^{(0)}$. Now suppose that (5.4) holds for all $|\chi\rangle \in \mathbf{V}^{(0)}$. Since $\mathcal{K}|\psi\rangle \in \mathbf{V}^{(0)}$, as stated above,

$$\mathcal{L}^{(n)}_i\mathcal{K}|\psi\rangle = 0\,. \tag{5.7}$$

Inserting (4.6), one finds

$$\mathscr{M}_{ij}^{(n)}(L_0)\mathscr{K}^{(n)j}{}_k(L_0)a_{(n)}^k(L_0)|\chi\rangle = 0\,. \quad (5.8)$$

Since this has to be true for *all* $|\chi\rangle \in \mathbf{V}^{(0)}$—especially for eigenvectors of L_0 corresponding to generic eigenvalues h—we conclude

$$\mathscr{K}_{ij}^{(n)}(h)a_{(n)}^j(h) = 0\,, \quad (5.9)$$

where

$$\mathscr{K}_{ij}^{(n)}(h) = \mathscr{M}_{ik}^{(n)}(h)\mathscr{K}^{(n)k}{}_j(h)\,. \quad (5.10)$$

Since $\mathscr{M}_{ij}^{(n)}(h)$ is generically invertible, one arrives at Eq. (4.12). Note that in the generic sense, $\mathscr{M}_{ij}^{(n)}$ plays the role of a metric which may serve to raise and lower indices. The matrix (5.10) may also be defined as

$$\mathscr{K}_{ij}^{(n)}(h) = \langle\chi|\mathscr{L}_i^{(n)}\mathscr{K}\mathscr{L}_{-j}^{(n)}|\chi\rangle\,, \quad (5.11)$$

where $|\chi\rangle$ satisfies (A26) and $\langle\chi|\chi\rangle = 1$. Hence

$$\mathscr{K}_{ij}^{(n)}(h) = \mathscr{K}_{ji}^{(n)}(h)\,, \quad (5.12)$$

for all h.

What we have shown so far is that, if (5.4) is valid for all $|\chi\rangle \in \mathbf{V}^{(0)}$, the $a_{(n)}^i(h)$ solve Eq. (4.12). Since (5.4) is a generic equation, it is equivalent to

$$\mathscr{L}_i^{(n-p)}\mathscr{A}_p|\psi\rangle = 0\,, \quad (5.13)$$

for $2 \leqslant p \leqslant n$. These are in fact $\Sigma_{p=2}^n \ell_{n-p}$ ordinary linear equations for the ℓ_n unknowns $a_{(n)}^i$.

Studying the algebra of the operators $\mathscr{A}_p$, one discovers another set of operators $\mathscr{B}_p$ which in some respect are more convenient. From (5.3) it follows

$$L_m = \frac{m}{(m-1)(m+1)}\mathscr{A}_m - \frac{m}{(m-1)(m+1)}L_1L_{m-1}\,. \quad (5.14)$$

Again replacing L_{m-1} by means of (5.3) and iterating the procedure, one finds an expression for L_m in terms of $\mathscr{A}_p$ ($p \leqslant m$) and powers of L_1. Defining

$$\mathscr{B}_m = L_m + [2m(-)^m/(m+1)!]\,L_1^m\,, \quad (5.15)$$

for $m \geqslant 2$, the expression reads

$$\mathscr{B}_m = \sum_{p=2}^m (-)^{m-p}\frac{m}{(m+1)!}\frac{p!}{p-1}L_1^{m-p}\mathscr{A}_p\,. \quad (5.16)$$

Clearly, this relation may be inverted to express $\mathscr{A}_m$ in terms of all $\mathscr{B}_p$ ($p \leqslant m$). Thus one may substitute $\mathscr{B}_p$ for $\mathscr{A}_p$ in Eqs. (5.4) and (5.13). It is worth noting that, upon the use of the Virasoro algebra, one finds that the commutators of the $\mathscr{B}$'s close in the sense ($2 \leqslant p < m$)

$$[\mathscr{B}_p, \mathscr{B}_m] = \sum_{q=p+1}^{p+m} c_{pmq}L_1^{p+m-q}\mathscr{B}_q\,, \quad (5.17)$$

where c_{pmq} are nonzero real numbers. From this it follows that (5.4)—as it stands or with $\mathscr{B}_p$ instead of $\mathscr{A}_p$—is already valid for all p if it holds for $p = 2$, 3, and 4. All higher $\mathscr{B}_p$ or $\mathscr{A}_p$ may then be expressed in terms of the lower ones by (5.17). This observation might be useful for practical calculations of the higher $\mathscr{S}_{-n}$.

The existence proof runs as follows. Consider the set of operators $\mathscr{L}_i^{(n-p)}\mathscr{B}_p$ ($2 \leqslant p \leqslant n$). They lie in the ℓ_n-dimensional space spanned by the $\mathscr{L}_i^{(n)}$. Although they are larger in number than ℓ_n, they actually span only an $(\ell_n - 1)$-dimensional subspace. This can be seen by noting that it is not possible to find real numbers $\lambda^i_{(n-p)}$ such that

$$\sum_{p=2}^n \lambda^i_{(n-p)}\mathscr{L}_i^{(n-p)}\mathscr{B}_p = L_n\,. \quad (5.18)$$

Hence (5.13) only provides a set of $\ell_n - 1$ linear equations for ℓ_n unknowns. Since the $\mathscr{K}^{(n)i}{}_j(h)$ are of degree one in h, the $a_{(n)}^i(h)$ may be chosen as polynomials of degree $\leqslant \ell_n - 1$.

It is actually not difficult to prove the uniqueness of solutions to (5.13), which is equivalent to showing that the $\ell_n - 1$ equations mentioned above are linearly independent. Imposing (5.4)—now using the $\mathscr{A}_p$—and requiring the single additional equation

$$L_n|\psi\rangle = 0\,, \quad (5.19)$$

one may check that the only solution is the trivial one, $a_{(n)}^i(h) \equiv 0$. To do this, one takes into account the definition (5.3) of the $\mathscr{A}_p$ and again uses the Virasoro algebra to show successively $L_{n-1}|\psi\rangle = L_{n-2}|\psi\rangle = \cdots = 0$. Ending up with $|\psi\rangle \in \mathbf{V}^{(0)}$ for all $|\chi\rangle \in \mathbf{V}^{(0)}$, one concludes $a_{(n)}^i \equiv 0$.

Thus we have proved the existence of gauge transformations defined by $\mathscr{S}_{-n}$ at each level n if $D = 26$. The solutions $a_{(n)}^i(h)$ arising from Eqs. (5.4) are unique (up to common factors). We have *not* proved that these are the only solutions satisfying (5.6). Nevertheless, a lot of playing around with (5.4) and expressions like (5.1) and (5.2) has convinced ourselves that the systems (5.4) and (5.6)—as generic equations for the $a_{(n)}^i(h)$—are equivalent. It also seems to be true that (5.6) together with (5.19) leads to $a_{(n)}^i \equiv 0$, which would again prove the equivalence. However, we have no rigorous proof but only the definite statement that everything works up to $n = 7$.

In the following we assume that our conjecture holds. In other words this means that the rank of the matrix $\mathscr{K}^{(n)i}{}_j(h)$ is equal to $\ell_n - 1$ except possibly for a finite number of values of h. These critical values may be found by computing the characteristic polynomial

$$\det(\lambda - \mathscr{K}^{(n)}(h)) = \lambda p^{(n)}(h) + \cdots + \lambda^{\ell_n}\,, \quad (5.20)$$

for $D = 26$; $\lambda = 0$ is always an eigenvalue. The zeros of the polynomial $p^{(n)}(h)$ are those values where a second eigenvector with eigenvalue zero *might* arise (in the next section we will conjecture that this is not the case).

Let us at the end of this section mention an interesting property of the $a_{(n)}^i$. Evaluating Eqs. (5.13) explicitly, one gets

$$A_{(n)i}^{j(n-p)}\mathscr{M}_{jk}^{(n)}(L_0)a_{(n)}^k(L_0)|\chi\rangle = 0\,, \quad (5.21)$$

for $2 \leqslant p \leqslant n$, where

$$\mathscr{L}_i^{(n-p)}\mathscr{A}_p = \mathscr{L}_j^{(n)}A_{(n)i}^{j(n-p)}\,. \quad (5.22)$$

The $\mathscr{A}_p$ contain only L_m with positive m, hence A does not depend on L_0. Since $\mathscr{M}_{jk}^{(n)}$ is generically nonsingular, (5.21) may be read as an equation for $a_j^{(n)} = \mathscr{M}_{jk}^{(n)}a_{(n)}^k$, the "covariant" components of $a^{(n)}$. The arguments given above apply to this interpretation as well. One finds a solution $b_j^{(n)}$ that is independent of L_0. Thus the generic solution $a_{(n)}^j(L_0)$ satisfies

$$a_j^{(n)}(h) \equiv \mathscr{M}_{jk}^{(n)}(h) a_{(n)}^k(h) = q^{(n)}(h) b_j^{(n)}, \qquad (5.23)$$

for some polynomial $q^{(n)}(h)$. The real numbers $b_j^{(n)}$ are a solution for the equation adjoint to (4.12),

$$b_j^{(n)} \mathscr{K}^{(n)j}{}_i(h) = 0. \qquad (5.24)$$

In order to compute $a_{(n)}^i$ explicitly, one may solve (5.24) for *some* noncritical value of h to find $b_i^{(n)}$. Then

$$a_{(n)}^i(h) = q^{(n)}(h) \mathscr{M}_{(n)}^{ij}(h) b_j^{(n)}, \qquad (5.25)$$

where $q^{(n)}(h)$ is a polynomial with minimal degree that eats up all singularities of $\mathscr{M}_{(n)}^{ij}(h) b_j^{(n)}$ [cf. Eq. (A31)].

VI. EXISTENCE OF THE VIRASORO GAUGE

Once having an infinite sequence of gauge transformations $\mathscr{S}_{-n}$ and the conjecture that they are essentially unique, the question arises whether these may actually be used to obtain the Virasoro conditions as a special gauge fixing. Neglecting the possibility of nongeneric states, the answer is trivial: In the generic sense, the decomposition (4.23) *is* the solution of the field equation (4.3). However, we must be a little bit more careful in order not to oversee the nongeneric states that lie in the intersections $\mathbf{V}^{(n)} \cap \mathbf{V}^{(m)}$.

Let $|\psi\rangle$ be any state that contains only Verma level contributions up to some positive n, i.e., $|\psi\rangle \in \mathbf{W}^{(n)}$. Then $|\psi\rangle$ is of the form (A32). The existence of the gauge conditions is ensured if $\mathscr{K}|\psi\rangle = 0$ implies

$$|\psi\rangle = |\xi\rangle + \sum_{p=1}^{n} \mathscr{S}_{-p}|\chi_p\rangle, \qquad (6.1)$$

for $|\xi\rangle$ and $|\chi_p\rangle$ in $\mathbf{V}^{(0)}$. Note that from

$$\mathscr{K}|\psi\rangle \equiv 2(L_0 - 1)|\chi_0\rangle + \sum_{p=1}^{n} \mathscr{L}_{-j}^{(p)} \mathscr{K}^{(p)j}{}_i(L_0)|\chi_p^i\rangle = 0, \qquad (6.2)$$

one cannot straightforwardly conclude that all the expressions at the different levels vanish separately. Multiplying by $\mathscr{L}_i^{(n)}$, we find

$$\mathscr{L}_i^{(n)} \mathscr{K}|\psi\rangle \equiv \mathscr{K}_{ij}^{(n)}(L_0)|\chi_n^j\rangle = 0. \qquad (6.3)$$

If there exists some $|\chi\rangle \in \mathbf{V}^{(0)}$ such that

$$\mathscr{M}_{ij}^{(n)}(L_0)(|\chi_n^j\rangle - a_{(n)}^j(L_0)|\chi\rangle) = 0, \qquad (6.4)$$

one finds, taking into account the definition (A27) of $\mathscr{M}^{(n)}$ that

$$\mathscr{L}_{-i}^{(n)}|\chi_n^i\rangle - \mathscr{S}_{-n}|\chi\rangle \in \mathbf{W}^{(n-1)}. \qquad (6.5)$$

Identifying $|\chi\rangle$ with $|\chi_n\rangle$ in (6.1), one has already confirmed the highest term. The remaining state

$$|\psi'\rangle = |\psi\rangle - \mathscr{S}_{-n}|\chi\rangle, \qquad (6.6)$$

which is an element of $\mathbf{W}^{(n-1)}$, is treated analogously. By induction, one arrives at (6.1). Thus, in order to complete the argument, one has to deduce (6.4) from (6.3). This is the second point where we have not been able to give a rigorous proof.

The conjecture we are forced to make simply states that the rank of the matrices $\mathscr{K}^{(n)i}{}_j(h)$ is $\ell_n - 1$ for *all* values h—including the zeros of the polynomial $p^{(n)}(h)$ in (5.20). Thus, at the critical values, $\mathscr{K}^{(n)i}{}_j(h)$ becomes nondiagonalizable since the eigenvalue zero appears twice in the characteristic polynomial. This situation is illustrated by (4.17) for the case $n = 2$. The critical values of h seem to be just those where two generic eigenvectors coincide.

Taking this statement for granted, one concludes by means of elementary linear algebra that there exist polynomials $B^{(n)i}{}_j(h)$ and real numbers $f_{(n)}^j$ such that

$$f_{(n)}^j b_j^{(n)} = 1 \qquad (6.7)$$

and

$$\mathscr{K}^{(n)i}{}_j(h) B^{(n)j}{}_k(h) = \delta_k^i - f_{(n)}^i b_k^{(n)}. \qquad (6.8)$$

Thus, from

$$\mathscr{K}^{(n)i}{}_j(L_0)|\xi_i\rangle = 0, \qquad (6.9)$$

for $|\xi_i\rangle \in \mathbf{V}^{(0)}$, it follows upon multiplication by $B^{(n)j}{}_i(L_0)$ that

$$|\xi_i\rangle = b_i^{(n)}|\tilde{\chi}\rangle \qquad (6.10)$$

with

$$|\tilde{\chi}\rangle = f_{(n)}^i|\xi_i\rangle. \qquad (6.11)$$

Now consider an equation of the type (6.3),

$$\mathscr{K}_{ij}^{(n)}(L_0)|\chi^j\rangle \equiv \mathscr{K}^{(n)k}{}_i(L_0)\mathscr{M}_{kj}^{(n)}(L_0)|\chi^j\rangle = 0, \qquad (6.12)$$

where one has made use of (5.12). Denoting

$$|\xi_k\rangle = \mathscr{M}_{kj}^{(n)}(L_0)|\chi^j\rangle, \qquad (6.13)$$

one arrives at (6.9), which implies via (6.10)

$$\mathscr{M}_{kj}^{(n)}(L_0)|\chi^j\rangle = b_k^{(n)}|\tilde{\chi}\rangle, \qquad (6.14)$$

for some $|\tilde{\chi}\rangle \in \mathbf{V}^{(0)}$. Since the polynomial $q^{(n)}(h)$ appearing in (5.23) is nonvanishing, one may always find a $|\chi\rangle \in \mathbf{V}^{(0)}$ such that

$$|\tilde{\chi}\rangle = q^{(n)}(L_0)|\chi\rangle. \qquad (6.15)$$

Hence, using (5.23),

$$\mathscr{M}_{kj}^{(n)}(L_0)(|\chi^j\rangle - a_{(n)}^j|\chi\rangle) = 0, \qquad (6.16)$$

which is just the statement (6.4).

We should summarize now in a simpler notation what we have proved under the assumption that the rank of $\mathscr{K}^{(n)i}{}_j(h)$ equals $\ell_n - 1$ for all h. Let $n \geqslant 1$ and $|v\rangle \in \mathbf{V}^{(n)}$. Then $\mathscr{K}|v\rangle \in \mathbf{W}^{(n-1)}$ if and only if there exists a $|\chi\rangle \in \mathbf{V}^{(0)}$ such that $|v\rangle - \mathscr{S}_{-n}|\chi\rangle \in \mathbf{W}^{(n-1)}$. In this formulation, the state $|v\rangle$ is just the highest term in the decomposition (A32), and the statement that $\mathscr{K}|v\rangle$ lies in $\mathbf{W}^{(n-1)}$ is identical to Eq. (6.3).

Applying this result to each mass level of a given state satisfying the field equation (4.3), one arrives at the infinite decomposition (4.23).

There is another algorithm that yields the Virasoro gauge conditions. It has been checked for $n \leqslant 4$, and its general validity would prove the conjecture made above. Take any $|\psi\rangle \in \mathbf{W}^{(n)}$, subject to the field equation (4.3). Trying to find a gauge in which $L_n|\psi\rangle = 0$, one obtains the equation

$$\delta(L_n|\psi\rangle) \equiv L_n \mathscr{S}_{-n}|\chi_n\rangle \equiv f_n(L_0)|\chi_n\rangle = -L_n|\psi\rangle, \qquad (6.17)$$

where $f_n(L_0)$ is a nonvanishing polynomial. Since $L_n|\psi\rangle$ lies in $\mathbf{V}^{(0)}$, this equation can be solved for $|\chi_n\rangle \in \mathbf{V}^{(0)}$. Now we conjecture that, from

$$\mathcal{K}|\psi\rangle = L_n|\psi\rangle = L_{n-1}|\psi\rangle = \cdots = L_{p+1}|\psi\rangle = 0\,, \tag{6.18}$$

it follows by pure application of the Virasoro algebra (A22) that $L_p|\psi\rangle$ is annihilated by all L_m ($m\geqslant 1$), hence an element of $\mathbf{V}^{(0)}$. Then the equation

$$\delta(L_p|\psi\rangle) \equiv L_p \mathcal{S}_{-p}|\chi_p\rangle \equiv f_p(L_0)|\chi_p\rangle = -L_p|\psi\rangle \tag{6.19}$$

is solvable for some $|\chi_p\rangle \in \mathbf{V}^{(0)}$. Iterating this process, one ends up with the gauge condition (A16).

