# A unified theory of point groups. VI. The projective corepresentations of the magnetic point groups of infinite order

Shoon K. Kim
*Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19006*

(Received 6 December 1983; accepted for publication 10 February 1984)

This paper provides all the $p$-inequivalent projective irreducible unitary corepresentations of all the magnetic point groups of infinite order with full use of their isomorphisms.

PACS numbers: 02.20. + b

In a series of papers[1–5] (referred to as I–V) we have developed a theory of representation of point groups in a unified manner regarding the corresponding double point groups as subgroups of the SU(2) group ( $= G_s$). In particular, in the last of these, we have constructed the general expressions of the projective irreducible unitary corepresentations (counirreps) of the magnetic (or antiunitary or Shubnikov) point groups of finite order. The present work is its extension to the magnetic point groups of infinite order denoted as $H^z_\infty$. Here $H_\infty$ is the halving subgroup which is a double point group of infinite order and $z$ is a unitary operator which defines the augmenting antiunitary operator $a = \theta z$ together with the time inversion operator $\theta$. By definition, $H^z_\infty$ is a mixed continuous group and thus construction of its representation group $H^{z\prime}_\infty$ requires algebraic manipulations which are different from those used in V. However, we still have the advantage that the representation group of a double point group is much simpler in structure than that of the corresponding single point group,[2] since the parameter space of the SU(2) group is simply connected while that of the SO(3) group is doubly connected. In fact, all proper double point groups continuous or otherwise have only one class of factor systems.[2]

In the present paper we shall first discuss the method of constructing the representation groups $H^{z\prime}_\infty$ of the magnetic point groups $H^z_\infty$ through a typical example of a grey group using the approach which is used for constructing the representation group of an ordinary continuous group whose parameter space is simply connected.[6] Then, the representation groups $H^{z\prime}_\infty$ will be constructed for a characteristic set of a total of eight $H^z_\infty$; any one of the remaining $H^z_\infty$ is isomorphic to one of them. Then, the vector counirreps of $H^{z\prime}_\infty$ will provide all the $p$-inequivalent projective counirreps of the characteristic set of $H^z_\infty$.

We shall now discuss how to construct the representation group of a typical example of the grey group $H^e_\infty$. It is assumed that the halving subgroup $H_\infty = \{x\}$ is a continuous symmetry group whose parameter space is simply connected. The antiunitary operator $a$ is the time inversion operator itself; $e$ being the identity operator. The grey group $H^e_\infty$ may be characterized by $H_\infty$ and the defining relations for $a(=\theta)$ as follows,

$$H^e_\infty: x \in H_\infty, \quad ax = xa, \quad a^2 = \bar{e}, \quad \bar{e}^2 = e, \tag{1}$$

where $\bar{e}$ is the $2\pi$ rotation. In constructing $H^{e\prime}_\infty$ we shall limit the discussion for the finite-dimensional representations.

Let $D$ be a $n$-dimensional general projective corepresentation of $H^e_\infty$. Then

$$D(x)D(y) = \exp[i\beta(x,y)]D(xy), \tag{2}$$

$$D(x)D(a) = \exp[i\xi(x)]D(a)D(x)^*, \tag{3}$$

$$D(a)D(a)^* = \tau D(\bar{e}), \tag{4}$$

TABLE I. The representation groups of the antiunitary and unitary point groups (of infinite order).[a]