VII. GAUGE TRANSFORMATIONS WITH UNCONSTRAINED PARAMETER

Instead of performing an infinite sequence of gauge transformations

$$\delta|\psi\rangle = \sum_{p=1}^{\infty} \mathcal{S}_{-p}|\chi_p\rangle \tag{7.1}$$

with parameters $|\chi_p\rangle \in \mathbf{V}^{(0)}$, one might try to formulate (7.1) in terms of a simple operator

$$\delta|\psi\rangle = \mathcal{S}|\Lambda\rangle \tag{7.2}$$

with an unconstrained parameter $|\Lambda\rangle$. In the following, we will construct an operator $\mathcal{S}$ satisfying

$$\mathcal{K}\mathcal{S} = 0 \tag{7.3}$$

and acting upon the different Verma levels according to

$$\mathcal{S}: \mathbf{V}^{(n)} \to \mathbf{V}^{(n+1)}\,, \tag{7.4}$$

for $n\geqslant 0$. Repeated application of $\mathcal{S}$ should give (with $|\chi\rangle \in \mathbf{V}^{(0)}$)

$$\mathcal{S}^p|\chi\rangle = \mathcal{S}_{-p}|\xi_p\rangle\,, \tag{7.5}$$

for some $|\xi_p\rangle \in \mathbf{V}^{(0)}$. Thus $|\Lambda\rangle$ in (7.2) may be chosen as

$$|\Lambda\rangle = \sum_{p=0}^{\infty} \mathcal{S}^p|\eta_p\rangle\,, \tag{7.6}$$

for appropriate elements $|\eta_p\rangle$ of $\mathbf{V}^{(0)}$ in order to achieve (7.1). Clearly, the gauge transformations in the form (7.2) are highly redundant because $|\Lambda\rangle$ is completely arbitrary.

Making the ansatz (using $\mathcal{L}^{(0)} = \mathbf{1}$ by convention)

$$\mathcal{S} = \sum_{p=0}^{\infty} \mathcal{L}^{(p+1)}_{-i} s_{(p+1,p)}{}^{ij}(L_0) \mathcal{L}^{(p)}_j \tag{7.7}$$

and requiring that $\mathcal{K}\mathcal{S}$ annihilates a general element of $\mathbf{V}^{(n)}$, one finds the recursion formula

$$\begin{aligned} s_{(n+1,n)}{}^{\ell j}(h)\mathcal{M}^{(n)}_{jk}(h) &= a^{\ell}_{(n+1)}(h)c^{(n)}_k(h) \\ &\quad - \sum_{p=0}^{n-1} s_{(p+1,p)}{}^{ij}(h-p+n)C^{\ell}{}_{(n+1)}{}^{(p+1,p,n)}{}_{ijk}(h)\,, \end{aligned} \tag{7.8}$$

where $c^{(n)}_k(h)$ is arbitrary and the C's arise from the identity

$$\begin{aligned} &\mathcal{L}^{(p+1)}_{-i}\mathcal{L}^{(p)}_j\mathcal{L}^{(n)}_{-k}|\chi\rangle \\ &= \mathcal{L}^{(n+1)}_{-\ell}C^{\ell}{}_{(n+1)}{}^{(p+1,p,n)}{}_{ijk}(L_0)|\chi\rangle\,, \end{aligned} \tag{7.9}$$

for $|\chi\rangle \in \mathbf{V}^{(0)}$. Solving (7.8) with respect to $s_{(n+1,n)}{}^{\ell j}$, one has to multiply by the inverse $\mathcal{M}^{ki}_{(n)}(h)$ and finds operators $\mathcal{S}$ that satisfy (7.3) but are in general nonlocal.

However, there are some hints that the arbitrary functions $s_{(1,0)}(h)$ and $c^{(n)}_i(h)$ may be adjusted in such a way that all $s_{(n+1,n)}{}^{ij}(h)$ become polynomial. In order to get a local action of $\mathcal{S}$ only upon $\mathbf{W}^{(n)}$, one just has to solve (7.8) for any choice of the arbitrary functions up to the nth order and multiply the result by the overall denominator. In order to get more insight into the structure of $\mathcal{S}$, we introduce the nonlocal operators[13]

$$\Pi^{(0)} = 0, \quad \Pi^{(n)} = \mathbf{1} - \mathcal{L}^{(n)}_{-i}\mathcal{M}^{ij}_{(n)}(L_0)\mathcal{L}^{(n)}_j\,, \tag{7.10}$$

for $n\geqslant 1$, and

$$\mathcal{P}^{(n)} = \Pi^{(n)}\Pi^{(n+1)}\Pi^{(n+2)}\cdots\,. \tag{7.11}$$

Denoting further

$$c^{(n)}_i(h) = \mathcal{M}^{(n)}_{ij}(h)c^j_{(n)}(h)\,, \tag{7.12}$$

$$\mathcal{C}_n \equiv c^i_{(n)}(L_0)\mathcal{L}^{(n)}_i\,, \tag{7.13}$$

a closed expression for $\mathcal{S}$—representing the general solution of (7.8)—is given by

$$\mathcal{S} = \sum_{n=0}^{\infty} \mathcal{S}_{-n-1}\mathcal{C}_n(1-\Pi^{(n)})\mathcal{P}^{(n+1)}\,, \tag{7.14}$$

where $c_{(0)}$ is to be identified with $s_{(1,0)}$. This equation is easily confirmed by using (7.10) to compute the action of $\mathcal{S}$ upon $\mathbf{V}^{(n)}$:

$$\mathcal{S}\mathcal{L}^{(n)}_{-i}|\chi\rangle = \mathcal{S}_{-n-1}\mathcal{C}_n\mathcal{L}^{(n)}_{-i}|\chi\rangle\,. \tag{7.15}$$

Iterative application of $\mathcal{S}$ upon an element of $\mathbf{V}^{(0)}$ gives

$$\mathcal{S}^p|\chi\rangle = \mathcal{S}_{-p}g_p(L_0)|\chi\rangle\,, \tag{7.16}$$

where

$$g_p(h) = f_{p-1}(h)f_{p-2}(h)\cdots f_2(h)f_1(h)c_{(0)}(h) \tag{7.17}$$

and

$$f_n(h) = c^i_{(n)}(h)\mathcal{M}^{(n)}_{ij}(h)a^j_{(n)}(h) \equiv c^{(n)}_j(h)a^j_{(n)}(h)\,. \tag{7.18}$$

Thus, if $c_{(0)}$ and all f_n are nontrivial polynomials, we recover (7.5) with

$$|\xi_p\rangle = g_p(L_0)|\chi\rangle\,.$$

Choosing $s_{(1,0)}(h) = 1$ and setting all subsequent $c^{(n)}_i(h)$ equal to zero, the result for $\mathcal{S}$ is just the operator $L_{-1}P$, where P is the projection onto $\mathbf{V}^{(0)}$. This can be seen by comparing (7.14) with the equation $P = \mathcal{P}^{(1)}$, which was given in Ref. 13. Clearly, this solution is neither local nor realizes (7.5) nontrivially. This was to be expected, because setting all $c^{(n)}_i$ equal to zero, no detailed information about $\mathcal{K}$ enters in $\mathcal{S}$.

In order to obtain a local solution for $\mathcal{S}$ that realizes all gauge transformations (7.5) occurring in the present formulation of free bosonic string field theory, we have to choose the $c^{(n)}_j$ in such a way that all singularities arising from the matrices $\mathcal{M}^{ij}_{(n)}(h)$ cancel, thereby keeping the $f_n(h)$ nontrivial. From (7.8) or (7.15) one concludes that the $c^{(n)}_i(h)$ have to be polynomials as well.

At the first three levels it is possible to adjust the arbitrary functions such that no L_0 appears explicitly. The result is unique up to a constant factor:

$$\mathcal{S} = L_{-1} - \tfrac{1}{3}L_{-2}L_1 - \tfrac{1}{13}L_{-3}(L_2 - \tfrac{2}{3}L_1^2) + \cdots\,, \tag{7.19}$$

where the dots denote higher-order terms. Inserting this expression into the equations for the next level, one inevitably encounters polynomials in L_0. Apparently, *any* local solution for $\mathscr{S}$ up to the level $\mathscr{L}_{-i}^{(p+1)}\mathscr{L}_j^{(p)}$ admits a family of local solutions at the next higher level. This has been checked up to $\mathscr{L}_{-i}^{(4)}\mathscr{L}_j^{(3)}$. The calculation of some solution for $\mathscr{S}$ to arbitrarily high levels is then straightforward but very tedious.

The general solution contains infinitely many arbitrary polynomials in L_0. This can also be seen by noting that, along with $\mathscr{S}$, also all operators $\mathscr{S}' = \mathscr{S}\mathscr{T}$ do the job if the nontrivial realization of (7.5) remains valid—and this is the case for various choices of $\mathscr{T}$.

If one prefers, however, to have an even more explicit expression for the unification of the gauge transformations into one closed formula, one may proceed as follows. Let

$$|\Lambda\rangle = \sum_{n=0}^{\infty} |\Lambda\rangle_n \tag{7.20}$$

be the (unique) decomposition of an arbitrary state with respect to the mass level operator [cf. Eqs. (A11) and (A12)]. Then define an operator $\mathscr{R}$ on Fock space by its action upon the nth mass level,

$$\mathscr{R}|\Lambda\rangle_n = r_n(L_0)P|\Lambda\rangle_n\,, \tag{7.21}$$

where the r_n are polynomials that cancel the singularities contained in the relevant part of P at the nth mass level. They are essentially the overall denominator of the products $\Pi^{(1)}\Pi^{(2)}\cdots\Pi^{(n)}$. Although $\mathscr{R}$ may not be expressed as a sum over products of Virasoro operators, it acts locally upon the space-time fields in $|\Lambda\rangle$, leaves the mass level invariant, and puts any state to Verma level zero,

$$\mathscr{R}: \mathbf{F}\to\mathbf{V}^{(0)}\,. \tag{7.22}$$

Of course, $\mathscr{R}$ is no longer a projector.

As a consequence of (7.22), any of the operators $\mathscr{S}_{-n}\mathscr{R}$ generates a gauge transformation of the type (4.7). Defining the operator

$$\hat{\mathscr{S}} = \sum_{n=1}^{\infty} \mathscr{S}_{-n}\mathscr{R}L_n\,, \tag{7.23}$$

we again recover *all* transformations (4.7) by setting

$$\delta|\psi\rangle = \hat{\mathscr{S}}|\Lambda\rangle \tag{7.24}$$

for unconstrained $|\Lambda\rangle$. Note that the L_n at the right-hand side of (7.23) may be replaced by any local operator $u_i^{(n)}(L_0)\mathscr{L}_i^{(n)}$ [or likewise by $u_i^{(n-1)}(L_0)\mathscr{L}_i^{(n-1)}$ in order to make the level zero part of $|\Lambda\rangle$ contribute to the variation of $|\psi\rangle$]. This construction is essentially a straightforward generalization of (3.14) to all space-time fields contained in a general Fock state. Since it makes explicit reference to the mass level decomposition [whereas the expression (4.22) for the kinetic operator does not] it seems artificial as compared to (7.7), but it is sufficient to merely reproduce the correct gauge transformation law for each space-time field.

One could think of using $2\mathscr{R}(L_0-1)$ as an alternative kinetic operator, which contains only a finite order of space-time derivatives at a definite mass level. As a result of the general property

$$\mathscr{R}L_{-n} = 0 \tag{7.25}$$

($n\geqslant 1$), one could always achieve the Virasoro conditions (A16) as a special choice of gauge. However, the remaining field equation would contain the polynomials $r_n(L_0)$ at the nth mass level in addition to the factor $L_0 - 1$. The mathematical reason for $\mathscr{R}$ being applicable in the transformation law (7.24) but not serving as kinetic operator is the fact that the restricted function

$$\mathscr{R}: \mathbf{V}^{(0)}\to\mathbf{V}^{(0)} \tag{7.26}$$

is surjective (as long as boundary conditions are omitted) but not injective.

Having established the forms (7.2) or (7.24) of the gauge symmetry, we are left with a single operator $\mathscr{S}$ (or $\hat{\mathscr{S}}$) instead of L_{-1} and L_{-2} (and hence all L_{-n}) as in the usual gauge-covariant approach. Choosing $|\psi\rangle$ itself as the gauge parameter, we find that the action functional admits "global" symmetries

$$\delta|\psi\rangle = \mathscr{S}|\psi\rangle \tag{7.27}$$

(or $\hat{\mathscr{S}}$ instead of $\mathscr{S}$), replacing the freedom of reparametrizations on the world sheet

$$\delta|\psi\rangle = L_{-n}|\psi\rangle \tag{7.28}$$

in the gauge-covariant formulations. The physical interpretation of these symmetries is not yet clarified.

We finally remark that the redundancy in the gauge transformation law (7.2) [resp. (7.24)] gives rise to "ghosts for ghosts" in the BRST quantization formalism, a feature that is well-known for string theories.

VIII. THE RESIDUAL GAUGE FREEDOM

Once having specified the Virasoro gauge condition for a solution to (4.3), the remaining gauge freedom is contained in the set of all states

$$|\psi\rangle \equiv \sum_{p=1}^{\infty} \mathscr{S}_{-p}|\chi_p\rangle \in \mathbf{V}^{(0)}\,, \tag{8.1}$$

where $|\chi_p\rangle \in \mathbf{V}^{(0)}$. Considering only a certain mass level n, the series (8.1) stops at $p = n$. Applying $\mathscr{L}_i^{(n)}$, one finds

$$\mathscr{M}_{ij}^{(n)}(L_0)a_{(n)}^j(L_0)|\chi_n\rangle = 0\,, \tag{8.2}$$

thus—using (5.23)—

$$q^{(n)}(L_0)|\chi_n\rangle = 0\,, \tag{8.3}$$

which implies

$$q^{(n)}(L_0 - n)\mathscr{S}_{-n}|\chi_n\rangle = 0\,. \tag{8.4}$$

Equation (8.3) is necessary but not sufficient for $|\chi_n\rangle$ to be contained in a state of the type (8.1). An example of such a state is given by Eqs. (4.24) and (4.25). The polynomials $q^{(n)}(h)$ are contained as factors in the Kac determinants [this follows from (5.25)]. Thus only part of the zeros of $\det\mathscr{M}^{(n)}$ is actually responsible for the existence of a residual gauge symmetry.

This occurrence of a nonphysical leftover in the covariant approach to string theories is not only known in principle but has also been worked out in detail: A general physical state is the sum of a DDF state—carrying the true physical degrees of freedom—and a physical spurious state.[2]

IX. CLOSED STRINGS

In the theory of oriented closed strings, the number of degrees of freedom is doubled as compared to the open strings. As a result of the existence to two independent sets of oscillator variables α_n^μ and $\bar{\alpha}_n^\mu$ (subject to the single relation $\frac{1}{2}p^\mu = \alpha_0^\mu = \bar{\alpha}_0^\mu$), the constraints are realized by means of two sets of Virasoro operators L_n and $\bar{L}_n$ in the closed-string Fock space. Different type operators commute with each other. The conditions defining the physical states are

$$(L_0 + \bar{L}_0 - 2)|\psi\rangle = 0\,, \tag{9.1}$$

$$L_n|\psi\rangle = \bar{L}_n|\psi\rangle = 0\,, \tag{9.2}$$

for all $n \geqslant 1$, supplemented by

$$(L_0 - \bar{L}_0)|\psi\rangle = 0\,, \tag{9.3}$$

which can also be written as

$$(N - \bar{N})|\psi\rangle = 0\,, \tag{9.4}$$

where N and $\bar{N}$ are the respective mass level operators [cf. Eq. (A10)]. Our task is again to find an action functional that produces the above conditions as field equations in a special gauge. In order to achieve (9.4), we define a projection operator Q onto the set of states satisfying (9.4). Expanding an arbitrary state with respect to the two types of mass level operators,

$$|\psi\rangle = \sum_{n,m=0}^{\infty} |\psi\rangle_{nm}\,, \tag{9.5}$$

where

$$N|\psi\rangle_{nm} = n|\psi\rangle_{nm}, \quad \bar{N}|\psi\rangle_{nm} = m|\psi\rangle_{nm}\,, \tag{9.6}$$

the action of Q is given by

$$Q|\psi\rangle = \sum_{n=0}^{\infty} |\psi\rangle_{nn}\,. \tag{9.7}$$

The operator Q acts locally on the field components (it just puts some of them to zero and leaves the others unchanged) and is Hermitian:

$$Q^\dagger = Q\,. \tag{9.8}$$

Although we have not gone so far into the details of the closed string case, a candidate for an appropriate action is

$$S = -(\psi|\mathscr{H}Q|\psi)\,, \tag{9.9}$$

where

$$\mathscr{H} = L_0 + \bar{L}_0 - 2 - \sum_{n=1}^{\infty} \frac{1}{n} L_{-n}L_n - \sum_{n=1}^{\infty} \frac{1}{n} \bar{L}_{-n}\bar{L}_n\,, \tag{9.10}$$

$$\mathscr{H}^\dagger = \mathscr{H}\,, \tag{9.11}$$

and

$$[\mathscr{H}, Q] = 0\,. \tag{9.12}$$

It admits the gauge symmetry

$$\delta|\psi\rangle = (1 - Q)|\Lambda\rangle \equiv \sum_{n,m=0,\, n\neq m}^{\infty} |\Lambda\rangle_{nm}\,, \tag{9.13}$$

which may be used to achieve (9.4) as the gauge-fixing condition. One may, alternatively, omit the Q in (9.9) and insert (9.4) by hand.

Moreover, setting $D = 26$, the action (9.9) is invariant under the two type of symmetries

$$\delta|\psi\rangle = \mathscr{S}_{-m}|\chi\rangle \tag{9.14}$$

and

$$\delta|\psi\rangle = \bar{\mathscr{S}}_{-m}|\chi\rangle\,, \tag{9.15}$$

where

$$L_p|\chi\rangle = \bar{L}_p|\chi\rangle = 0\,, \tag{9.16}$$

for all $p \geqslant 1$. In order not to destroy (9.4), one must choose $|\chi\rangle$'s of the correct mass levels, e.g., for (9.14)

$$\delta|\psi\rangle_{nn} = \mathscr{S}_{-m}|\chi\rangle_{n-m,n}\,. \tag{9.17}$$

A typical state $|\chi\rangle_{0,1}$, subject to (9.16), is of the form

$$|\chi\rangle_{0,1} = -i\bar{\alpha}_{-1}^\mu \chi_\mu(x)|0\rangle\,, \tag{9.18}$$

with

$$\partial_\mu \chi^\mu = 0\,. \tag{9.19}$$

Applied to

$$|\psi\rangle_{1,1} = \alpha_{-1}^\mu \bar{\alpha}_{-1}^\nu k_{\mu\nu}(x)|0\rangle\,, \tag{9.20}$$

it gives rise to the gauge transformation law $\delta|\psi\rangle_{1,1} = L_{-1}|\chi\rangle_{0,1}$ or, written in components,

$$\delta k_{\mu\nu} = \tfrac{1}{2}\partial_\mu \chi_\nu\,. \tag{9.21}$$

Of course, the role of operators with and without bars may be interchanged in (9.17) and (9.18), leading to the additional gauge symmetry at the mass level $|\psi\rangle_{1,1}$:

$$\delta k_{\mu\nu} = \tfrac{1}{2}\partial_\nu \Lambda_\mu, \tag{9.22}$$

with

$$\partial_\mu \Lambda^\mu = 0\,. \tag{9.23}$$

Decomposing the space-time field $k_{\mu\nu}(x)$ contained in (9.20) into its symmetric traceless, pure trace, and antisymmetric parts, the transformation laws (9.21) and (9.22) split into the symmetries of the well-known graviton, dilaton, and antisymmetric tensor field at the first excited mass level of the closed string.

Let us now examine whether the gauge symmetries (9.14) and (9.15) are enough to admit the Virasoro constraints (9.2) as a special gauge condition and the mass-shell condition (9.1) as the remainder of the field equation

$$\mathscr{H}|\psi\rangle = 0 \tag{9.24}$$

in that special gauge. In the following we assume that the symmetry (9.13) has already been exploited and (9.4) is valid from the outset.

At the mass level $|\psi\rangle_{0,0}$ (the tachyon field) there is nothing to prove, because the physical state condition follows immediately. Next consider the states $|\psi\rangle_{1,1}$, as given by (9.20). Trying to achieve (9.2) for $n = 1$, one sets

$$\delta(L_1|\psi\rangle_{1,1}) \equiv L_1\mathscr{S}_{-1}|\chi\rangle_{0,1} \equiv 2L_0|\chi\rangle_{0,1} = -L_1|\psi\rangle_{1,1}\,,$$
$$\delta(\bar{L}_1|\psi\rangle_{1,1}) \equiv \bar{L}_1\bar{\mathscr{S}}_{-1}|\Lambda\rangle_{1,0} \equiv 2\bar{L}_0|\Lambda\rangle_{1,0} = -\bar{L}_1|\psi\rangle_{1,1}\,. \tag{9.25}$$

These equations admit a solution for $|\chi\rangle$ and $|\Lambda\rangle$ if and only if the integrability conditions

$$L_1\bar{L}_1|\psi\rangle_{1,1} = 0\,, \tag{9.26}$$

or in terms of the component field,

$$\partial^{\mu\nu}k_{\mu\nu} = 0\,, \tag{9.27}$$

are valid. Applying L_1 or $\bar{L}_1$ upon (9.24), one gets

$$L_{-1}L_1\bar{L}_1|\psi\rangle_{1,1} = \bar{L}_{-1}L_1\bar{L}_1|\psi\rangle_{1,1} = 0\,, \tag{9.28}$$

or equivalently

$$\partial_{\rho\mu\nu}k^{\mu\nu} = 0\,, \tag{9.29}$$

stating that $\partial^{\mu\nu}k_{\mu\nu}$ is not strictly zero but only a constant, and this is the best one can do locally. In order to deduce (9.27) from (9.29), one has to impose an appropriate boundary condition, e.g., that $k_{\mu\nu}$ vanishes at spatial infinity (no matter how fast). This is in contrast to the open string case, where the existence of the gauge parameters necessary to achieve the Virasoro constraints was guaranteed by means of pure differential manipulations alone. The reason for this new feature is that the two types of Virasoro operators only combine into equations like (9.28), which are absent in the open string case. However, once having imposed the necessary boundary condition, the correct gauge fixing is possible.

At the second mass level $|\psi\rangle_{2,2}$, we meet an analogous situation. The integrability conditions necessary to achieve

$$L_2|\psi\rangle_{2,2} = \bar{L}_2|\psi\rangle_{2,2} = 0 \tag{9.30}$$

turn out to be

$$\bar{L}_1L_2|\psi\rangle_{2,2} = \bar{L}_2L_1|\psi\rangle_{2,2} = \bar{L}_2L_2|\psi\rangle_{2,2} = 0\,. \tag{9.31}$$

Expanding

$$\begin{aligned} \bar{L}_1L_2|\psi\rangle_{2,2} &= \bar{\alpha}^{\mu}_{-1}\alpha_{\mu}|0\rangle\,, \\ \bar{L}_2L_1|\psi\rangle_{2,2} &= \alpha^{\mu}_{-1}\bar{\alpha}_{\mu}|0\rangle\,, \\ \bar{L}_2L_2|\psi\rangle_{2,2} &= b\,|0\rangle\,, \end{aligned} \tag{9.32}$$

one concludes from the field equation (9.24) by pure differential manipulations that the combinations $\alpha_\mu + (i/6)\partial_\mu b$ and $\bar{\alpha}_\mu + (i/6)\partial_\mu b$ are constant. Again upon appropriate boundary conditions, these constants have to be zero. Assuming $\alpha_\mu = \bar{\alpha}_\mu = -(i/6)\partial_\mu b$, one concludes by further exploitation of (9.24) that $\alpha_\mu = \bar{\alpha}_\mu = b = 0$. Hence the gauge condition given by (9.31) is possible. In order to obtain the remaining Virasoro constraints

$$L_1|\psi\rangle_{2,2} = \bar{L}_1|\psi\rangle_{2,2} = 0\,, \tag{9.33}$$

we have to deduce the integrability conditions

$$L_1^2|\psi\rangle_{2,2} = \bar{L}_1^2|\psi\rangle_{2,2} = L_1\bar{L}_1|\psi\rangle_{2,2} = 0\,. \tag{9.34}$$

These follow from (9.24), (9.30), and (9.31) without imposing any further boundary conditions.

Although we have not checked this mechanism for the higher mass levels, the computations done so far provide a certain evidence that the action (9.9) applies to all space-time fields contained in $|\psi\rangle$. The question of whether there is a better one—which works without imposing boundary conditions—remains open.

X. CONCLUSION

The conjectures made so far in the open-string case are summarized by the statement that the matrices $\mathscr{K}^{(n)i}{}_j(h)$ have a one-dimensional kernel for all h. Under the assumption that it holds we have shown that a free theory of open bosonic string fields may be formulated without the use of supplementary fields. The action is local in the string fields and contains space-time derivatives only up to the second order. Let us finally raise the question of whether there exist other kinetic operators constructed by the Virasoro operators alone. The most general candidate is

$$\tilde{\mathscr{K}} = 2(L_0 - 1) + \sum_{n=1}^{\infty} \mathscr{L}^{(n)}_{-i}\tilde{k}^{ij}_{(n)}(L_0)\mathscr{L}^{(n)}_j\,, \tag{10.1}$$

with

$$\tilde{k}^{ij}_{(n)} = \tilde{k}^{ji}_{(n)} \tag{10.2}$$

to ensure the Hermiticity of $\tilde{\mathscr{K}}$. If $\tilde{\mathscr{K}}$ is to contain at most second-order space-time derivatives, the $\tilde{k}^{ij}_{(n)}$ have to be constants, and only the combinations $L_{-n}L_n$ have nonzero coefficients. Thus we are led to the ansatz (4.1) made already in Sec. IV. In this sense, the solution of the problem as given in (4.22) together with $D = 26$ seems to be the *unique* one.

However, in order to understand the significance of the conjecture in this paper, we may omit the restriction to second-order space-time derivatives and ask for a general criterion for (10.1) to be applicable for string field theory. Following essentially the same arguments as given in Sec. VI, it turns out that $\mathscr{K}$ admits the Virasoro gauge conditions if and only if the dimension of the kernels

$$\dim\ker \tilde{\mathscr{K}}^{(n)i}{}_j(h) = r_n \tag{10.3}$$

does not depend on h. The matrices $\tilde{\mathscr{K}}^{(n)i}{}_j$ are defined analogously to (4.11). The operator $\mathscr{K}$ suggested in this paper has $r_n = 1$ for all $n \geqslant 1$. An operator $\tilde{\mathscr{K}}$ with $r_n = 0$ for all $n \geqslant 1$ would not admit any gauge symmetry at all. However, $r_n = 0$ contradicts (10.1). If, on the other hand, the matrices $\tilde{\mathscr{K}}^{(n)i}{}_j(h)$ were nonsingular only for generic h—which is probably the case for (4.1) with $a_n \neq 1/n$ at $n \geqslant 2$—one would find zero-norm states corresponding to critical values of h, satisfying the field equations but being neither physical nor gauge. The equality of the dimensions in (10.3) for *all h* is thus necessary to absorb also these exceptional states into the decomposition (4.23).

Let us finally note that the actions defined by the kinetic operators $\mathscr{K}$ and $\mathscr{H}$ agree with those given by Neveu, Nicolai, and West[15] when all supplementary fields are set equal to zero. From this point of view, the symmetries provided by the $\mathscr{S}_{-n}$ are possibly just those of Ref. 15, which leave all supplementary fields invariant. Then the variation of the field should as well be given by

$$\delta|\psi\rangle \equiv \sum_{n=1}^{\infty} \mathscr{S}_{-n}|\chi_n\rangle = \sum_{n=1}^{\infty} L_{-n}|\Lambda_n\rangle\,, \tag{10.4}$$

for appropriately *constrained* parameters $|\Lambda_n\rangle$. On the other hand, one should be able to reintroduce the supplementary fields into the scheme presented here by relaxing the constraints $|\chi_n\rangle \in \mathbf{V}^{(0)}$ in (10.4).

However, the question of to what extent it is necessary or convenient to restore the supplementary fields is to be answered at the level of interacting string fields.

APPENDIX: FOCK SPACE, PHYSICAL STATES, AND VIRASORO OPERATORS

We use units in which $c = \hbar = 2\alpha' = 1$. The dynamics of a bosonic open string propagating in D-dimensional Minkowski space with metric $\eta_{\mu\nu} = \mathrm{diag}(-1,1,\ldots,1)$ is described by a set of oscillator variables $\alpha_n{}^\mu$ ($n \neq 0$ and integer), the center of mass coordinate x^μ, and the total momentum $p^\mu \equiv \alpha_0{}^\mu$. Quantization along the lines of the old covariant approach[1,2] gives rise to the operator algebra

$$[\alpha_n{}^\mu,\alpha_m{}^\nu] = \eta^{\mu\nu} n \delta_{n+m,0}\,,$$
$$[x^\mu,p_\nu] = i\delta^\mu_\nu\,, \tag{A1}$$
$$[x^\mu,\alpha_n{}^\nu] = 0, \quad \text{for } n\neq 0\,,$$

and the Hermiticity conditions

$$(\alpha_n{}^\mu)^\dagger = \alpha_{-n}{}^\mu, \quad (x^\mu)^\dagger = x^\mu\,. \tag{A2}$$

In the usual Fock representation the above algebra is realized by postulating a formal vacuum state $|0\rangle$, subject to

$$\alpha_n{}^\mu|0\rangle = 0\,, \tag{A3}$$

for $n\geqslant 0$. Since $n=0$ is included in this condition, some authors would refer to $|0\rangle$ as the "vacuum with respect to momentum $p^\mu = 0$." Acting upon $|0\rangle$ by products of the operators $i\alpha_{-n}{}^\mu$ $(n\geqslant 1)$ and forming linear combinations with real x-dependent coefficients, one arrives at the total Fock space **F**. The momentum operator is represented by

$$p_\mu = -i\frac{\partial}{\partial x^\mu} \equiv -i\,\partial_\mu\,. \tag{A4}$$

the action of the $\alpha_n{}^\mu$ $(n\geqslant 1)$ is evaluated by means of (A1) and (A3). Thus the general element of **F** has the form

$$|\psi\rangle = (\phi(x) - iA_\mu(x)\alpha_{-1}{}^\mu - iv_\mu(x)\alpha_{-2}{}^\mu - \tfrac{1}{2}h_{\mu\nu}(x)\alpha_{-1}{}^\mu\alpha_{-1}{}^\nu + \cdots)|0\rangle\,, \tag{A5}$$

where $h_{\mu\nu} = h_{\nu\mu}$ and the dots denote higher-order products of the α's. As far as the open-string Fock space is concerned, it will never be necessary to impose boundary conditions on the component fields ϕ, A_μ, etc. We merely assume them to be sufficiently differentiable. Formal adjunction leads to

$$\langle\psi| = \langle 0|(\phi(x) + iA_\mu(x)\alpha_1{}^\mu + iv_\mu(x)\alpha_2{}^\mu - \tfrac{1}{2}h_{\mu\nu}(x)\alpha_1{}^\mu\alpha_1{}^\nu + \cdots) \tag{A6}$$

and

$$\langle\psi|\psi'\rangle = \phi(x)\phi'(x) + A^\mu(x)A'_\mu(x) + 2v^\mu(x)v'_\mu(x) + \tfrac{1}{2}h^{\mu\nu}(x)h'_{\mu\nu}(x) + \cdots, \tag{A7}$$

where

$$\langle 0|0\rangle = 1 \tag{A8}$$

has been used. Note that the formal "scalar product" (A7) still depends on x^μ. In order to have a suitable notation for action functionals, we define

$$(\psi|\psi') = \int d^Dx\langle\psi|\psi'\rangle\,, \tag{A9}$$

which will in general not be finite.

The mass level operator is defined by

$$N = \sum_{n=1}^{\infty} \alpha_{-n}{}^\mu\alpha_{n\mu}\,. \tag{A10}$$

It splits **F** into a direct sum such that any $|\psi\rangle \in$ **F** may be decomposed uniquely as

$$|\psi\rangle = \sum_{n=0}^{\infty} |\psi\rangle_n\,, \tag{A11}$$

where

$$N|\psi\rangle_n = n|\psi\rangle_n\,. \tag{A12}$$

Any contribution to $|\psi\rangle$ of the form

$$\alpha_{-n_1}{}^{\mu_1}\cdots\alpha_{-n_k}{}^{\mu_k}A_{\mu_1\cdots\mu_k}(x)|0\rangle$$

has a definite value of N, namely $\Sigma_{i=1}^k n_i$. Note that (A11) is a formal sequence involving infinitely many space-time fields, rather than an actual sum. Thus any operator whose action upon any definite mass level is finite is well defined in the total Fock space. As a consequence, we will never run into convergence problems.

The Virasoro operators are defined by

$$L_n = \frac{1}{2}:\sum_{m=-\infty}^{\infty} \alpha_{n-m}{}^\mu\alpha_{m\mu}: = (L_{-n})^\dagger\,, \tag{A13}$$

where the normal ordering with respect to $\alpha_n{}^\mu$ $(n\geqslant 1)$ as annihilation operator is only necessary for L_0, leading to

$$L_0 = \tfrac{1}{2}p^\mu p_\mu + N \equiv -\tfrac{1}{2}\Box + N\,, \tag{A14}$$

where

$$\Box = \partial^\mu\partial_\mu = -\partial_{00} + \sum_{i=1}^{D-1}\partial_{ii}\,. \tag{A15}$$

The Virasoro operators constitute the quantum version of the classical constraints. They select the physical states by a set of Gupta–Bleuler type conditions (the so-called Virasoro conditions)

$$L_n|\psi\rangle = 0\,, \tag{A16}$$

for all $n\geqslant 1$, and

$$(L_0 - 1)|\psi\rangle = 0\,. \tag{A17}$$

As is well-known,[1,2] the absence of ghosts and a consistent theory of interacting strings is only guaranteed in the critical dimension $D = 26$.

The set of states satisfying only (A16) is denoted by $\mathbf{V}^{(0)}$. They are also called states of Verma level zero (cf. Ref. 13). The action of the L_n $(n\geqslant 1)$ upon the lowest mass levels of any state in **F** is given by

$$L_1|\psi\rangle_1 = -\partial_\mu A^\mu|0\rangle\,, \tag{A18}$$

$$L_1|\psi\rangle_2 = -2i(v_\mu - \tfrac{1}{2}\partial^\rho h_{\rho\mu})\alpha_{-1}{}^\mu|0\rangle\,,$$
$$L_2|\psi\rangle_2 = -2(\delta_\mu v^\mu + \tfrac{1}{4}h^\mu{}_\mu)|0\rangle\,, \tag{A19}$$

$|\psi\rangle_0$ being automatically subject to (A16), hence an element of $\mathbf{V}^{(0)}$.

The mass-shell condition (A17) reads, for the nth mass level,

$$(\Box - 2(n-1))|\psi\rangle_n = 0\,, \tag{A20}$$

thus giving rise to a squared mass

$$m_n^2 = 2(n-1) \tag{A21}$$

for the fields at this level. Here ϕ is the well-known tachyon and A_μ becomes a massless vector field.

The L_n's satisfy the Virasoro algebra

$$[L_n,L_m] = (n-m)L_{n+m} + (D/12)n(n^2-1)\delta_{n+m,0} \tag{A22}$$

including the anomalous term responsible for the crucial role played by the space-time dimension D.

It is convenient to denote the set of products $L_{n_1}\cdots L_{n_k}$ with $1\leqslant n_1\leqslant\cdots\leqslant n_k$ and $\Sigma_{i=1}^k n_i = n$ collectively by $\mathscr{L}_i^{(n)}$, where the index i ranges from 1 to a certain number ℓ_n. They altogether form a so-called "universal envelopping algebra" (cf. Ref. 12). The adjoint operators are denoted as

$$\mathscr{L}^{(n)}_{-i} = (\mathscr{L}^{(n)}_i)^\dagger \tag{A23}$$

and constitute an analogous set of products of Virasoro operators with negative indices. According to the relation

$$[N, \mathscr{L}^{(n)}_{\pm i}] = \mp n \mathscr{L}^{(n)}_{\pm i}, \tag{A24}$$

the $\mathscr{L}^{(n)}_{\pm i}$ change the mass level number by $\mp n$. We adopt the Einstein summation convention for indices of the type i.

As an example, we write down the operators $\mathscr{L}^{(3)}_i$,

$$\begin{aligned} \mathscr{L}^{(3)}_1 &= L_3, \\ \mathscr{L}^{(3)}_2 &= L_1 L_2, \\ \mathscr{L}^{(3)}_3 &= L_1^3, \end{aligned} \tag{A25}$$

and quote the lowest values of ℓ_n,

$$(\ell_0, \ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6, \ldots) = (1,1,2,3,5,7,11,\ldots),$$

where the convention $\ell_0 = 1$ is useful in some formulas.