1. $C'_\infty (= C_\infty)$: $x$,
2. $C^{e\prime}_\infty$: $x \in C_\infty$, $xa = ax$, $a^2 = \tau\bar{e}$, $\tau^2 = e$,
3. $C^{u\prime}_\infty$: $x \in C_\infty$, $xax = a$, $a^2 = \tau\bar{e}$, $\tau^2 = e$,
4. $C'_{\infty i}$: $x \in C_\infty$, $x\hat{i} = \hat{i}x$, $\hat{i}^2 = e$,
5. $C^{e\prime}_{\infty i}$: $x, \hat{i} \in C_{\infty i}$, $xa = ax$, $\hat{i}a = \zeta a\hat{i}$, $a^2 = \tau\bar{e}$, $\zeta^2 = \tau^2 = e$,
6. $C^{u\prime}_{\infty i}$: $x, \hat{i} \in C_{\infty i}$, $xax = a$, $\hat{i}a = \zeta a\hat{i}$, $a^2 = \tau\bar{e}$, $\zeta^2 = \tau^2 = e$,
7. $D'_\infty (= D_\infty)$: $x \in C_\infty$, $y^2 = (xy)^2 = \bar{e}$,
8. $D^{e\prime}_\infty$: $x, y \in D_\infty$, $xa = ax$, $ya = \eta ay$, $a^2 = \tau\bar{e}$, $\eta^2 = \tau^2 = e$,
9. $D'_{\infty i}$: $x, y \in D_\infty$, $x\hat{i} = \hat{i}x$, $y\hat{i} = \gamma\hat{i}y$, $\hat{i}^2 = e$, $\gamma^2 = e$,
10. $D^{e\prime}_{\infty i}$: $D'_{\infty i}(\gamma)$, $xa = ax$, $ya = \eta ay$, $\hat{i}a = \zeta a\hat{i}$, $a^2 = \tau\bar{e}$, $\eta^2 = \zeta^2 = \tau^2 = e$,
11. $G'_s (= G_s)$: $x$,
12. $G^{e\prime}_s$: $x \in G_s$, $xa = ax$, $a^2 = \tau\bar{e}$,
13. $G'_{si}$: $x \in G_s$, $x\hat{i} = \hat{i}x$, $\hat{i}^2 = e$,
14. $G^{e\prime}_{si}$: $x, \hat{i} \in G'_{si}$, $xa = ax$, $\hat{i}a = \zeta a\hat{i}$, $a^2 = \tau\bar{e}$, $\zeta^2 = \tau^2 = e$.

[a] Note: For the notations, see Table I of Paper V.

 0022-2488/84/072125-03$02.50  

for all $x$ and $y \in H_\infty$. Here $*$ denotes the complex conjugate, $\beta(x, y)$ and $\xi(x)$ are real continuous functions of the elements over the entire parameter space of the group and are called the local exponents.[6] For uniqueness we take the standard factor system such that $D(e) = 1$ and fix the local exponents uniquely by

$$\beta(x,e) = \beta(e, y) = \xi(e) = 0, \quad \forall x, y \in H_\infty . \tag{5}$$

It is a simple matter now to map off the local exponents completely by a gauge transformation. Let the determinant of the unitary matrix $D(x)$ be

$$\det D(x) = \exp[i\delta(x)], \quad \forall x \in H_\infty , \tag{6}$$

where $\delta(x)$ is a real continuous function of $x$ and $\delta(e) = 0$. Taking the determinants of both sides of both equations (2) and (3) we obtain

TABLE II. The projective counirreps (unirreps) of the antiunitary (unitary) point groups (of infinite order).[a]