Given a state $|\chi\rangle$ of Verma level zero that is at the same time an eigenvector of L_0 to a real eigenvector h (a "highest weight vacuum vector"),

$$L_0 |\chi\rangle = h |\chi\rangle, \tag{A26}$$

the linear space spanned by all states of the form $\mathscr{L}^{(n)}_{-i} |\chi\rangle$ is called a Verma module.[24] By making use of the Virasoro algebra (A22), one finds

$$\mathscr{L}^{(n)}_i \mathscr{L}^{(n)}_{-j} |\chi\rangle = \mathscr{M}^{(n)}_{ij}(h) |\chi\rangle, \tag{A27}$$

where $\mathscr{M}^{(n)}_{ij}$ is an $\ell_n \times \ell_n$ matrix of polynomials in h (the so-called Shapovalov matrix). It is generically invertible,[25] which means that its determinant (the Kac determinant) is a nontrivial polynomial and hence admits only a finite number of zeros. Given a value h for which all matrices $\mathscr{M}^{(n)}_{ij}(h)$ are nonsingular, the whole Verma module splits into a direct sum distinguished by the Verna level number n. For the purpose of the work presented here it is convenient to collect all states generated by the action of the $\mathscr{L}^{(n)}_{-i}$ upon any state of Verma level zero. By $\mathbf{V}^{(n)}$ we denote the set of states $\mathscr{L}^{(n)}_{-i} |\chi^i\rangle$ for which $|\chi^i\rangle \in \mathbf{V}^{(0)}$. As a result of the existence of zeros in the Kac determinants, the intersections $\mathbf{V}^{(m)} \cap \mathbf{V}^{(n)}$ are nontrivial. Let $m < n$ and $|\psi\rangle$ be an element of both $\mathbf{V}^{(m)}$ and $\mathbf{V}^{(n)}$,

$$|\psi\rangle = \mathscr{L}^{(n)}_{-i} |\chi^i\rangle = \mathscr{L}^{(m)}_{-j} |\xi^j\rangle, \tag{A28}$$

with $|\chi^i\rangle$ and $|\xi^j\rangle$ elements of $\mathbf{V}^{(0)}$. Acting upon this equation by $\mathscr{L}^{(n)}_k$, one obtains

$$\mathscr{M}^{(n)}_{ki}(L_0) |\chi^i\rangle = 0; \tag{A29}$$

thus, upon multiplication by the algebraic adjoint of $\mathscr{M}^{(n)}$,

$$\det \mathscr{M}^{(n)}(L_0) |\chi^i\rangle = 0,$$

which implies

$$\det \mathscr{M}^{(n)}(L_0 - n) |\psi\rangle = 0. \tag{A30}$$

This illustrates that the elements of the intersections $\mathbf{V}^{(n)} \cap \mathbf{V}^{(m)}$ are related to the zeros of the Kac determinants. One may show further from (A28) that $|\psi\rangle$ has zero norm. One is sometimes interested in computations involving "generic" states that do not satisfy equations like (A30). In these cases one may formally compute the inverse matrices

$$\mathscr{M}^{(n)}_{ij}(L_0)^{-1} = \mathscr{M}^{ij}_{(n)}(L_0) \tag{A31}$$

whose components are rational functions of L_0. They have been used to construct the projector P onto $\mathbf{V}^{(0)}$ (cf. Refs. 11 and 13).

Let $\mathbf{W}^{(n)}$ denote the set of all states $|\psi\rangle$ in $\mathbf{F}$ that are annihilated by the operators $\mathscr{L}^{(n+1)}_i$. Any such element may be written as a sum of lower level states,

$$|\psi\rangle = |\chi_0\rangle + \sum_{p=1}^{n} \mathscr{L}^{(p)}_{-i} |\chi^i_p\rangle, \tag{A32}$$

where all $|\chi\rangle$'s are in $\mathbf{V}^{(0)}$. To prove this, one just has to interpret (A32) as a set of differential equations for the $|\chi\rangle$'s. After multiplication by $\mathscr{L}^{(n)}_k$, one obtains an equation in $\mathbf{V}^{(0)}$ for $|\chi^i_n\rangle$ which—using the fact that $\det \mathscr{M}^{(n)}(L_0)$ is a nontrivial polynomial—admits a solution. The rest follows by induction. Since the $\mathbf{V}^{(p)}$ do not form a direct product, the decomposition (A32) is not unique, the arbitrariness again being connected with the zeros of the Kac determinants.

Applying (A32) separately to the different mass level contributions of a given state $|\psi\rangle$, and using $|\psi\rangle_n \in \mathbf{W}^{(n)}$, one finds that any element of $\mathbf{F}$ has the form

$$|\psi\rangle = |\chi_0\rangle + \sum_{p=1}^{\infty} \mathscr{L}^{(p)}_{-i} |\chi^i_p\rangle \tag{A33}$$

with all $|\chi\rangle$'s in $\mathbf{V}^{(0)}$. Usually, the states given by the second expression in (A33) are called spurious. From the point of view of gauge-covariant string field theory, these states carry only gauge degrees of freedom [cf. Eq. (1.4)]. By virtue of the field equations (A16) and (A17), they are excluded except for certain zero-norm states of the type (A30). Their occurrence—and thus the ambiguity of the decomposition (A33)—reflects the fact that upon imposing covariant gauge conditions like (A16) there is still some gauge freedom left.

ACKNOWLEDGMENTS

I would like to thank Helmut Rumpf for useful comments on the subject.

This work was supported by "Fonds zur Förderung der wissenschaftlichen Forschung in Österreich," Project No. P6241P.

[1]J. Scherk, Rev. Mod. Phys. **47**, 123 (1975).
[2]M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U. P., Cambridge, 1987), Vol. 1.
[3]W. Siegel, Phys. Lett. B **151**, 391, 396 (1985).
[4]W. Siegel and B. Zwiebach, Nucl. Phys. B **263**, 105 (1985).
[5]W. Siegel, in *Unified String Theories*, edited by M. Green and D. Gross (World Scientific, Singapore, 1986), p. 583.
[6]B. Zwiebach, contained in Ref. 5, p. 607.
[7]K. Itoh, T. Kugo, H. Kunitomo, and H. Ooguri, Phys. Rev. D **34**, 2360 (1986).
[8]E. Witten, Nucl. Phys. B **268**, 253 (1986).
[9]G. T. Horowitz, J. Lykken, R. Rohm, and A. Strominger, Phys. Rev. Lett. **57**, 283 (1986).
[10]M. Kaku, Univ. New York preprint HEP-CCNY-14, 1986.

[11]M. Kaku, Nucl. Phys. B **267**, 125 (1985).
[12]M. Kaku, Phys. Lett. B **162**, 97 (1985).
[13]T. Banks and M. E. Peskin, Nucl. Phys. B **264**, 513 (1986).
[14]T. Banks and M. E. Peskin, contained in Ref. 5, p. 621.
[15]A. Neveu and P. C. West, Nucl. Phys. B **268**, 125 (1986).
[16]A. Neveu, H. Nicolai, and P. C. West, Nucl. Phys. B **264**, 573 (1986).
[17]A. Neveu, J. Schwarz, and P. C. West, Phys. Lett. B **164**, 51 (1985).
[18]A. Neveu and P. C. West, Phys. Lett. B **165**, 63 (1985).
[19]A. Neveu, H. Nicolai, and P. C. West, Phys. Lett. B **167**, 307 (1986).
[20]A. Neveu and P. C. West, Phys. Lett. B **168**, 192 (1986).
[21]A. Neveu and P. C. West, Nucl. Phys. B **278**, 601 (1986).
[22]A. Neveu and P. C. West, Phys. Lett. B **165**, 63 (1985).
[23]E. S. Fradkin and V. I. Vilkovisky, Phys. Lett. B **73**, 209 (1978).
[24]V. Kac, *Infinite Dimensional Lie Algebras* (Birkhäuser, Basel, 1983).
[25]V. Kac, in *Lecture Notes in Physics*, Vol. 94, edited by W. Beiglböck, A. Böhm, and E. Takasugi (Springer, Berlin, 1979), p. 441.

Self-energy operator for an electron in an external Coulomb potential. II

Levere Hostler
Physics Department, Wilkes College, Wilkes Barre, Pennsylvania 18766 [a] *and Physics Department, Yale University, New Haven, Connecticut 06511*

(Received 18 September 1987; accepted for publication 6 April 1988)

Relativistic Coulomb Sturmian matrix elements of the operator $O \equiv \ln(1-\rho)/\rho$, $\rho = -[\pi\cdot(1+i\sigma)\cdot\pi]/m^2$, in terms of which the self-energy operator for an electron in an external Coulomb potential has been expressed, are studied. The operator O is dealt with on a term by term basis in a Sturmian expansion. Each term of the Sturmian expansion is separated into a part whose matrix elements are analytic functions of $Z\alpha$, plus a remainder evaluated in closed form by use of the Cauchy residue theorem. All ignorance about the matrix element of the general term in the Sturmian expansion of O is thereby placed entirely in the analytic part, for which an explicit integral representation is derived.

I. INTRODUCTION

The self-energy operator for an electron in an external Coulomb potential has been investigated earlier[1] using a mass eigenfunction expansion concept.[2] A formal representation of the self-energy Σ_0, neglecting "shift corrections,"[3] was derived:

$$\begin{aligned}\Sigma_0 = \frac{-\alpha}{4\pi}\Big\{&4\Pi\cdot(1-\rho)\frac{\ln(1-\rho)}{\rho}\cdot\Pi\\&+\Pi\cdot(1-\rho)\frac{(1-\rho)\ln(1-\rho)+\rho}{\rho^2}(1-i\sigma)\cdot\Pi\\&+\Pi\cdot(1-i\sigma)\frac{(1-\rho)\ln(1-\rho)+\rho}{\rho^2}(1-\rho)\cdot\Pi\\&+\frac{1}{3}\Pi\cdot(1-i\sigma)(1-\rho)\\&\times\frac{(1-\rho)^2\ln(1-\rho)+\rho-\frac{3}{2}\rho^2}{\rho^3}(1-i\sigma)\cdot\Pi\\&+\frac{1}{12}(1-i\sigma)_{\mu\nu}m^2\rho(1-\rho)\\&\times\frac{(1-\rho)^2\ln(1-\rho)+\rho-\frac{3}{2}\rho^2}{\rho^3}(1-i\sigma)_{\nu\mu}\\&-\frac{1}{36}(1-i\sigma)_{\mu\nu}m^2(1-\rho)(1-i\sigma)_{\nu\mu}\\&+\frac{31}{9}m^2(1-\rho)-3m^2+\frac{11}{9}e\sigma_{\mu\nu}F_{\mu\nu}\Big\},\end{aligned} \tag{1.1}$$

in which $\rho \equiv [-\Pi\cdot(1+i\sigma)\cdot\Pi]/m^2$ is a "mass operator," measured in units of m^2, for the electron in the external Coulomb potential.

The entire calculation was set in the framework of the "scalar formalism" for QED. In this formalism the Dirac equation for the electron is expressed in the second-order form,

$$\{\Pi\cdot(1+i\sigma)\cdot\Pi+m^2\}\Phi=0,\qquad \Pi\equiv(p-eA), \tag{1.2}$$

in which Φ is a 2×1 Pauli spinor and σ is a Lorentz spin tensor made up of the ordinary 2×2 Pauli spin matrices. The scalar formalism for QED brings out a close parallel between the quantum theory of a Dirac particle and the quantum theory of a scalar particle, with the row matrix $\bar{\Phi}\equiv\Phi^\dagger(-i\overleftarrow{\Pi}_4-\vec{\sigma}\cdot\overleftarrow{\Pi})$ playing the role of the simple Φ^* of a scalar particle.

Interest in the expression (1.1) arises through its connection with the Lamb shift, obtained by forming the expectation value of the self-energy operator in the unperturbed state. The main difficulty of calculating the contribution of Σ_0 to the Lamb shift lies in treating the complicated operator structures appearing in Eq. (1.1), such as the $\ln(1-\rho)/\rho$ term. In an effort to cope with this problem, these terms have been reexpressed in terms of the relativistic Coulomb Green's function, $1/(\rho'-\rho)$. For example,

$$\begin{aligned}\left\{\frac{\ln(1-\rho)}{\rho}\right\}' &= -\int_0^1 d\zeta\,\frac{(2-\zeta)}{\zeta}\,\frac{1}{(2q_0)^2-2(2q_0)\zeta+\zeta^2}\\&\quad\times\left[\frac{1}{\rho'-\rho}-\frac{\mathscr{P}_0}{\rho'-1}\right],\\ \rho'&=\frac{4(q_0)^2-4(q_0)^2\zeta+\zeta^2}{\zeta^2}.\end{aligned} \tag{1.3}$$

Known properties of the relativistic Coulomb Green's function can thus be brought to bear on the problem.[2,4,5] The prime on the left-hand side of Eq. (1.3) signifies deleting the term in the spectral representation belonging to the unperturbed state, and $\mathscr{P}_0$ is a projection operator onto that state. The parameter q_0 is of order $Z\alpha$, and has the physical interpretation of the classical linear momentum of the electron in the unperturbed state,

$$q_0=\frac{Z\alpha/(\gamma_0+n_0)}{[1+[Z\alpha/(\gamma_0+n_0)]^2]^{1/2}}. \tag{1.4}$$

The discrete expansion[4]

[a] Permanent address.

 0022-2488/88/081930-07$02.50

$$\frac{1}{\rho'-\rho}=\frac{m}{2q_0}Sr^{1/2}\sum_A P_A \times\frac{\zeta|\zeta,A\rangle\langle\zeta,A|}{2(\gamma_A+n_A)-\zeta(\gamma_A+n_A+\gamma_0+n_0)} \times S^{-1}r^{1/2},\quad |\zeta-1|<1, \tag{1.5}$$

of the relativistic Coulomb Green's function is one of the key equations needed in this study. The shorthand notation A is used to signify the set of quantum numbers n,L,J. The states $|\zeta,A\rangle$ refer strictly to the radial degree of freedom of the electron, and P_A stands for the angular momentum projection operator $P_A\equiv\Sigma_M|L,J,M\rangle\langle L,J,M|$. The operator P_A is actually independent of n. An operator similar to

$$S\equiv\cosh(\theta/2)+i\vec{\sigma}\cdot\hat{r}\sinh(\theta/2),$$
$$\theta\equiv\tanh^{-1}[Z\alpha/(\vec{\sigma}\cdot\vec{L}+1)],$$

has been introduced by Biedenharn[6] and Martin and Glauber[7] to simplify the Kepler problem for the linear Dirac equation.

The states $|\zeta,A\rangle$ are the Sturmian basis states. For any fixed real value of ζ, $0<\zeta<2$, and for fixed L and J the states $|\zeta,A\rangle$, $n=1,2,3,\ldots$, form an orthonormal basis of radial functions. The unperturbed state corresponds to $|\zeta=1,n_0,L_0,J_0\rangle$. In the following the term Sturmian "representation" shall refer to a representation with respect to the basis set $|\zeta=1,A\rangle\otimes|L,J,M\rangle$, while the term Sturmian "expansion" shall be reserved for the expansion (1.5), in whatever representation.

When Eq. (1.5) is substituted into Eq. (1.3), Eq. (1.3) goes over into the form

$$\left\{\frac{\ln(1-\rho)}{\rho}\right\}'=Sr^{1/2}OS^{-1}r^{1/2}, \tag{1.6}$$

where O is the integral

$$O=-2mq_0\int_0^1 d\zeta\frac{(2-\zeta)}{\zeta}\frac{1}{(4(q_0)^2-4(q_0)^2\zeta+\zeta^2)} \times\left\{\sum_{\text{all }A}P_A\frac{\zeta|\zeta,A\rangle\langle\zeta,A|}{2(\gamma_A+n_A)-\zeta(\gamma_A+n_A+\gamma_0+n_0)} -\sum_{\gamma_A+n_A=\gamma_0+n_0}P_A\frac{\zeta^2|\zeta=1,A\rangle\langle\zeta=1,A|}{2(\gamma_0+n_0)(1-\zeta)}\right\}. \tag{1.7}$$

It was shown in Ref. 1 that the q_0 dependence explicitly exhibited in the factor $[(2q_0)^2-2(2q_0)\zeta+\zeta^2]^{-1}$ of the integrand of O carries the main $Z\alpha$ dependence of the integrand in a Sturmian representation. The only other $Z\alpha$ dependence in a Sturmian representation occurs through the "relativistic angular momentum" quantum number $\gamma\equiv[(J+\frac{1}{2})^2-(Z\alpha)^2]^{1/2}-1$, if $L=J-\frac{1}{2}$, and $\gamma\equiv[(J+\frac{1}{2})^2-(Z\alpha)^2]^{1/2}$, if $\gamma=J+\frac{1}{2}$. To order $(Z\alpha)^2$ the quantum number γ is a constant $\approx L$.

This relatively simple $Z\alpha$ dependence of the integrand of O is exploited here in a term by term study of the series (1.7). A technique is developed for separting the ζ integrals for a general term $O_{n,L,J}$ of the series (1.7) into a part analytic in $Z\alpha$, plus a remainder which is known in closed form, and which carries any singular behavior. All ignorance about the matrix element of the general term in the Sturmian expansion of O is thereby placed entirely in the analytic part, for which an explicit integral representation is derived.

Our results take their simplest form when expressed in terms of O', an operator which differs from O primarily by terms that are projected to zero by the factor $(1-\rho)$ in Eq. (1.1). Equation (3.9) provides a relatively simple representation exhibiting the analytic part and remainder for the n,L,Jth term in the Sturmian expansion of O'. This separation into an analytic part plus remainder in closed form does not carry over to the full L,Jth partial wave of O', defined as $O'_{L,J}\equiv\Sigma_{n=1}^{\infty}O'_{n,L,J}$. However, the methods explored here on a term by term basis are expected to be applicable in a future investigation of the analytic behavior of the full L,Jth partial wave, $O'_{L,J}$.

In the interest of brevity the present calculation will be treated largely as a continuation of Ref. 1. The reader is referred to this earlier work and the references cited therein for background material.

II. METHOD OF SEPARATION OF AN ANALYTIC PART

It is convenient to work as far as possible in terms of the components, $g_A(\zeta)\equiv g_{n,L,J}(\zeta)$, in the Coulomb Sturmian expansion of the Coulomb Green's function:

$$\frac{1}{\rho'-\rho}\equiv Sr^{1/2}\sum_A P_{L,J}g_A(\zeta)r^{1/2}S^{-1}, \tag{2.1}$$

$$g_A(\zeta)=\frac{m}{2q_0}\frac{\zeta|\zeta,A\rangle\langle\zeta,A|}{2(\gamma_A+n_A)-\zeta(\gamma_A+n_A+\gamma_0+n_0)}. \tag{2.2}$$

The usual radial partial wave Coulomb Green's function, $g_{L,J}(\zeta)$, is given by

$$g_{L,J}(\zeta)=\sum_{n=1}^{\infty}g_A(\zeta). \tag{2.3}$$

Corresponding to Eq. (2.1) there is an expansion

$$O=\sum_A P_{L,J}O_A, \tag{2.4}$$

in which

$$O_A=-4(q_0)^2\int_0^1 d\zeta\frac{(2-\zeta)}{\zeta} \times\frac{1}{4(q_0)^2-4(q_0)^2\zeta+\zeta^2}g_A(\zeta),$$
$$\text{if }\gamma_A+n_A\neq\gamma_0+n_0; \tag{2.5a}$$

and

$$O_A=-4(q_0)^2\int_0^1 d\zeta\frac{(2-\zeta)}{\zeta} \times\frac{1}{4(q_0)^2-4(q_0)^2\zeta+\zeta^2} \times\frac{m}{2q_0}\frac{\zeta|\zeta,A\rangle\langle\zeta,A|-\zeta^2|\zeta=1,A\rangle\langle\zeta=1,A|}{(\zeta-1)(-2(\gamma_0+n_0))},$$
$$\text{if }\gamma_A+n_A=\gamma_0+n_0. \tag{2.5b}$$

The converging effect of the subtraction in Eq. (2.5b) is evidenced in the structure of the numerator $\zeta|\zeta,A\rangle$

$\times\langle\zeta,A\,|-\zeta^2|\zeta=1,A\rangle\langle\zeta=1,A\,|$, which has a zero at the upper limit $\zeta=1$.