1. $C_\infty(K^0)$: $K^0$, $M_m$: $m = m^0$, $m^0 = 0, \pm\frac{1}{2}, \pm 1, \ldots, \pm\infty$,
2. $C^e_\infty(K)$
   $K$, $S(M_0)$, $S(M_m, M_{-m})$, $m = m^* = \frac{1}{2}, 1, \ldots, \infty$,
3. $C^u_\infty(K)$
   $K$, $S(M_m)$, $m = m^0$,
4. $C_{\infty i}(K^0)$
   $K^0$, $M^\pm_m$, $m = m^0$,
5. $C^e_{\infty i}(K_t, t = \{\zeta\})$
   $K_1$, $S(M^\pm_0)$, $S(M^\pm_m, M^\pm_{-m})$, $m = m^*$,
   $K_2$, $S(M^+_m, M^-_{-m})$, $m = m^0$,
6. $C^u_{\infty i}(K_t, t = \{\zeta\})$
   $K_1$, $S(M^\pm_m)$, $m = m^0$,
   $K_2$, $S(M^+_m, M^-_m)$, $m = m^0$,
7. $D_\infty(K^0)$: $K^0$, $A_1, A_2, E_m$, $m = m^*$,
8. $D^e_\infty(K_t, t = \{\eta\})$:
   $K_1$, $S(A_1)$, $S(A_2)$, $S(E_m; 1_2, \sigma_y)$, $m = m^*$,
   $K_2$, $S(A_1, A_2)$, $S(E_m, E_m; \sigma_y, 1_2)$, $m = m^*$,
9. $D_{\infty i}(K^0_s, s = \{\gamma\})$:
   $K^0_1$, $A^\pm_1$, $A^\pm_2$, $E^\pm_m$, $m = m^*$,
   $K^0_2$, $D_A = D(A_1, A_2)$, $D^{\pm y}_m = D(E_m; \pm\sigma_y)$, $m = m^*$,
10. $D^e_{\infty i}(K_{st}, s = \{\gamma\}, t = \{\eta, \zeta\})$:
   $K_{11}$, $S(A^\pm_1)$, $S(A^\pm_2)$, $S(E^\pm_m; 1_2, \sigma_y)$, $m = m^*$,
   $K_{12}$, $S(A^+_1, A^-_1)$, $S(A^+_2, A^-_2)$, $S(E^+_m, E^-_m; 1_2, \sigma_y)$, $m = m^*$,
   $K_{13}$, $S(A^\pm_1, A^\pm_2)$, $S(E^\pm_m, E^\pm_m; \sigma_y, 1_2)$, $m = m^*$,
   $K_{14}$, $S(A^\pm_1, A^\mp_2)$, $S(E^+_m, E^-_m; \sigma_y, 1_2)$, $m = m^*$,
   $K_{21}$, $S(D_A; 1_2)$, $S(D^{+y}_m, D^{-y}_m; 1_2, \sigma_y)$, $m = m^*$,
   $K_{22}$, $S(D_A; \sigma_z)$, $S(D^{\pm y}_m; 1_2, \sigma_y)$, $m = m^*$,
   $K_{23}$, $S(D_A; \sigma_x)$, $S(D^{+y}_m, D^{-y}_m; \sigma_y, 1_2)$, $m = m^*$,
   $K_{24}$, $S(D_A, D_A; \sigma_y)$, $S(D^{\pm y}_m, D^{\pm y}_m; \sigma_y, 1_2)$, $m = m^*$,
11. $G_s(K^0)$
   $K^0$, $D^{(j)}$, $j = 0, \frac{1}{2}, 1, \ldots, \infty$,
12. $G^e_s(K)$:
   $K$, $S(D^{(j)}; N^{(j)})$, $N^{(j)}_{nm} = (-1)^{j-m}\delta(n, -m)$, $n, m = j, j-1, \ldots, -j$,
13. $G_{si}(K)$:
   $K$, $D^{(j)\pm}$, $j = 0, \frac{1}{2}, 1, \ldots, \infty$,
14. $G^e_{si}(K_t, t = \{\zeta\})$:
   $K_1$, $S(D^{(j)\pm}, N^{(j)})$,
   $K_2$, $S(D^{(j)+}, D^{(j)-}; N^{(j)})$.

[a] Notes: (1) For the notations see Table II of Paper V. (2) $m^0$, $m^*$ are integers of half-integers defined by $m^0 = 0, \pm\frac{1}{2}, \pm 1, \ldots, \pm\infty$, $m^* = \frac{1}{2}, 1, \frac{3}{2}, \ldots, \infty$. (3) For $D^{(j)}$ of (11) and (12) see Eq. (3.10) of Ref. 4.

$$\delta(x) + \delta(y) = n\beta(x,y) + \delta(xy),$$
$$2\delta(x) = n\xi(x). \tag{7}$$