The only Sturmian matrix elements of $P_{L,J}g_A(\zeta)$ not equal to zero are the matrix elements diagonal in the angular momentum quantum numbers, L,J,M. Accordingly, the only nonzero radial Sturmian matrix elements of $g_A(\zeta)$ that need be considered are of the type $\langle\zeta=1,C\,|g_A(\zeta)|\zeta=1,D\rangle$, with $L_C=L_D=L_A\equiv L$, $J_C=J_D=J_A\equiv J$, and $M_C=M_D$. These nonzero matrix elements will be signified in the following by use of a bracket notation: $[g_A(\zeta)]$ denotes the matrix whose matrix elements are

$$[g_A(\zeta)]_{C,D}\equiv\langle\zeta=1,n_C,L,J\,|g_A(\zeta)|\zeta=1,n_D,L,J\rangle\,.$$

For convenience a similar bracket notation is used for bras and kets, for example $[\,|\zeta,A\rangle]$ denotes the column matrix whose matrix elements are

$$[\,|\zeta,A\rangle]_C=\langle\zeta=1,n_C,L,J\,|\zeta,n,L,J\rangle\,.$$

The latter matrix elements can be computed using the formula[8]

$$\begin{aligned}\langle\zeta=1,&n_2,L,J\,|\zeta,n,L,J\rangle\\&=(-1)^{n-1}\left[\frac{(n_2+2\gamma)!(n+2\gamma)!}{(n_2-1)!(n-1)!}\right]^{1/2}\\&\quad\times\frac{\zeta^{\gamma+1}(2-\zeta)^{\gamma+1}}{(2\gamma+1)!}(1-\zeta)^{|n_2-n|}\\&\quad\times{}_2F_1(-(n_<-1),2\gamma+1+n_>,2\gamma+2;2\zeta-\zeta^2),\\&\quad n_<\equiv\min(n_2,n),\quad n_>\equiv\max(n_2,n)\,,\end{aligned}\tag{2.6}$$

for the overlap integral between two Sturmian basis states corresponding to different values of ζ. The formula (2.6) is combined with Eq. (2.2) to obtain the matrix elements $[g_A(\zeta)]$.

In the following there is a need to continue analytically the matrix elements $[g_A(\zeta)]$ into the complex ζ plane. The matrix elements $[g_A(\zeta)]$ are rendered single-valued functions of the complex variable ζ through the prescriptions $0<\operatorname{arc}(\zeta)<2\pi$; $|\operatorname{arc}(2-\zeta)|<\pi$. Near $\zeta=0$ the behavior of $[g_A(\zeta)]$ is $[g_A(\zeta)\propto\zeta^{(2\gamma+3)}]$. Because of the cut $0\leqslant\zeta<\infty$, the integral $\int_0^1 d\zeta$ could have two different values, according as one integrates along the upper sheet or the lower sheet; the integral over the lower sheet being exactly $e^{2\pi i2\gamma}$ times the integral over the upper sheet. The original real integrals (2.5a), (2.5b) correspond to integrals along the upper sheet. The separation of an analytic part exploits the double valuedness by first changing the original integral $\int_0^1 d\zeta$ into the contour integral $\int_1^{(0+)}d\zeta$ by use of the identity

$$\int_0^1 d\zeta\cdots=\frac{1}{(e^{2\pi i2\gamma}-1)}\int_1^{(0+)}d\zeta\cdots\,.\tag{2.7}$$

The contour C_1 is as shown in Fig. 1. There is an assumption here that when the loop $(0+)$ about the origin is taken infinitely small, then the contribution to the integral over the loop $(0+)$ vanishes. This assumes sufficiently fast vanishing of $\zeta^{(2\gamma+3)}$ in the neighborhood of $\zeta=0$. Next the contour C_1 is deformed into the unit circle $|\zeta|=1$.

Corrections for this include residues from the enclosed poles. The details of this vary depending upon the quantum numbers of O_A as follows.

$\gamma_A+n_A>\gamma_0+n_0$: The method of separating an analytic part will be illustrated first in the simplest case, that of Eq. (2.5a) with the assumption $\gamma_A+n_A>\gamma_0+n_0$. This assumption places the pole of the expression on the right-hand side of Eq. (2.2) outside the integration region $0\leqslant\zeta\leqslant1$ at

$$\zeta_A\equiv\frac{2(\gamma_A+n_A)}{\gamma_A+n_A+\gamma_0+n_0}>1\,.$$

The simple poles

$$\zeta_p\equiv2q_0e^{i(\pi/2)}e^{-i\theta_p}\,,\tag{2.8}$$

and

$$\begin{aligned}&(\zeta_p)^*\equiv2q_0e^{i(3\pi/2)}e^{i\theta_p}\,,\\&\theta_p=\tan^{-1}(Z\alpha/(\gamma_0+n_0))\,,\quad0<\theta_p<\pi/2\,,\end{aligned}\tag{2.9}$$

within the unit circle, $|\zeta|=1$, are the poles of the factor $[4(q_0)^2-4(q_0)^2\zeta+\zeta^2]^{-1}$ in the integrand of Eq. (2.5a).

To organize the work for this and subsequent calculations, it is convenient to introduce partial fraction expansions in order to be able to deal with one pole at a time. In the present example, the expansion

$$\begin{aligned}\frac{1}{4(q_0)^2-4(q_0)^2\zeta+\zeta^2}&=\frac{1}{(\zeta-\zeta_p)(\zeta-(\zeta_p)^*)}\\&=\frac{1}{(\zeta-\zeta_p)(\zeta_p-(\zeta_p)^*)}\\&\quad+\frac{1}{((\zeta_p)^*-\zeta_p)(\zeta-(\zeta_p)^*)}\,,\end{aligned}\tag{2.10}$$

is needed. Now let $f(\zeta)$ have the general form

$$f(\zeta)=\sum_{m=0}^{\infty}C_m\zeta^{2\gamma+m}\,,$$

where the series converges for $|\zeta|<\zeta_A$. Let ζ_p be any complex number for which $|\zeta_p|<1$. Then the formula

$$\begin{aligned}\int_0^1\frac{d\zeta}{(\zeta-\zeta_p)}f(\zeta)&=\frac{-2\pi i}{(e^{2\pi i2\gamma}-1)}f(\zeta_p)\\&\quad+\frac{1}{(e^{2\pi i2\gamma}-1)}\int_{|\zeta|=1}\frac{d\zeta}{(\zeta-\zeta_p)}f(\zeta)\end{aligned}\tag{2.11}$$

can be derived along the lines indicated above. The change of contour is illustrated in Fig. 1. If for the moment ζ_p is considered as a free parameter in Eq. (2.11), the contour integral would represent an analytic function of ζ_p near $\zeta_p=0$. Then if ζ_p is allowed to vary with $Z\alpha$ in accordance with Eq. (2.8) while holding γ constant; the contour integral would represent a function of $Z\alpha$ which is analytic near $Z\alpha=0$. However, for application of Eq. (2.5a) $f(\zeta)$ is taken to be

$$f(\zeta)\equiv\frac{\zeta\,[\,|\zeta,A\rangle][\langle\zeta,A\,|]}{2(\gamma_A+n_A)-\zeta(\gamma_A+n_A+\gamma_0+n_0)}\,,$$

which has a $Z\alpha$ dependence through γ as well as through ζ_p. When both ζ_p and γ are allowed to vary with $Z\alpha$; then a pole at $Z\alpha=0$ in the contour integral part of Eq. (2.11) is introduced through the factor $[(e^{2\pi i2\gamma}-1)]^{-1}$. In order to in-

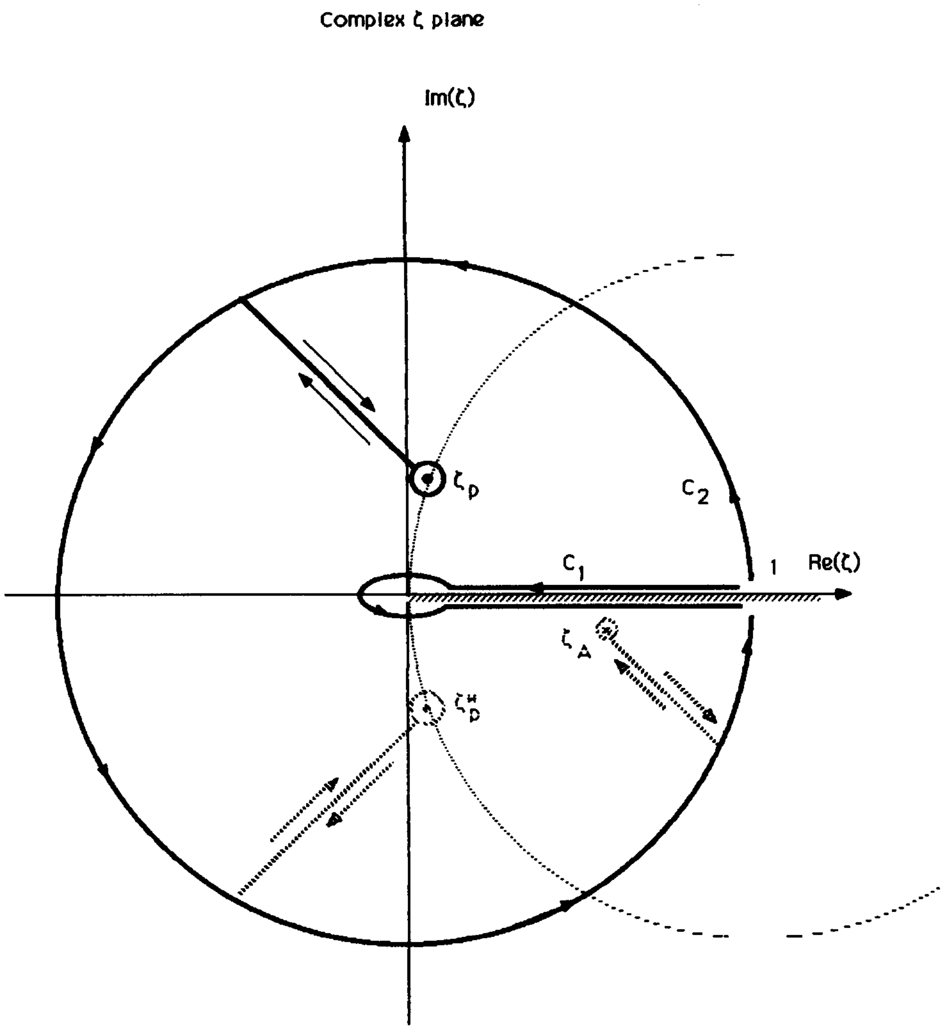


FIG. 1. Equivalent contours C_1 and C_2 used in the investigation of the integral (2.5a). Contour C_2 detours around the pole $\zeta = \zeta_p$ lying on the circle of convergence, $|\zeta - 1| = 1$, of the Sturmian expansion of the Green's function. Through the use of a partial fraction expansion the integral (2.5a) is split into a sum of integrals, each of which has only one pole within the unit circle $|\zeta| = 1$. For the integrals involving poles ζ_p^* or ζ_A the contour C_2 is changed by replacing the detour around ζ_p by the appropriate one of the two detours shown dimmed.

vestigate this the functions $[(\zeta - \zeta_p)]^{-1}$ and $f(\zeta)$ in Eq. (2.11) are expanded as

$$f(\zeta) = \sum_{m=0}^{\infty} C_m \zeta^{2\gamma + m},$$

and

$$\frac{1}{(\zeta - \zeta_p)} = \sum_{\kappa=0}^{\infty} \frac{(\zeta_p)^\kappa}{\zeta^{\kappa+1}};$$

and the contour integral is performed term by term. The resulting identity is

$$\int_0^1 \frac{d\zeta}{(\zeta - \zeta_p)} f(\zeta) = \frac{-2\pi i}{(e^{2\pi i 2\gamma} - 1)} f(\zeta_p) + \sum_{\kappa=0}^{\infty} \sum_{m=0}^{\infty} \frac{(\zeta_p)^\kappa C_m}{(2\gamma + m - \kappa)}. \quad (2.12)$$

The poles can be identified in Eq. (2.12): when $\kappa - m = 2L$, the denominator $(2\gamma + m - \kappa)$ is second-order small in $Z\alpha$. Next, the summation in (2.12) is rearranged according to the identity

$$\sum_{\kappa=0}^{\infty} \sum_{m=0}^{\infty} \frac{(\zeta_p)^\kappa C_m}{(2\gamma + m - \kappa)} = \sum_{\sigma=0}^{\infty} \frac{(\zeta_p)^\sigma}{(2\gamma - \sigma)} \sum_{m=0}^{\infty} (\zeta_p)^m C_m + \sum_{\sigma=1}^{\infty} \sum_{\kappa=0}^{\infty} \frac{(\zeta_p)^\kappa C_{\sigma+\kappa}}{(2\gamma + \sigma)}. \quad (2.13)$$

The offending terms are collected here in the first summation on the right-hand side of the equation. This summation takes the form of a factor times $(\zeta_p)^{-2\gamma} f(\zeta_p)$, and combines with the first term, $[-2\pi i/(e^{2\pi i 2\gamma} - 1)] f(\zeta_p)$, of Eq. (2.12) in such a way as to cancel the singularity of the factor $-2\pi i/(e^{2\pi i 2\gamma} - 1)$ in that term:

$$\int_0^1 \frac{d\zeta}{(\zeta - \zeta_p)} f(\zeta) = \left\{ \frac{-2\pi i}{(e^{2\pi i 2\gamma} - 1)} + \sum_{\sigma=0}^{\infty} \frac{(\zeta_p)^{\sigma - 2\gamma}}{(2\gamma - \sigma)} \right\} f(\zeta_p) + \sum_{\sigma=1}^{\infty} \sum_{\kappa=0}^{\infty} \frac{(\zeta_p)^\kappa C_{\sigma+\kappa}}{(2\gamma + \sigma)}. \quad (2.14)$$

The double sum remaining in Eq. (2.14) can be reexpressed as a contour integral by use of the Cauchy formula,

$$C_{\sigma+\kappa} = \frac{1}{2\pi i} \int_{(0+)} \frac{d\zeta}{\zeta^{\sigma+\kappa+1}} \zeta^{-2\gamma} f(\zeta), \quad (2.15)$$

for the Taylor expansion coefficient $C_{\sigma+\kappa}$ of the analytic function $\zeta^{-2\gamma} f(\zeta)$. In Eq. (2.15) (0+) may in principle be any contour about the origin which is within the region of analyticity $|\zeta| < \zeta_A$ of $\zeta^{-2\gamma} f(\zeta)$. However, we want to be able to interchange the order of summation and integration, and to reconstruct a closed expression for the integrand. This requires convergence of the resulting sums under the integral sign. The necessary convergence is achieved by taking the contour $(0+)$ to be a circle of radius R, where $1<R<\zeta_A$. We now have the identity

$$\int_0^1 \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) = \left\{ \frac{-2\pi i}{(e^{2\pi i 2\gamma}-1)} + \sum_{\sigma=0}^{\infty} \frac{(\zeta_p)^{\sigma-2\gamma}}{(2\gamma-\sigma)} \right\} f(\zeta_p) + \frac{1}{2\pi i} \int_{|\zeta|=R} \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) \zeta^{-2\gamma} \times \sum_{\sigma=1}^{\infty} \frac{(1/\zeta)^\sigma}{(2\gamma+\sigma)}; \tag{2.16}$$

in which

$$1<R<\zeta_{min}, \tag{2.17}$$

and $\zeta_{\min}$ signifies the least ζ_A for all levels with $\zeta_A>1$. With this way of choosing R, the same contour in Eq. (2.16) will suffice for all O_A. The new contour integral in Eq. (2.16) has a strictly analytic behavior as a function of $Z\alpha$ when both ζ_p and γ vary.

$\gamma_A+n_A<\gamma_0+n_0$: The processing of O_A for $\gamma_A+n_A<\gamma_0+n_0$, equivalently $\zeta_A<1$, proceeds along similar lines. This time the partial fraction expansion of $[(\zeta-\zeta_p)(\zeta-(\zeta_p)^*)(\zeta-\zeta_A)]^{-1}$ is used to separate the contribution into three terms, each of which has only one pole. Each of the three terms is treated using the same identity (2.16), even the term with the pole at ζ_A. The function $f(\zeta)$ in Eq. (2.16) is taken to be $f(\zeta)=\zeta[|\zeta,A\rangle][\langle\zeta,A|]$ for all three terms. The detours needed to process the two terms belonging to the poles ζ_p^* and ζ_A are shown dimmed in Fig. 1. The appropriate one of these detours replaces the detour around ζ_p when the terms belonging to the other poles are processed.

$\gamma+n_A=\gamma_0+n_0$: The case $\gamma_A+n_A=\gamma_0+n_0$ requires special treatment. Now O_A is given by the integral (2.5b). Only the first of the two terms $\zeta|\zeta,A\rangle\langle\zeta,A|$ and $\zeta^2|\zeta=1,A\rangle\langle\zeta=1,A|$ in the numerator behaves at $\zeta=0$ like $\zeta^{2\gamma+3}$ and seems subject to the above technique. On the other hand, direct separation of the integral into two parts corresponding to the two dissimilar terms leads to divergences at the upper limit $\zeta=1$. A limiting process can be used to deal with this complication of the $\gamma_A+n_A=\gamma_0+n_0$ case: the upper limit in Eq. (2.5b) is changed to h, where h is a number near 1, but less than 1. Then the two types of terms are processed separately, the results are combined, and a final limit is taken in which $h\to 1^-$. The integral with the new limit $h<1$ involving the first term $\zeta|\zeta,A\rangle\langle\zeta,A|$ behaves like the first case $\gamma_A+n_A>\gamma_0+n_0$ treated above; since ζ_p and $(\zeta_p)^*$ are again the only poles within the circle $|\zeta|=h$ into which the starting contour is deformed. All the steps in the above derivation of Eq. (2.14) can be followed for the $\zeta|\zeta,A\rangle\langle\zeta,A|$ term with the new upper limit, and one obtains just a slightly modified form of Eq. (2.14):

$$\int_0^h \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) = \left\{ \frac{-2\pi i}{(e^{2\pi i 2\gamma}-1)} + \sum_{\sigma=0}^{\infty} \frac{(\zeta_p/h)^{\sigma-2\gamma}}{(2\gamma-\sigma)} \right\} f(\zeta_p) + \sum_{\sigma=1}^{\infty}\sum_{\kappa=0}^{\infty} \frac{(\zeta_p)^\kappa h^{2\gamma+\sigma} C_{\sigma+\kappa}}{(2\gamma+\sigma)}. \tag{2.18}$$

Up to this point there is no basic difference between the new case and the $\gamma_A+n_A>\gamma_0+n_0$ case treated before. As before the next step to use the Cauchy integral formula (2.15) to represent the Taylor expansion coefficient $C_{\sigma+\kappa}$, and then to interchange the order of summation and integration. To ensure convergence of the sums that are formed under the integral sign the radius of the contour must be taken $>h$; while to ensure that the pole at $\zeta=1$ of the integrand is not enclosed the contour should have radius <1. These conditions are consistent; since $h<1$. Accordingly, we can get to the step corresponding to Eq. (2.16), in the new calculation; except that now the radius of the contour must lie between h and 1. On the other hand it is convenient to keep the radii of all the contour integrals the same. For this reason, the contour of integration in the analog of Eq. (2.16) is expanded beyond $h<|\zeta|<1$ to $|\zeta|=R$, where $R>1$ is the same as before. When this last step is taken, a residue correction for passing over the pole of the integrand at $\zeta=1$ is encountered. The final formula needed to process the first term of Eq. (2.5b) is similar to Eq. (2.16), but has this additional residue term besides incorporating the effect of the new upper limit:

$$\int_0^h \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) = \left\{ \frac{-2\pi i}{(e^{2\pi i 2\gamma}-1)} + \sum_{\sigma=0}^{\infty} \frac{(\zeta_p/h)^{\sigma-2\gamma}}{(2\gamma-\sigma)} \right\} f(\zeta_p) + \frac{1}{2\pi i} \int_{|\zeta|=R} \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) \sum_{\sigma=1}^{\infty} \frac{(h/\zeta)^{2\gamma+\sigma}}{(2\gamma+\sigma)} - \frac{1}{(1-\zeta_p)} \operatorname{Res}_{\zeta=1} f(\zeta) \sum_{\sigma=1}^{\infty} \frac{h^{2\gamma+\sigma}}{(2\gamma+\sigma)}. \tag{2.19}$$

Equation (2.19) assumes that $f(\zeta)\zeta^{-2\gamma}$ is analytic within the circle $|\zeta|<\zeta_{\min}$ except for a simple pole at $\zeta=1$, and that $|\zeta_p|<h<1<R<\zeta_{\min}$. For application to the first term of Eq. (2.5b) $f(\zeta)$ is taken to be

$$f(\zeta) = \frac{(2-\zeta)}{\zeta} \frac{\zeta[|\zeta,A\rangle][\langle\zeta,A|]}{(\zeta-1)}.$$

Notice that in the limit $h\to 1^-$ all terms on the right-hand side of Eq. (2.19) approach definite limits, except the residue term, where the divergent infinite series $\Sigma_{\sigma=1}^{\infty}[1/(2\gamma+\sigma)]$ is encountered.

The final type of term that must be treated is the second term of Eq. (2.5b). Even though the integral of this term is elementary, there seems to be an advantage in using an analysis that follows as closely as possible the treatment of the other terms. This is achieved by use of the identity,[9] analogous to Eq. (2.7),

$$\int_0^h \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) = \frac{1}{2\pi i} \int_h^{(0+)} \frac{d\zeta}{(\zeta-\zeta_p)} \ln(\zeta) f(\zeta), \quad 0<\operatorname{arc}(\zeta)<2\pi, \tag{2.20}$$

which assumes that the function $f(\zeta)$ is analytic in the circle $|\zeta|<1$. If the general method of deriving Eq. (2.19) is followed here, the identity

$$\int_0^h \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) = \left\{ -\ln\left(\frac{\zeta_p}{h}\right) + \ln\left(1-\frac{\zeta_p}{h}\right) + \pi i \right\} f(\zeta_p) - \frac{1}{2\pi i}\int_{|\zeta|=c;\ h<c<1} \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) \ln\left(1-\frac{h}{\zeta}\right), \tag{2.21}$$

emerges. Now let $f(\zeta)$ be analytic in the larger circle $|\zeta| < \zeta_{\min}$ except for a pole at $\zeta = 1$. In the application at hand $f(\zeta) = [(2-\zeta)/\zeta][\zeta^2/(\zeta-1)]$. Then the contour in Eq. (2.21) can be enlarged to $|\zeta| = R$, provided that, again, a residue correction for passing over the pole at $\zeta = 1$ is incorporated. In this way the contour for Eq. (2.21) can be given the same radius as for the other cases treated above,

$$\int_0^h \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) = \left\{ -\ln\left(\frac{\zeta_p}{h}\right) + \ln\left(1-\frac{\zeta_p}{h}\right) + \pi i \right\} f(\zeta_p) - \frac{1}{2\pi i}\int_{|\zeta|=R} \frac{d\zeta}{(\zeta-\zeta_p)} f(\zeta) \ln\left(1-\frac{h}{\zeta}\right) + \frac{1}{(1-\zeta_p)} \operatorname{Res}_{\zeta=1} f(\zeta)\ln(1-h) . \tag{2.22}$$

Again, each term on the right-hand side of the identity separately approaches a definite limit as $h \to 1^-$, except the residue correction term.

III. SUMMARY OF RESULTS FOR THE n,L,Jth TERM IN THE STURMIAN EXPANSION

$n_A + \gamma_A > n_0 + \gamma_0$: Equation (2.16) is used with

$$f(\zeta) = \frac{\zeta[|\zeta,A\rangle][\langle\zeta,A|]}{2(n_A+\gamma_A) - \zeta(n_A+\gamma_A+n_0+\gamma_0)}$$

to transform Eq. (2.5a) into

$$O_A = \left\{ \frac{-4(q_0)^2}{(\zeta_p - (\zeta_p)^*)} \rho(\zeta_p) \frac{(2-\zeta_p)}{\zeta_p} g_A(\zeta_p) + \zeta_p \leftrightarrow (\zeta_p)^* \right\} - 4(q_0)^2 \frac{1}{2\pi i} \int_{|\zeta|=R} \frac{d\zeta}{(4(q_0)^2 - 4(q_0)^2\zeta + \zeta^2)} \frac{(2-\zeta)}{\zeta} g_A(\zeta)\tau(\zeta) , \tag{3.1}$$

$$\rho(\zeta) \equiv \frac{-2\pi i}{(e^{2\pi i 2\gamma}-1)} + \sum_{\sigma=0}^{\infty} \frac{(\zeta)^{\sigma-2\gamma}}{2\gamma-\sigma} , \tag{3.2}$$

$$\tau(\zeta) \equiv \sum_{\sigma=1}^{\infty} \frac{(1/\zeta)^{\sigma+2\gamma}}{2\gamma+\sigma} . \tag{3.3}$$

$n_A + \gamma_A < n_0 + \gamma_0$: This time Eq. (2.16) is used with $f(\zeta) = \zeta[|\zeta,A\rangle][\zeta,A|]$, and, as mentioned above, the partial fraction expansion of $[(\zeta-\zeta_p)(\zeta-(\zeta_p)^*)(\zeta-\zeta_A)]^{-1}$ is used to separate the integral into three parts having only one pole each. The partial fraction expansion can be reversed to some extent after Eq. (2.16) is applied, with the result

$$O_A = \left\{ \frac{-4(q_0)^2}{(\zeta_p - (\zeta_p)^*)} \rho(\zeta_p) \frac{(2-\zeta_p)}{\zeta_p} g_A(\zeta_p) + \zeta_p \leftrightarrow (\zeta_p)^* \right\} - 4(q_0)^2 \frac{1}{2\pi i} \int_{|\zeta|=R} \frac{d\zeta}{(4(q_0)^2 - 4(q_0)^2\zeta + \zeta^2)} \frac{(2-\zeta)}{\zeta} g_A(\zeta)\tau(\zeta) + \frac{2q_0 m}{4(q_0)^2 - 4(q_0)^2\zeta_A + (\zeta_A)^2} \rho(\zeta_A)(2-\zeta_A) \frac{\zeta_A[|\zeta_A,A\rangle][\langle\zeta_A,A|]}{2(\gamma_A+n_A)} . \tag{3.4}$$

$\gamma_A + n_A = \gamma_0 + n_0$: This time Eq. (2.19) is needed with

$$f(\zeta) = \frac{(2-\zeta)}{\zeta} \frac{\zeta[|\zeta,A\rangle][\langle\zeta,A|]}{(\zeta-1)}$$

and Eq. (2.22) with

$$f(\zeta) = (2-\zeta)\zeta^2/\zeta(\zeta-1) .$$

After taking the limit $h \to 1^-$, Eq. (2.5b) goes over into

$$O_A = \left\{ \frac{-4(q_0)^2}{(\zeta_p - (\zeta_p)^*)} \rho(\zeta_p) \frac{(2-\zeta_p)}{\zeta_p} g_A(\zeta_p) + \zeta_p \leftrightarrow (\zeta_p)^* \right\} - 4(q_0)^2 \frac{1}{2\pi i} \int_{|\zeta|=R} \frac{d\zeta}{(4(q_0)^2 - 4(q_0)^2\zeta + \zeta^2)} \frac{(2-\zeta)}{\zeta} g_A(\zeta)\tau(\zeta)$$

$$+\left\{\frac{4(q_0)^2}{(\zeta_p-(\zeta_p)^*)}\frac{m}{2q_0}\frac{[|\zeta=1,A\rangle][\langle\zeta=1,A|]}{(-2(\gamma_0+n_0))}\left[\frac{\Psi(1)-\Psi(2\gamma+1)}{(1-\zeta_p)}\right.\right.$$

$$\left.\left.+\frac{\zeta_p(2-\zeta_p)}{(\zeta_p-1)}(-\ln(\zeta_p)+\ln(1-\zeta_p)+\pi i)+1\right]+\zeta_p\leftrightarrow(\zeta_p)^*\right\}. \tag{3.5}$$

Details of the limiting process $h\to 1^-$ are as follows. As noted above, only the residue terms of Eqs. (2.19) and (2.22) fail to have limits individually. When these residue terms for $h<1$ are combined, the following expression is obtained:

$$\frac{4(q_0)^2}{(\zeta_p-(\zeta_p)^*)}\frac{m}{2q_0}\frac{1}{(-2(\gamma_0+n_0))}\frac{[|\zeta=1,A\rangle][\langle\zeta=1,A|]}{(1-\zeta_p)}\left\{\frac{h^{2\gamma}}{2\gamma}{}_2F_1(1,2\gamma,2\gamma+1;h)-\frac{h^{2\gamma}}{2\gamma}+\ln(1-h)\right\}. \tag{3.6}$$

Now the limit can be taken with the help of the relation[10]

$$(1/2\gamma)_2F_1(1,2\gamma,2\gamma+1;h)\to-(-\Psi(1)+\Psi(2\gamma))-\ln(1-h)+\cdots, \tag{3.7}$$

in which the dots signify terms vanishing as $h\to 1^-$. The relation (3.5) incorporates the identity $\Psi(z+1)=\Psi(z)+1/z$ obeyed by the logarithmic derivative of the gamma function.

Simplified O: When Eq. (3.5) is applied in the expression (1.1) for the self-energy operator, the factor $(1-\rho)$ occurring there will project the objects

$$r^{1/2}S[|\zeta=1,A\rangle][\langle\zeta=1,A|]r^{1/2}S^{-1}$$

to zero. Accordingly, the $[|\zeta=1,A\rangle][\langle\zeta=1,A|]$ terms in Eq. (3.5) can simply be dropped. A new operator O' is thereby defined which is quite a bit simpler than O and which will serve equally as well. If the bound state contributions corresponding to states having $\gamma_A+n_A<\gamma_0+n_0$ are further treated separately, we can write O' in a form

$$O'=\sum_{L,J}P_{L,J}\sum_{n=1}^{\infty}O'_{n,L,J}+\sum_{\gamma_A+n_A<\gamma_0+n_0}P_{L,J}B_A,$$

$$B_A=\sum_{\gamma_A+n_A<\gamma_0+n_0}\frac{2q_0m}{4(q_0)^2-4(q_0)^2\zeta_A+(\zeta_A)^2}\rho(\zeta_A)(2-\zeta_A)\frac{\zeta_A[|\zeta_A,A\rangle][\langle\zeta_A,A|]}{2(\gamma_A+n_A)}, \tag{3.8}$$

for which there is a relatively simple uniform prescription for the separation of the n,L,J th term into an analytic part plus a remainder in closed form:

$$O'_{n,L,J}=-4(q_0)^2\frac{1}{2\pi i}\int_{|\zeta|=R}\frac{d\zeta}{(4(q_0)^2-4(q_0)^2\zeta+\zeta^2)}\frac{(2-\zeta)}{\zeta}g_{n,L,J}(\zeta)\tau(\zeta)$$

$$+\left[\frac{-4(q_0)^2}{(\zeta_p-(\zeta_p)^*)}\rho(\zeta_p)\frac{(2-\zeta_p)}{\zeta_p}g_{n,L,J}(\zeta_p)+\zeta_p\leftrightarrow(\zeta_p)^*\right]. \tag{3.9}$$

The densities ρ and τ are as defined above in Eqs. (3.2) and (3.3).

It will be noted that, by virture of Eq. (2.3), the sum on n can be performed in closed form for the remainder terms of Eq. (3.9), with a result proportional to the full L,J th partial wave Green's function, evaluated at $\zeta=\zeta_p$, together with a similar term having ζ_p and ζ_p^* interchanged. Reference to the defining equation of ζ_p shows that $\rho'=0$. Accordingly, the corresponding Green's function is evaluated at zero mass and at energy E_0. We are prevented from simplifying the sum over analytic parts in a similar way: in the sum on n of the analytic parts, interchange of the order of summation and integration leads to a series under the integral sign that diverges for ζ values on the part of the integration contour lying outside the circle $|\zeta-1|=1$, the circle of convergence of the Sturmian expansion (1.5).

ACKNOWLEDGMENTS

This work was carried out during the summer of 1987 while I was a Research Affiliate in the Physics Department at Yale University. I am grateful to the Physics Department at Yale University for their hospitality.

This research was supported by the National Science Foundation under Grant No. PHY-8603769.

[1]L. Hostler, J. Math. Phys. **28**, 720 (1987).
[2]L. Hostler, J. Math. Phys. **26**, 124 (1985).
[3]G. W. Erickson and D. R. Yennie, Ann. Phys. (NY) **35**, 271, 447 (1965).
[4]L. Hostler, J. Math. Phys. **24**, 2366 (1983).
[5]L. Hostler, J. Math. Phys. **28**, 2984 (1987).
[6]L. C. Biedenharn, Phys. Rev. **126**, 845 (1962).
[7]P. C. Martin and R. J. Glauber, Phys. Rev. **109**, 1307 (1958).
[8]Equation (4.7) of Ref. 1.
[9]This identity might be used to apply our method to O in a "Furry approximation," $\gamma\approx L$. Such an approach would hold the promise of providing a separation into an analytic part plus a remainder in closed form for the full O, as opposed to just the individual terms of a Sturmian expansion. Details of this will perhaps be the subject of future research.
[10]H. Bateman, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 1, p. 110, Eq. (12).

On the monotonic dependence of atomic excitation energies on the nuclear charge parameter Z

H. Hogreve[a)]
Centre de Physique Théorique,[b)] *CNRS-Luminy-Case 907, F-13288 Marseille Cedex 9, France*

(Received 21 April 1987; accepted for publication 23 March 1988)

The dependence of atomic excitation energies, i.e., differences of two energy levels, on the nuclear charge parameter Z is examined. For various transitions of heliumlike atoms the associated excitation energies are shown to be monotonic functions of Z. These results and physical arguments suggest the conjecture that such a monotony is a general fact in the nonrelativistic Schrödinger theory of atoms. The proofs given here use upper and lower energy bounds that for two-electron systems are available as explicit expressions in closed form.

I. INTRODUCTION

The mathematical analysis of Schrödinger operators has become a rather active field during the past decades. But in most cases, these investigations have concentrated on rather general and abstract aspects of the theory. As a consequence, whereas these parts of the field have seen substantial progress, for the Schrödinger atom only a relatively small number of rigorous results exist and many interesting questions remain open. Despite the fact that often problems considered have answers that are more or less obvious on physical grounds or can be verified by numerical computations, mathematical *ab initio* proofs assuming as their starting point the Schrödinger equation only turn out to be, in general, remarkably difficult. Examples of such problems that are not yet satisfactorily solved include the proof of the nonexistence of doubly negative atomic or molecular ions[1] and the monotonicity of the ionization energies.[2]

Here, we will formulate and analyze a further problem in this context. Within the nonrelativistic Schrödinger theory, an N-electron atom (in the infinite nuclear mass approximation) is described by the Hamiltonian

$$H=\sum_{i=1}^{N}\left(-\frac{1}{2}\Delta_i-\frac{Z}{r_i}\right)+\sum_{i<j}\frac{1}{r_{ij}}. \quad (1)$$

The operator H acts on the antisymmetric tensor product $A\{L^2(\mathbb{R}^{3N})\otimes \otimes_N \mathbb{C}^2\}$ and is known to be self-adjoint on an appropriate domain constructed with the second Sobolev space $W^2(\mathbb{R}^{3N})$. Furthermore, for sufficiently large nuclear charge Z, the Hamiltonian H has a nonempty discrete spectrum $\sigma_{\rm dis}(H)\subset\sigma(H)$. The elements of $\sigma_{\rm dis}(H)$ are denoted by E_k (arranged such that $E_1\leqslant E_2\leqslant\cdots$) and interpreted as the discrete energy levels of the atom. As a function of Z, these energies $E_k(Z)$ are easily shown to be monotone decreasing, which is just the manifestation of stronger binding when the nuclear charge is increased. Moreover, if the $E_{k,i}$ denote the eigenvalues within a subspace of definite symmetry of H, the sum $\Sigma_{i=1}^{r}E_{k,i}$ of the r-lowest energies has been demonstrated to be concave in Z.[3] In particular, this implies that each ground-state energy itself is concave in Z. On the other hand, the concavity of any excited energy level is not known.