Then, the gauge transformation

$$D'(x) = \exp[-i\delta(x)/n] D(x) \tag{8}$$

leads to the required result

$$D'(x)D(y)' = D'(xy), \quad D'(x)D(a) = D(a)D'(x)^*. \tag{9}$$

To determine the phase factor $\tau$ in (4), we take the equivalent transformations of both sides of (4) with respect to $D(a)$. Then we have

$$\tau^2 = 1. \tag{10}$$

Now we regard $\tau$ as a second-order element which commutes with all the elements of $H^e_\infty$ and arrive at the representation group $H^{e\prime}_\infty$ which may be defined by

$$x \in H_\infty, \quad xa = ax, \quad a^2 = \tau\bar{e}, \quad \tau^2 = e, \tag{11}$$

where $\tau$ is in the center of $H^{e\prime}_\infty$. In an analogous manner one can construct all the representation groups $H^{z\prime}_\infty$ given in Table I.

There exists a total of 14 magnetic point groups $H^z_\infty$ of infinite order. On account of their isomorphisms, however, it is only necessary to construct the representation groups of a characteristic set of the magnetic point groups which may be chosen to be

$$C^e_\infty, \quad C^u_\infty, \quad C^e_{\infty i}, \quad C^u_{\infty i}, \quad D^e_\infty, \quad D^e_{\infty i}, \quad G^e_s, \quad G^e_{si} \tag{12}$$

(for the notations see IV). Any one of the remaining $H^z_\infty$ is isomorphic to one of these as follows[5]:

$$C^i_\infty \simeq C^e_\infty, \quad C^v_\infty \simeq C^u_\infty,$$
$$C^i_{\infty v} \simeq C^e_{\infty v} \simeq D^i_\infty \simeq D^e_\infty, \quad G^i_s \simeq G^e_s \tag{13}$$

through the one-to-one correspondence $\theta\hat{i} \leftrightarrow \theta$ and $\bar{c}_2' \leftrightarrow c_2'$. The representation groups $H^{z\prime}_\infty$ of the above set (12) are given in Table I together with the representation of groups $H'_\infty$ of their halving subgroups $H_\infty$ for convenience of presentation. Then, from their vector counirreps we have obtained the general expressions of all $p$-inequivalent projective counirreps of the corresponding magnetic point groups in terms of the unirreps of the proper point groups given in the previous work I. These are presented in Table II together with the projective unirreps of the halving unitary groups $H_\infty$. Thus Table II provides all the projective counirreps (unirreps) of any antiunitary (unitary) group of infinite order directly or through isomorphisms. It is noted that these results given in Table II can be obtained by the limiting procedure from those of $H^z$ of finite order. It is also noted that in general a class of the factor systems $K$ and its dual $K'$ are always $p$-inequivalent without exception for $H^z_\infty$. This is not in general true for the magnetic groups of finite order.[5]


[1] S. K. Kim, J. Math. Phys. **22**, 2101 (1981).
[2] S. K. Kim, J. Math. Phys. **24**, 411 (1983).
[3] S. K. Kim, J. Math. Phys. **24**, 414 (1983).
[4] S. K. Kim, J. Math. Phys. **24**, 419 (1983).
[5] S. K. Kim, J. Math. Phys. **25**, 189 (1984).
[6] M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison-Wesley, Reading, MA, 1964), p. 469.

# Signatures of finite SU( $p,q$ ) representations[a)]

J. Patera
*Centre de recherches de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada*
R. T. Sharp
*Physics Department, McGill University, Montreal, Quebec, H3A 2T8, Canada*

(Received 12 May 1983; accepted for publication 6 January 1984)

The signature $S$ of a finite-dimensional representation of SU( $p,q$ ) is the difference between the number of positive and negative signs in the bilinear invariant in its diagonal form. An expression for $S$ is derived starting from the Weyl character formula for U( $p,q$ ) representations.

PACS numbers: 02.20. + b

## I. INTRODUCTION

Noncompact simple Lie groups/algebras are today well established among mathematical tools of theoretical physics. Originally they found their way into physics through the special relativity theory and since then their interest to physicists had its ups and downs, but it is beyond doubt that, in general, their applications have been growing in variety and frequency. Although most of the representations which are being used are either the lowest-dimensional defining representations or, on the contrary, the infinite-dimensional unitary ones, it appears to be only a matter of time until nontrivial information about other finite-dimensional (nonunitary) representations will be needed. One of the very first questions to be answered about many of them is what is the signature, i.e., the number of positive and negative signs in the bilinear invariant in its diagonal form. Equivalently, one may ask what is the maximal number of linearly independent "spacelike, timelike, or lightlike" vectors in that representation space. It turns out that the answer is nowhere to be found except for the lowest cases which are obvious and Ref. 1, which deals with representations of SU( $p,q$ ), $p+q\leqslant 4$.

The purpose of this paper is to provide the answer for SU( $p,q$ ) with any value of $p+q$, and to set up a general method which can be applied to representations of other groups.

The method of Ref. 1 makes use of known generating functions and therefore cannot easily be extended to higher $p$ and $q$. Here we evaluate Weyl's U( $p,q$ ) character formula for the element of the U( $p,q$ ) group whose character is the signature. The present approach could be used to derive character formulas for other elements of SU( $p+q$ ) of finite order.

The signature $S_\lambda$ of an irreducible representation $\lambda$ of SU( $p,q$ ) of dimension $N_\lambda$ is the difference between the number $p_\lambda$ of positive signs and the number $q_\lambda$ of negative signs in the bilinear invariant ( $x,y$ ) taken in diagonal form, i.e.,

$$(x,y)=x^+M_\lambda\, y,\quad M_\lambda=I_p\oplus(-I_q),$$

where $I_n$ is the $n\times n$ identity matrix. Thus $S_\lambda=\operatorname{Tr} M_\lambda$. For the defining representation $\lambda=(1,0,...,0)$, $p_\lambda=p$, and $q_\lambda=q$ so that $S=\operatorname{Tr} M=p-q$. The matrix $M$ is an element of U( $p,q$ ) and also of U( $p+q$ ). The signature $S_\lambda$ is the character of the element $M_\lambda$ in the representation $\lambda$ of U( $p,q$ ); we thereby fix the phase of $S_\lambda$. Therefore our task here is to evaluate the character of the element $M_\lambda$ for the representation $\lambda$. For that purpose we use Weyl's character formula for an element of the group U( $p,q$ ) which is a diagonal ( $p+q$ )×( $p+q$ ) matrix with the variables fixed such that $p$ of its elements are $+1$ and $q$ are $-1$. Any finite representation $\lambda$ of U( $p,q$ ) contains a unique representation $\lambda$ of SU( $p,q$ ) and the character of $M_\lambda$ is the signature $S_\lambda$.

The signature $S_\lambda$ of a representation $\lambda$ of dimension $N_\lambda$ of the group U( $p,q$ ) has some obvious properties. Let $p_\lambda$ and $q_\lambda$ denote the number of positive and negative signs in the bilinear invariant of $\lambda$. Then we have

$$p_\lambda=\tfrac{1}{2}(N_\lambda+S_\lambda),\quad q_\lambda=\tfrac{1}{2}(N_\lambda-S_\lambda). \tag{1}$$

For the direct sum and product $\lambda_1\oplus\lambda_2$, $\lambda_1\otimes\lambda_2$ we have

$$\begin{aligned}&p_{\lambda_1\oplus\lambda_2}=p_{\lambda_1}+p_{\lambda_2},\quad p_{\lambda_1\otimes\lambda_2}=p_{\lambda_1}p_{\lambda_2}+q_{\lambda_1}q_{\lambda_2},\\&q_{\lambda_+\oplus\lambda_2}=q_{\lambda_1}+q_{\lambda_2},\quad q_{\lambda_1\otimes\lambda_2}=p_{\lambda_1}q_{\lambda_2}+q_{\lambda_1}p_{\lambda_2},\\&S_{\lambda_1\oplus\lambda_2}=S_{\lambda_1}+S_{\lambda_2},\quad S_{\lambda_1\otimes\lambda_2}=S_{\lambda_1}S_{\lambda_2}.\end{aligned}\tag{2}$$