Physically more interesting than the sum of the energies are their differences E_i-E_j which for $i<j$ give the excitation energies corresponding to a transition between the ith and kth level. Although there is no convincing reason to expect that such differences should also show a concave or convex behavior, the comparison of experimental energy levels indicates that these excitation energies are monotonic in Z. In fact, in all of those examples that we have extracted from the extensive amount of available spectroscopic data,[4] it turned out that the wave numbers associated with corresponding transitions within fixed isoelectronic series are strictly monotonically increasing in Z. Hence the question whether this observed monotonicity is a universal property of all those systems that are modeled by (1) is certainly an interesting and physically well motivated problem. A heuristical argument supporting the monotonic behavior is based on the assumption that the atomic electrons are distributed according to the shell model: By increasing Z, the inner shells are more attracted to the center, whereas the influence of the additional nuclear charge on the outer shells (associated with more excited states) is less strong due to the larger distance from the nucleus.

For the hydrogen atom the solution is explicit: for $n_1<n_2$ the energy difference $E_{n_2lm}(Z)-E_{n_1lm}(Z)=(\frac{1}{2})(n_2^2-n_1^2)Z^2/(n_2^2\times n_1^2)$ is monotonic increasing and convex in Z. The nontrivial geometrical configuration caused by the electron-electron repulsion and the occurrence of screening effects renders the situation less clear for more complex atoms. Below we shall succeed in proving the monotonic increase of various different excitation energies for the simplest nontrivial multielectron system, i.e., a helium-type atom with two electrons. The method that we employ consists of reducing the estimates of the Z derivatives of the excitation energies to relations involving the corresponding energies and to apply energy bounds that are sufficiently sharp but still analytically tractable.

Since selection rules do not allow the direct observation of electromagnetic radiation associated to transitions within symmetry subspaces of, e.g., constant total angular momentum, the behavior of excitation energies for transitions between subspaces with different symmetries also needs to be examined. For heliumlike systems, the Hamiltonian (1) commutes with the spin operators of each electron, the parity operator, and the total angular momentum $\mathbf{L}=\mathbf{L}_1+\mathbf{L}_2$.

[a)] Current address: Hahn–Meitner Institut, Postfach 390128, D-1000 Berlin 39, West Germany.
[b)] Laboratoire propre du CNRS–LP 7061.

 0022-2488/88/081937-06$02.50

Accordingly, the eigenstates of H can be classified into the usual sectors ${}^{2S+1}L$, where $S = s_1 + s_2$ is the total spin and L is the total angular momentum. Using similar arguments as for the transitions within fixed symmetry sectors, we shall demonstrate the monotony for some transitions with $\Delta L = \pm 1$ and $\Delta S = \pm \frac{1}{2}$. Because the excitation energies are proportional to the frequencies of the emitted or absorbed photons, this implies that the radiation frequencies for the corresponding transitions are increasing as one moves to higher ionized members of the helium isoelectronic series H^-, He, Li^+, Be^{2+},... .

II. SYMMETRY CONSERVING TRANSITIONS

In the rest of this paper we consider only heliumlike systems characterized by the Hamiltonian (1) with $N=2$. Furthermore, we assume that the nuclear charge parameter Z is such that the discrete spectrum $\sigma_{\rm dis}(H|^{2S+1}L)$ in each of the considered symmetry subspaces is not empty, i.e., there exists at least a ground-state energy $E_1(Z|^{2S+1}L)$ and a ground state $\psi_1 \in \mathcal{A}\{L^2(\mathbb{R}^6)\otimes \mathbb{C}^2 \otimes \mathbb{C}^2\}$ such that $H\psi_1 = E_1\psi_1$, $\mathbf{L}^2\psi_1 = L(L+1)\psi_1$, and $\mathbf{S}^2\psi_1 = S(S+1)\psi_1$. Let $E_2(Z|^{2S+1}L) = \inf(\sigma(H|^{2S+1}L)\setminus\{E_1(Z|^{2S+1}L)\})$ be the next higher point of the spectrum of H restricted to the respective symmetry sector ${}^{2S+1}L$.

Our aim is to show that the excitation energy

$$\Delta E(Z|^{2S+1}L) := E_2(Z|^{2S+1}L) - E_1(Z|^{2S+1}L) \qquad (2)$$

as a function of Z is monotone increasing.

To this end we will try to prove the non-negativity

$$\frac{d\Delta E(Z|^{2S+1}L)}{dZ} \geqslant 0. \qquad (3)$$

Applying the Feynman–Hellman and virial theorems,[5] the above derivative can be expressed as

$$\frac{d\Delta E(Z|^{2S+1}L)}{dZ} = \frac{1}{Z}\Big\{2\Delta E(Z|^{2S+1}L) - \Big\langle \psi_2, \frac{1}{r_{12}}\psi_2\Big\rangle + \Big\langle \psi_1, \frac{1}{r_{12}}\psi_1\Big\rangle\Big\}. \qquad (4)$$

For (4) we have assumed also that $E_2(Z|^{2S+1}L)\in\sigma_{\rm dis}(H|^{2S+1}L)$ and therefore its associated eigenstates ψ_2 are L^2 functions. If this is not the case, then $E_2(Z|^{2S+1}L) = \inf \sigma_{\rm ess}(H|^{2S+1}L)$ is an ionization threshold and thus of the form $E_2(Z|^{2S+1}L) = -Z^2/(2n^2)$ with n depending on the respective symmetry subspace. Since

$$0 > E_2(Z|^{2S+1}L) > E_1(Z|^{2S+1}L)$$

and

$$dE_2(Z|^{2S+1}L)/dZ = 2\times E_2(Z|^{2S+1}L),$$

the derivative

$$\frac{d\Delta E(Z|^{2S+1}L)}{dZ} = Z^{-1}\Big\{2\Delta E(Z|^{2S+1}L) + \Big\langle \psi_1, \frac{1}{r_{12}}\psi_1\Big\rangle\Big\} > 0,$$

is positive, and hence we have proved the following proposition.

Proposition 1: If $E_2(Z|^{2S+1}L)$ does not belong to the discrete spectrum of $H|^{2S+1}L$, then the corresponding ionization energies $\Delta E(Z|^{2S+1}L)$ are strongly monotonic increasing in Z.

Remark 2: Obviously Proposition 1 remains true if we do not restrict E_2 to the same symmetry sector as E_1.

Now, to show the non-negativity (3) for E_2 in the discrete spectrum, we have to estimate the quantities in (4) appropriately. Whereas we do not know of a possibility to compare directly the matrix elements of $1/r_{12}$ within the ground state ψ_1 and the excited state ψ_2, it is possible to estimate them in terms of energies; then for energies we can fall back on variational and projector bounds. We need an upper bound on $\langle\psi_2,(1/r_{12})\psi_2\rangle$ and a lower bound on $\langle\psi_1,(1/r_{12})\psi_1\rangle$. In the sequel, concavity arguments will play an important role; so instead of $\langle\psi_2,(1/r_{12})\psi_2\rangle$ we shall estimate the sum $\Sigma_{i=1}^r\langle\psi_2,(1/r_{12})\psi_2\rangle$ from above.

Lemma 3: Let E_i for $i=1,...,r$ be the r-lowest eigenvalues of H in a fixed symmetry subspace, and ψ_i the associated eigenfunctions. Then

$$\text{(i)}\quad \langle\psi_1,(1/r_{12})\psi_1\rangle \geqslant 2\big(\sqrt{|E_1^{UB}|\times|E_1^B|} + E_1^{LB}\big), \qquad (5)$$

where E_1^{LB} and E_1^{UB} are lower and upper bounds on E_1 and the E_k^B are the eigenvalues of a noninteracting hydrogenic two-electron problem

$$\Big\{-\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 - \frac{Z}{r_1} - \frac{Z}{r_2}\Big\}\psi_k^B = E_k^B\psi_k^B \qquad (6)$$

restricted to the respective symmetry sector;

$$\text{(ii)}\quad \sum_{i=1}^r \Big\langle \psi_i, \frac{1}{r_{12}}\psi_i\Big\rangle \leqslant \Big[\sum_{j=1}^r |E_j^{LB}| \Big(\sum_{k=1}^r |E_j^B|\Big)^{-1}\Big]^{1/2} \times \sum_{i=1}^r \Big\langle \psi_i^B, \frac{1}{r_{12}}\psi_i^B\Big\rangle \qquad (7)$$

where the E_j^{LB} denote lower bounds to the E_j, and E_k^B and ψ_k^B are determined by (6).

Proof: (i) For $\alpha\geqslant 0$ the first eigenvalue $E_1(\alpha)$ of

$$H(\alpha) := \sum_{i=1}^2\Big(-\frac{1}{2}\Delta_i - \frac{Z}{r_i}\Big) + \frac{\alpha}{r_{12}} \qquad (8)$$

is concave in α. Hence, for fixed Z, the α derivative $E_1'(\alpha) = \langle\psi_1^{(\alpha)},(1/r_{12})\psi_1^{(\alpha)}\rangle$ can be bounded [cf. (4.3.21) of Ref. 5] by

$$\langle\psi_1^{(\alpha)},(1/r_{12})\psi_1^{(\alpha)}\rangle \geqslant 2\big(-|E_1(\alpha)| + \sqrt{|E_1(\alpha)|\times|E_1(\alpha_0)|}\big)/(\alpha-\alpha_0) \qquad (9)$$

with $\alpha_0<\alpha$. Taking $\alpha_0=0$, $\alpha=1$, replacing E_1 by upper and lower bounds, and observing that $E_1(0) = E_1^B$, relation (5) results.

(ii) Again, let $E_k(\alpha)$ be the eigenvalues and $\psi_k^{(\alpha)}$ be the eigenfunctions of (8) in the considered symmetry sector. Then $f_r(\alpha) := -\sqrt{-\Sigma_{k=1}^r E_k(\alpha)}$ is concave in $\alpha\geqslant 0$ [cf. (4.3.19) of Ref. 5] and therefore $f_r''(\alpha)\leqslant 0$. Hence

$$f_r'(\alpha) = \frac{1}{2}\sum_{k=1}^r\Big\langle\psi_k^{(\alpha)}, \frac{1}{r_{12}}\psi_k^{(\alpha)}\Big\rangle\Big(-\sum_{j=1}^r E_j(\alpha)\Big)^{-1/2}$$

is a monotone decreasing function of α. Since $\psi_k^{(0)} = \psi_k^B$ and $\psi_k^{(1)} = \psi_k$, the inequality (7) is a consequence of $f_r'(1) \leqslant f_r'(0)$ and $|E_j^{LB}| \geqslant |E_j|$ for $j = 1,...,k$.

Remark 4: If we choose the trivial lower bounds $E_i^{LB} = E_i^B$, the inequality (7) says that the sum of the first r matrix elements $\langle 1/r_{12} \rangle$ taken within the exact eigenstates is bounded from above by the corresponding sum of the matrix elements calculated within hydrogenic-type functions. We would have also arrived at this weaker relation if in the proof of (7), instead of the concavity of $-\sqrt{-\Sigma_{k=1}^r E_k(\alpha)}$, we would have only used the concavity of $\Sigma_{k=1}^r E_k(\alpha)$. Although in particular for small Z the square root factor on the right-hand side of (7) improves the bound (depending of course on the quality of the lower bounds E_i^{LB}), we suspect that in this Z region the estimate (7) is nevertheless still far from being really sharp.

After these more general preparations, we now start with the investigation of specific transitions. First we will treat the transition between the ground and lowest excited state in the $L = 0$ sector of parahelium, i.e., in the conventional notation we are fixing the subspace belonging to the ^1S symmetry.

Lower bounds to the E_i can be derived with the help of the so-called "intermediate" techniques.[6] For our purpose, the Bazley–Fox procedure[7] with one- or two-dimensional projections onto the base problem given by (6) will work. In the ^1S-symmetry subspace the base levels are determined by

$$E_k^B(Z) = -(Z/2)(1 + 1/k^2), \tag{10}$$

where $k = 1,2,...$, and the ψ_k^B consist of the symmetric tensor products $S(\phi_1 \otimes \phi_{k00})$ times the antisymmetric spin part (here the ϕ_{nlk} denote the usual hydrogenic functions with quantum numbers n,l,m).

For nontrivial lower bounds on E_2 we need at least a two-dimensional projector. This yields a lower bound (cf. Refs. 5 and 8),

$$E_2^{LB}(Z) = \min\{E_+(Z), E_3^B(Z)\}, \tag{11}$$

where E_+ is the second eigenvalue of the intermediate Hamiltonian and is given by the expression

$$E_\mp(Z) = (E_1^B(Z) + E_2^B(Z))/2 + (M_{11} + M_{22})Z/2 \mp [(E_1^B(Z) - E_2^B(Z) + M_{11}Z - M_{22}Z)^2/4 + M_{12}^2 Z^2]^{1/2}. \tag{12}$$

The matrix elements M_{ij} can be calculated exactly (cf. the Appendix and the references there); in fractional form they are given by

$$M_{11} = 22405152737744/47537460297079,$$
$$M_{12} = 2313309487104/47537460297079, \tag{13}$$
$$M_{22} = 7942800465810/47537460297079.$$

There exists a unique $Z_{\text{cross}} > 0$ where the curves of E_3^B and E_+ cross, $E_3^B(Z_{\text{cross}}) = E_+(Z_{\text{cross}})$. For $Z \geqslant Z_{\text{cross}}$ we have $E_+(Z) \leqslant E_3^B(Z)$, so E_+ serves as a lower bound on E_2; for $Z \leqslant Z_{\text{cross}}$ the base level E_3^B has to be taken as a lower bound. This Z_{cross}, too, can be calculated exactly, but because the explicit expression involves too many digits we do not display it here. The estimate

$$1821/740 < Z_{\text{cross}} = 9M_{11}/8 + 36M_{22}/5 + (81/5)((5M_{11}/72 + 4M_{22}/9)^2 + 10(M_{12}^2 - M_{11}M_{22})/81)^{1/2} < 1319/536$$

(cf. Ref. 8) will be sufficient for our application.

Now we are ready to prove the following theorem.

Theorem 5: For all $Z \geqslant Z_0 = 674/449 \approx 1.50$ the excitation energy $\Delta E(Z|^1S)$ is strictly monotonic increasing.

Proof: First we consider those Z with $Z \geqslant Z_{\text{cross}}$ as defined above. Then, using (4) and Lemma 3, the left-hand side of (3) is estimated by

$$Z\frac{d\Delta E(Z)}{dZ} \geqslant 2(E_2^{LB}(Z) - E_1^{UB}(Z)) + 4(\sqrt{|E_1^{UB}(Z)| \times |E_1^B(Z)|} + E_1^{LB}(Z)) - \sum_{k=1}^{2}\left\langle \psi_k^B, \frac{1}{r_{12}}\psi_k^B \right\rangle. \tag{14}$$

(Because we are now working in the fixed symmetry subspace ^1S, in order to smooth out the notation we suppress the marking of the respective quantities with the symmetry symbols.) As upper bound E_1^{UB} we employ the standard variational expression

$$E_1^{UB}(Z) = -(Z - \tfrac{5}{16})^2. \tag{15}$$

In this Z region, it is sufficient to take for the lower bounds the trivial $E_2^{LB}(Z) = E_2^B(Z)$ and the linear one

$$E_1^{LB}(Z) = -Z^2 + \tfrac{16}{33}Z, \tag{16}$$

Eq. (16) being valid for $Z \geqslant \frac{128}{105}$ [see (4.3.22) of Ref. 5].

The matrix elements are determined as

$$\sum_{k=1}^{2}\left\langle \psi_k^B, \frac{1}{r_{12}}\psi_k^B \right\rangle = \left(\frac{5}{8} + \frac{17}{81} + \frac{16}{729}\right)Z = \frac{4997}{5832}Z. \tag{17}$$

Using these relations in (14), it remains to verify

$$\frac{3}{4}Z^2 - \frac{311947}{204120}Z + \frac{25}{128} \geqslant 0. \tag{18}$$

Now, since the coefficient of Z in (18) is bounded below by $-\frac{124}{81}$, for the validity of (18) it is sufficient to show the nonnegativity of $P(Z) := \frac{3}{4}Z^2 - \frac{124}{81}Z + \frac{25}{128}$. But, for large Z, $P(Z)$ is obviously positive and its greatest root can be calculated and estimated by $Z_+ \leqslant 500/243 \approx 2.06$. Because $Z_+ < Z_{\text{cross}}$, the positivity of $P(Z)$ implies the positivity in relation (3) for all $Z \geqslant Z_{\text{cross}}$.

Next we deal with the case $Z \leqslant Z_{\text{cross}}$. Here, we have to use $E_2^{LB}(Z) = E_2^B(Z)$ as a lower bound on $E_2(Z)$, but instead of the linear bound (16) for $E_1(Z)$ we choose its parabolic version

$$E_1^{LB}(Z) = -(Z - \tfrac{128}{105}(1 - \sqrt{\tfrac{5}{8}}))^2. \tag{19}$$

Actually, for the considered Z domain, (19) is sharper than the corresponding bound $E_-(Z)$ derived from a two-dimensional projector [and as given in (12)]. Since inserting trivial energy bounds into (7) would not be sufficient for our purpose, but on the other hand we want to keep the expressions analytically tractable, we estimate the nominator of the square root on the right-hand side of (7) as

$$\begin{aligned}\sqrt{|E_1^{LB}(Z)|+|E_2^{LB}(Z)|}&=\left[\frac{14}{9}\left(Z-\frac{9}{14}\gamma\right)^2+\frac{5}{14}\gamma^2\right]^{1/2}\\&\leqslant\left(\frac{14}{9}\right)^{1/2}\left(Z-\frac{9}{14}\gamma\right)+\left(\frac{5}{14}\right)^{1/2}\gamma,\end{aligned}\tag{20}$$

where we have set $\gamma := (128/105)\times(1-\sqrt{5/8})$. With (19), (20), and (4), proceeding analogously as in the derivation of (14), we arrive at the estimate

$$Z\frac{d\Delta E(Z)}{dZ}\geqslant\frac{4}{9}Z^2+BZ+C,\tag{21}$$

where B and C are given by

$$\begin{aligned}B&=\frac{1523}{420}-\frac{256}{105}\left(\frac{5}{2}\right)^{1/2}-\frac{4997}{8748}\left(\frac{7}{13}\right)^{1/2},\\C&=\frac{327678}{11025}\left(\frac{5}{2}\right)^{1/2}\\&\quad+\frac{79952}{59535}\left(\frac{1}{26}\right)^{1/2}+\frac{39976}{15309}\left(\frac{1}{182}\right)^{1/2}\\&\quad-\frac{13355863}{2822400}-\frac{19988}{59535}\left(\frac{5}{13}\right)^{1/2}\\&\quad-\frac{79952}{76545}\left(\frac{5}{91}\right)^{1/2}.\end{aligned}\tag{22}$$

Because we have no interest in a proliferation of digits, we bound the terms in B and C by fractions in which the denominator is restricted to at most three digits: e.g., $\frac{256}{105}\sqrt{\frac{5}{2}}\leqslant\frac{3801}{986}$ and $\frac{4997}{8748}\leqslant\frac{70}{167}$. In this way we obtain

$$\frac{4}{9}Z^2+BZ+C\geqslant\frac{4}{9}Z^2-\frac{646}{997}Z-\frac{3}{104}=:P(Z).\tag{23}$$

Clearly $P(Z)>0$ for large enough Z and its greatest root Z_+ is estimated by

$$Z_+\leqslant\frac{9}{8}\left(\frac{646}{997}+\frac{650}{947}\right)<\frac{674}{449}\approx 1.50,$$

concluding the proof of (3).

Remark 6: (i) The proof of Theorem 5 assumes the existence of $E_2(Z)\in\sigma_{\rm dis}(H(Z)|{}^1S)$ for $Z\geqslant Z_0$. Even if, for some Z, this were not true, Proposition 1 would guarantee that the assertion of Theorem 5 would still be correct. The same argument will also apply to the next theorems that follow.

(ii) Inequality (20) seems not to be very sharp. With the given energy bounds, but without falling back to this estimate, the relation (3) can be verified numerically for all $Z\geqslant 1.41$.

The other symmetry conserving transition that we want to examine involves the ground and first excited state of the $L=0$-sector of orthohelium. Thus we now fix the subspace with ${}^3S$ symmetry. Then we have the following theorem.

Theorem 7: The excitation energy $\Delta E(Z|{}^3S)$ is strictly monotonic increasing for all $Z\geqslant Z_0=\frac{2834352}{1232465}\approx 2.30$.

Proof: We apply the same strategy as in the proof of Theorem 5. In Ref. 8 it is shown that the crossing between the third base level E_3^B and the second eigenvalue E_+ of the intermediate Hamiltonian constructed with a two-dimensional Bazley–Fox projector is bounded by $\frac{3249}{992}\leqslant Z_{\rm cross}\leqslant\frac{1559}{476}$. Here the nth base level E_n^B is given by the $(n+1)$th level of the parasituation, i.e., $E_n^B(Z)=-\frac{1}{2}Z^2(1+(n+1)^{-2})$.

Again, we divide the proof into two parts. First we treat the case $Z\geqslant Z_{\rm cross}$. As before, in this Z region it is sufficient for us to use the trivial lower bound for E_2, namely $E_2^{LB}(Z)=E_2^B(Z)$. Now E_1 is estimated by

$$E_1^{LB}(Z)=-\tfrac{5}{8}(Z-Z_0(1-\tfrac{2}{3}\sqrt{2}))^2,\tag{24}$$

which is the parabolic version of the lower bound obtained from a one-dimensional projector (cf. Refs. 5 and 8), valid for $Z\geqslant Z_0$.

As an upper bound on E_1 we take the variational one

$$E_1^{UB}(Z)=-\tfrac{5}{8}(Z-\tfrac{548}{3645})^2\tag{25}$$

and the inequality (7) becomes here

$$\sum_{k=1}^{2}\left\langle\psi_k,\frac{1}{r_{12}}\psi_k\right\rangle\leqslant\left(\frac{17}{81}-\frac{16}{729}+\frac{815}{8192}-\frac{189}{32768}\right)Z.\tag{26}$$

With (24)–(26) the relation corresponding to (14) reads

$$\begin{aligned}Z\frac{d\Delta E(Z)}{dZ}&\\&\geqslant 2\left\{-\frac{5}{9}Z^2+\frac{5}{8}\left(Z-\frac{548}{3645}\right)^2\right\}\\&\quad-4\left\{\left[\frac{5}{8}\left(Z-\frac{548}{3645}\right)^2\times\frac{5}{8}Z^2\right]^{1/2}\right.\\&\quad\left.-\frac{5}{8}\left(Z-Z_0\left(1-\frac{2}{3}\sqrt{2}\right)\right)^2\right\}-\frac{6727975}{23887872}Z\\&\geqslant\frac{5}{36}Z^2-\frac{33434}{88981}Z-\frac{4133}{275616}=:P(Z).\end{aligned}\tag{27}$$

For (27) the same type of estimates as those in the derivation of (21)–(23) have been employed. The greatest root Z_+ of $P(Z)$ can be calculated exactly; again, because of the many digits involved, instead of giving it explicitly we only note that $Z_+\leqslant\frac{2720}{991}<Z_{\rm cross}$. Thus the expressions in (27) are definitely positive for all $Z\geqslant Z_{\rm cross}$.

For $Z\leqslant Z_{\rm cross}$, a lower bound to E_2 is given by $E_2^{LB}(Z)=E_3^B(Z)$. If again (24)–(26) are used, the only change in (27) concerns the coefficient Z^2 that is now $\frac{3}{16}$, and the greatest root of the corresponding lower bound polynomial can be estimated by $Z_+\leqslant\frac{1943}{951}$.

Remark 8: (i) Note that the smallest Z value for which we have proved the monotony of $\Delta E(Z|{}^3S)$ is not determined by the quality of the bounds but by the validity of the lower bound expression (24). Numerical experience with lower bounds from a two-dimensional projector shows that the monotony could be verified until $Z_0\approx 2.02$. Therefore since the physically interesting $Z=2$ value cannot be reached in this way, and on the other hand the exact lower bound expression for E_- contains numbers with more than 70 digits, we abstain here from trying to improve Z_0 via higher-dimensional lower bound projectors.

(ii) Transitions between the ground and first excited states in the $L=1$ sectors ${}^1P$ and ${}^3P$ can be handled analogously. But, as it turns out, if one uses expressions for the upper bounds E_1^{UB} that are similar to those of (15) and (25), then the resulting estimates are not sharp enough to prove positivity in (27) for interesting Z values. With upper and lower energy bounds derived with the help of two-dimensional projectors, the positivity in (3) can be verified for $Z \geqslant Z_0 \approx 2.3$ for parasystems and for $Z \geqslant Z_0 \approx 1.8$ for orthosystems. Although an analytical treatment analogous to the proofs of the preceding theorems would be possible, certainly this would become rather complex. So instead of entering into this, it seems to be more rewarding to find improvements or alternative bounds on the matrix elements $\langle r_{12}^{-1} \rangle$.

III. SYMMETRY NONCONSERVING TRANSITIONS

Physically more important than transition-within-symmetry subspaces are those in which the symmetry is changed. Selection rules favor, in particular, transitions with $\Delta L = \pm 1$. In our next theorem transitions between ground states of various symmetry sectors are investigated. If $E_1(Z\,|^{2S+1}L)$ denotes the ground state in the ${}^{2S+1}L$-symmetry subspace, the considered excitation energies are of the form

$$\Delta E(Z\,|^{2S_a+1}L_a \to {}^{2S_b+1}L_b) = E_1(Z\,|^{2S_a+1}L_a) - E_1(Z\,|^{2S_b+1}L_b).$$

Unfortunately, our available bounds are not sharp enough to derive interesting results for the ${}^3P \to {}^3S$ and ${}^1P \to {}^3S$ transitions, but we can show the following theorem.

Theorem 9: The excitation energies $\Delta E(Z\,|^1P \to {}^1S)$, $\Delta E(Z\,|^3P \to {}^1S)$, and $\Delta E(Z\,|^3S \to {}^1S)$ are strictly monotonic increasing for all $Z \geqslant Z_0 = \frac{128}{105}$.

Proof: Obviously $\Delta E(Z\,|^{2S_a+1}L_a \to {}^1S) \geqslant 0$, hence (4) implies

$$\frac{d\Delta\, E(Z\,|^{2S_a+1}L_a \to {}^1S)}{dZ} \geqslant \frac{1}{Z}\left\{\left\langle \psi_1, \frac{1}{r_{12}}\psi_1\right\rangle - \left\langle \psi_a, \frac{1}{r_{12}}\psi_a\right\rangle\right\}, \tag{28}$$

where ψ_a is the ground state in the ${}^{2S_a+1}L_a$ symmetry subspace. Now $\langle \psi_1, (1/r_{12})\psi_1\rangle$ is estimated by part (i) of Lemma 3 and $\langle \psi_a, (1/r_{12})\psi_a\rangle$ by part (ii) with $r = 1$. Choosing E_1^{UB} and E_1^{LB} as in (15) and (16) we get

$$\langle \psi_1, (1/r_{12})\psi_1\rangle \geqslant \tfrac{81}{280}Z$$

for all $Z \geqslant \frac{128}{105}$. The bound $\langle \psi_a, (1/r_{12})\psi_a\rangle \geqslant \langle \psi_a^B, (1/r_{12})\psi_a^B\rangle$ depends of course on the respective transition. In the ${}^1P$ sector we have $\psi_a^B = S(\phi_{100} \otimes \phi_{210}) \otimes$ *antisymmetric spin part,* while for ${}^3P$-symmetry $\psi_a^B = A(\phi_{100} \otimes \phi_{210}) \otimes$ *symmetric spin part* and for ${}^3S$-symmetry $\psi_a^B = A(\phi_{100} \otimes \phi_{200}) \otimes$ *symmetric spin part.* Consequently, the right-hand side of (26) is bounded below by

$$\frac{81}{280} - \left\{\begin{matrix} \frac{1705}{6561}, & \text{for } {}^1P \to {}^1S, \\ \frac{1481}{6561}, & \text{for } {}^3P \to {}^1S, \\ \frac{137}{729}, & \text{for } {}^3S \to {}^1S, \end{matrix}\right\} > 0,$$

which is the desired result.

The last transition that will concern us here is between the first excited state in the ${}^3S$ sector and the ground state in the ${}^1S$ sector. From Theorems 7 and 9 we can already infer the monotony of the corresponding transition energy

$$E_2(Z\,|^3S) - E_1(Z\,|^1S) \tag{29}$$

for all $Z \geqslant Z_0$ with $Z_0 \approx 2.30$ given in Theorem 7. But we can do better by examining this transition directly.

Theorem 10: The excitation energy (29) is strictly monotonic increasing for all $Z \geqslant \frac{128}{105}$.

Proof: Let ψ_{21} and ψ_{22} denote the ground and the first excited state, respectively, in the ${}^3S$ symmetry subspace. Then, by Lemma 2 the right-hand side of the relation (28) is bounded below by

$$\frac{1}{Z}\left\{\left\langle \psi_1, \frac{1}{r_{12}}\psi_1\right\rangle + \left\langle \psi_{21}, \frac{1}{r_{12}}\psi_{21}\right\rangle - \sum_{k=1}^{2}\left\langle \psi_{2k}, \frac{1}{r_{12}}\psi_{2k}\right\rangle\right\}$$
$$\geqslant \frac{1}{Z}\left\{\left\langle \psi_1, \frac{1}{r_{12}}\psi_1\right\rangle - \sum_{k=1}^{2}\left\langle \psi_{2k}^B, \frac{1}{r_{12}}, \psi_{2k}^B\right\rangle\right\}$$
$$\geqslant \frac{81}{280} - \left(\frac{137}{729} + \frac{3071}{32768}\right) > 0.$$

Here we have estimated $\langle \psi_1, (1/r_{12})\psi_1\rangle$ as in the preceding proof of Theorem 9, and the nonspin parts of ψ_{2k}^B are given by $A(\phi_{100} \otimes \phi_{200})$ for $k = 1$ and by $A(\phi_{100} \otimes \phi_{300})$ for $k = 2$.

IV. CONCLUDING REMARKS

In a rescaled $x \to Zx$ version, the Hamiltonian (1) becomes $H(Z) = Z^2\widetilde{H}(Z)$, where

$$\widetilde{H}(Z) = \sum_{i=1}^{N}\left(-\frac{1}{2}\Delta_i - \frac{1}{r_i}\right) + \frac{1}{Z}\sum_{i<j}\frac{1}{r_{ij}}.$$

The sign of the Z derivatives of the excitation energies $\Delta\widetilde{E}$ associated with $\widetilde{H}$ is determined by $\langle \psi_1, (1/r_{12})\psi_1\rangle - \langle \psi_2, (1/r_{12})\psi_2\rangle$. As in the original case, because the electrons tend to be on the average further away from each other in more excited states, physical intuition expects a positive difference. Unfortunately, because the positive contribution from the energy difference is missing, with our methods the positivity of this difference cannot be shown for the symmetry conserving transitions considered in Sec. II. This seems to be an effect of the poor quality of the estimates based on the inequality (7). As already mentioned before, we believe that the derivation of better or alternative bounds for matrix elements of $1/r_{12}$ would be an essential step towards further results, not only in the context considered here.

Our theorems of Secs. II and III and their physical background motivate the following general conjecture: *Let $E_1(Z)$, $E_2(Z)$ be eigenvalues of the Hamiltonian* (1) [*or $E_2(Z)$ an ionization threshold, if it is not an eigenvalue*] *such that $E_2(Z) > E_1(Z)$ for all $Z \geqslant Z_0$ with a certain $Z_0 > 0$. Then for all $Z \geqslant Z_0$ the excitation energy $E_2(Z) - E_1(Z)$ is monotone increasing in Z.*

At the moment we have no precise idea for a proof of this conjecture in the general situation. Since for atoms with more than two electrons no nontrivial analytical energy bounds are available, a generalization of the method employed in this paper seems not very promising. Let us stress again that further progress depends crucially on the development of sharper bounds on $\langle 1/r_{12}\rangle$.

APPENDIX: CALCULATION OF THE LOWER BOUNDS E^{LB}

For the sake of completeness let us briefly outline here the methods that we have employed for the calculation of the lower bounds E^{LB} and, in particular, for the numerical computation of the required matrix elements $\langle r_{12}\rangle$; further details can be found in Refs. 8 and 9, and in the references cited there.

The Bazley–Fox procedure[7] rests on the fact that $H^{LB} := H^B + r_{12}^{-1/2} P r_{12}^{-1/2}$ is a lower bound operator for H, i.e., $H^{LB} \leqslant H$, and that consequently its (ordered) eigenvalues E_k^{LB} are lower bounds to the corresponding eigenvalues E_k of H. Here the "base operator" H^B is given on the left-hand side of Eq. (6) and the projector

$$P = \sum_{i,j} r_{12}^{1/2} |\psi_i^B\rangle M_{ij} \langle \psi_j^B| r_{12}^{1/2}$$

contains the matrix elements M_{ij} of the matrix $M = W^{-1}$. The matrix W itself is composed of the inverse interaction matrix elements $W_{ij} = \langle \psi_i^B, r_{12} \psi_j^B\rangle$. For a one- or a two-dimensional projector P the diagonalization of H^B is either trivial or given by Eq. (12), respectively, and it remains to calculate W and its inverse W^{-1}.

As already mentioned in Sec. II, the eigenfunctions ψ_k^B of the base operator H^B are of hydrogenic type; explicit expressions for them are included in every textbook on quantum mechanics. Apart from the explicit knowledge of the ψ_k^B, the relation

$$r_{12} = (r_1^2 + r_2^2) \sum_{l=0}^{\infty} \frac{r_<^l}{r_>^{l+1}} P_l(\cos\Omega) - 2 r_1 r_2 \times \sum_{l=0}^{\infty} \left(\frac{l}{2l-1} \frac{r_<^{l-1}}{r_>^l} + \frac{l+1}{2l+3} \frac{r_<^{l+1}}{r_>^{l+2}} \right) P_l(\cos\Omega) \tag{A1}$$

is crucial for the possibility to calculate the W_{ij} exactly. In (A1), $r_< = \min(r_1, r_2)$, $r_> = \max(r_1, r_2)$, and Ω denotes the angle between the vectors r_1 and r_2. This relation follows from the so-called "Laplace expansion" of $1/r_{12}$ and the recursion properties of the Legendre polynomials P_l [cf. Eq. (18) in Ref. 9]. Although (A1) has the form of an infinite series, the angular integration in $\langle r_{12}\rangle$ renders almost all terms into zero so that the matrix elements W_{ij} are actually reduced to a sum over a relatively small number of terms (depending of course on the magnitude of the given quantum numbers for the ψ_k^B). The remaining radial integration can also be performed in closed form; as it turns out, the final result involves only rational expressions and, eventually, square roots of rationals.

While along these lines for the smallest quantum numbers the W_{ij} can still be computed with the help of a pocket calculator, the computations are tedious and become rather soon impractical if base functions contain a nontrivial angular part. Therefore we have developed[9] a package of REDUCE procedures that for any given set of quantum numbers calculate automatically the wanted matrix elements. Since the symbolic language REDUCE allows integer arithmetics with arbitrary precision, the results are exact and free of any round-off errors. On the other side, the final expressions for the W_{ij} and, in particular, those of the matrix elements of the inverse W^{-1} (which can also be computed exactly by REDUCE) contain more and more digits as the quantum numbers are increased. Thus for our purpose it was necessary to estimate appropriately most of the terms containing some M_{ij} and entering into the lower bounds E^{LB}. To this end we have chosen the criterion that only denominators with at most three digits are allowed. Hence the resulting expressions are still relatively simple whereas nevertheless these additional estimates do not induce any significant loss of accuracy in comparison with the sharpness of the original bounds. Moreover, since only rational manipulations are involved, no uncontrolled approximations are entering into our results and full mathematical rigor is maintained.

ACKNOWLEDGMENTS

It is a great pleasure to thank the members of the Centre de Physique Théorique of the CNRS Marseille for their kind hospitality. In particular, discussions with P. Duclos and Ph. Briet have been very stimulating for this work.

[1]E. H. Lieb, Phys. Rev. A **29**, 3018 (1984).
[2]B. Simon, "Fifteen problems in mathematical physics," in *Perspectives in Mathamatics*, Anniversary of Oberwalfach 1984 (Birkhäuser, Basel, 1984), p. 423; Ph. Briet, P. Duclos, and H. Hogreve, Lett. Math. Phys. **13**, 137 (1987).
[3]H. Narnhofer and W. Thirring, Acta. Phys. Austriaca **41**, 281 (1975).
[4]St. Bashkin, and J. O. Stoner, *Atomic Energy Levels and Grotian Diagrams I–IV* (North-Holland, Amsterdam, 1975); R. L. Kelly and L. J. Palumbo, "Atomic and ionic emission lines below 2000 Ångstroms, hydrogen through krypton," NRL report 7599 (Naval Research Laboratory, Washington, DC, 1973).
[5]W. Thirring, *A Course in Mathematical Physics*: *Vol.* 3, *Quantum Mechanics of Atoms and Molecules* (Springer, New York, 1981).
[6]A. Weinstein and W. Stenger, "Methods of intermediate problems for eigenvalues," *Mathematics in Science and Engineering* (Academic, New York, 1972), Vol. 89.
[7]N. W. Bazley, and D. W. Fox, Phys. Rev. **129**, 483 (1961); J. Math Phys. **3**, 469 (1962).
[8]H. Hogreve, "Rigorous bounds on eigenvalues of two-electron atomic Hamiltonians," J. Phys. A (to be published).
[9]N. Bogdanova and H. Hogreve, Comput. Phys. Commun. **48**, 319 (1988).