An irreducible representation $\lambda$ of SU( $p+q$ ) is ordinarily labeled by the $p+q-1$ nonnegative integers

$$\lambda_n=2(\lambda,\alpha_n)/(\alpha_n,\alpha_n),\quad n=1,2,...,p+q-1, \tag{3}$$

where $\lambda$ denotes the highest weight of the representation and $\alpha_i$ are the simple roots of U( $p,q$ ). For our purpose it is convenient to use an equivalent set of $p+q$ integers,

$$\begin{aligned}&l_j=\sum_{k=j}^{p+q-1}\lambda_k+p+q-j,\quad j=1,2,...,p+q-1;\\&l_{p+q}=0.\end{aligned}\tag{4}$$

In Sec. II the general formula for $S_\lambda$ is presented. It turns out to be a product of two expressions. The first contains only trivial factors while the second is a sum of products of two determinants depending separately on the even- and odd-valued labels $l_i$ of the representation and otherwise only on the difference $p-q$. The cases $p-q<5$ are worked out in detail. We assume that $p\geqslant q$. If $q>p$, interchange $p$ and $q$ in the formula for $S_\lambda$ and multiply by $(-1)^{\Sigma_n n\lambda_n}$.

The signature formula is derived in Sec. III.

## II. THE SIGNATURE FORMULA

Consider an irreducible representation of U( $p,q$ ) labeled by the integers $l_n$. In Eq.(4) they are in decreasing order $l_1>l_2>\cdots>l_{p+q}=0$. Let $s$ be the number of odd $l_n$ and $t$ the number of even $l_n$. Then $s+t=p+q$. It is convenient to number the $l_n$ so the odd ones are $l_1^o>...>l_s^o$ and the even ones are $l_1^e>...>l_t^e$.

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

  

The chief result of this paper is that the signature $S_\lambda$ is zero whenever $s > p$ (and $t < q$) or $t > p$ (and $s < q$) and otherwise is given by

$$S_\lambda = \frac{\epsilon_\lambda(-1)^{s(p-1)+(1/2)q(q-1)}(\Pi_{1<n<m\leq s}(l_n^o - l_m^o))(\Pi_{1<n<m<t}(l_n^e - l_m^e))F_{p-q,s-q}(l^o,l^e)}{2^{q(p-q)}(\Pi_{n=1}^{p-1} n!)(\Pi_{m=1}^{q-1} m!)}. \tag{5}$$

An intuitive reason for the vanishing of $S_\lambda$ with the above inequalities is that when one of them is satisfied all choices of the $s$ (or $t$) determinant vanish when a Laplace expansion is made in the first $s$ and last $t$ columns; $\epsilon_\lambda$ in (5) is $\pm 1$ according to whether the permutation from $l_1,...,l_{p+q}$ to $l_1^o,...,l_s^o,l_1^e,...,l_t^e$ is even or odd. $F_{p-q,s-q}(l)$ is given by the formula

$$F_{p-q,s-q}(l^o,l^e) = \sum_{(\beta_i)} (-1)^{\Sigma_k\beta_k - (1/2)(s-q)(s+q-1)} \times |p_{\beta i} - s + j(l^o)|\ |p_{\gamma i} - t + j(l^e)|. \tag{6}$$

Here $(\beta_i)$ stands for $s-q$ integers $\beta_1 > ... > \beta_{s-q}$ chosen from the set $p-1, p-2,...,q$ and $(\gamma_i)$ stands for the remaining $t-q$ integers from the set, also numbered in decreasing order. The sum $\Sigma_{(\beta_i)}$ is over the $(p-q)!/(s-q)!(t-q)!$ choices of the integers $(\beta_i)$. The factor $|p_{\beta_i - s + j}(l^o)|$ is the $(s-q)\times(s-q)$ determinant whose $ij$ element is exhibited; $|p_{\gamma_i - t + j}(l^e)|$ is a similar $(t-q)\times(t-q)$ determinant. The $n$th degree symmetric function $p_n(l_1,...,l_s)$ is defined by

$$\prod_{i=1}^{s}(l - zl_i)^{-1} = \sum_{n=0}^{\infty} p_n(l)z^l, \quad n \geqslant 0,$$

$$p_n(l) = 0, \quad n < 0. \tag{7}$$

For the following trivial special cases the function (6) is unity:

$$F_{p-q,0}(l) = F_{p-q,p-q}(l) = 1. \tag{8}$$

The form of the function (6) depends only on $p-q$ and $s-q$. Therefore we evaluate it explicitly for a few low values of $p-q$, namely, $p-q<5$. In order that $F \neq 0$ one must have $0 < s-q < p-q$. Furthermore a symmetry relation (25) below allows us to cut the range of $s-q$ values by half. When $p-q=0$ or 1, then $s-q=0$ or $p-q$ and $F$ is given by (8). Consequently the nontrivial cases we list below have $2 < p-q < 5$ and $0 < s-q < \frac{1}{2}(p-q)$. In order to simplify the notation we use $p_\beta^0$ and $p_\beta^e$ for $p_\beta(l^o)$ and $p_\beta(l^e)$, respectively; in the summations in (9)–(12) distinct dummies $i,j,...$ never take the same values when the variables $l_i,l_j,...$ are raised to different powers, and satisfy inequalities $i<j<...$ when the variables $l_i,l_j,...$ are raised to the same power. Thus $\Sigma\, l_i^2 l_j^3$ means $\Sigma_{i\neq j} l_i^2 l_j^3$ while $\Sigma\, l_i^2 l_j^2$ means $\Sigma_{i<j} l_i^j l_j^j$.

$$F_{2,1}(l) = p_1^e - p_1^o = \sum_i l_i^e - \sum_i l_i^o, \tag{9}$$

$$\begin{aligned} F_{3,1}(l) &= (p_1^e)^2 - p_2^e - p_1^e p_1^o + p_2^o \\ &= \sum_{i<j} l_i^e l_j^e - \Big(\sum_i l_i^e\Big)\Big(\sum_j l_j^o\Big) + \sum_i (l_i^o)^2 + \sum_{i<j} l_i^o l_j^o, \end{aligned} \tag{10}$$

$$\begin{aligned} F_{4,1}(l) &= (p_1^e)^3 + p_3^2 - 2p_1^e p_2^e - p_1^o((p_1^e)^2 - p_2^e) + p_2^o p_1^e - p_1^o \\ &= \sum_{i<j<k} l_i^e l_j^e l_k^e - \Big(\sum_i l_i^o\Big)\Big(\sum_{i<j} l_i^e l_j^e\Big) \\ &\quad + \Big(\sum_i (l_i^o)^2 + \sum_{i<j} l_i^o l_j^o\Big)\Big(\sum_i l_i^e\Big) - \sum_i (l_i^o)^3 \\ &\quad - \sum_{ij} (l_i^o)^2 l_j^o - \sum_{i<j<k} l_i^o l_j^o l_k^o, \end{aligned} \tag{11}$$

$$\begin{aligned} F_{4,2}(l) &= (p_2^e)^2 - p_1^e p_3^e - p_1^o(p_1^e p_2^e - p_3^e) + p_2^o((p_1^e)^2 - p_2^e) + ((p_1^o)^2 - p_2^o)p_2^e - (p_1^o p_2^o - p_3^o)p_1^e + (p_2^o)^2 - p_1^o p_3^o \\ &= \sum_{\substack{j\neq i\neq k \\ j<k}} (l_i^e)^2 l_j^e l_k^e + \sum_{i<j} (l_i^e)^2 (l_j^e)^2 + 2\sum_{i<j<k<r} l_i^e l_j^e l_k^e l_r^e - \Big(\sum_i l_i^o \sum_{i\neq j} (l_i^e)^2 l_j^e + 2\sum_{i<j<k} l_i^e l_j^e l_k^e\Big) \\ &\quad + \Big(\sum_i (l_i^o)^2 + \sum_{i<j} l_i^o l_j^o\Big)\sum_{i<j} l_i^e l_j^e + (l_i^e \leftrightarrow l_i^o). \end{aligned} \tag{12}$$

## III. DERIVATION OF THE SIGNATURE FORMULA

The signature of any finite irreducible representation $\lambda$ of U($p,q$) is given in terms of the character $\chi_\lambda(\eta)$ by

$$S_\lambda = \chi_\lambda(1,...,1,-1,...,-1), \tag{13}$$

where the first $p$ $\eta$'s have been set equal to 1, the last $q$ to $-1$. For the character $\chi_\lambda(\eta)$ Weyl gives the formulas

$$\chi_\lambda(\eta) = \xi_\lambda(\eta)/\xi_0(\eta) = |p_{l_i - p - q + j}(\eta)|, \tag{14}$$

where

$$\xi_\lambda(\eta) = |\eta_i^{\,l_j}|, \tag{15}$$

$$\xi_0(\eta) = |\eta_i^{\,l_j^o}| = |\eta_i^{p+q-j}| = \prod_{1<i<j\leq p+q} (\eta_i - \eta_j). \tag{16}$$

In Eqs. (14) – (16) $|A_{ij}|$ denotes the $(p+q)\times(p+q)$ determinant whose $ij$ element is $A_{ij}$. We see, by (13) and (14) that $|p_{l_i - p - q + j}(1,...,1,-1,...,-1)|$ is an explicit expression for the signature. However, the expression (2.1) is far simpler to evaluate. This section is devoted to its derivation.

We start with $\xi_\lambda(\eta)/\xi_0(\eta)$ and set

$$\eta_i = \begin{cases} e^{\zeta_i}, & 1 \leqslant i \leqslant p, \\ -e^{\zeta_i}, & p+1 \leqslant i \leqslant p+q. \end{cases} \tag{17}$$

Then, according to (13) and (14),

$$S_\lambda = \lim_{\zeta_i \to 0} \xi_\lambda / \xi_0. \tag{18}$$

Keeping only lowest degree terms when $\zeta_i \to 0$ we find, using (16) and (17),

$$\xi_0 \simeq (-1)^{(1/2)p(-1)} 2^{pq} \Big( \prod_{1 \leqslant i < j \leqslant p} (\zeta_j - \zeta_i) \Big) \times \Big( \prod_{p+1 \leqslant i < j \leqslant p+q} (\zeta_j - \zeta_i) \Big). \tag{19}$$

With the substitution (17) we get

$$\xi_\lambda = \epsilon_\lambda \left| \begin{array}{c|c} e^{l_j \zeta_i} & e^{l_j \zeta_i} \\ \hline -e^{l_j \zeta_i} & e^{l_j \zeta_i} \end{array} \right|, \tag{20}$$

the vertical line separates the first $s$ columns from the last $t$ columns while the horizontal line divides the first $p$ rows from the last $q$. Now repeat the following operation $p-1$ times, giving $i$ in succession the values $1,2,\dots,p-1$: subtract the $i$th row from each row $k$ for which $i+1 \leqslant k \leqslant p$ and bring outside a factor $(\zeta_k - \zeta_i)/i$ from the $k$ th row. Then repeat the following operation $q-1$ times, giving $i$ the values $p+1, p+2,\dots,p+q-1$: subtract the $i$th row from each row $k$ for which $i+1 \leqslant k \leqslant p+q$ and bring outside a factor $(\zeta_k - \zeta_i)/(i-p)$ from the $k$ th row. The result is

$$\xi_\lambda \simeq \epsilon_\lambda \Big( \prod_{1 \leqslant i < k \leqslant p} (\zeta_k - \zeta_i) \Big) \Big( \prod_{p+1 \leqslant i < k \leqslant p+q} (\zeta_k - \zeta_i) \Big) \times \Big( \prod_{i=1}^{p-1} i! \Big)^{-1} \Big( \prod_{i=1}^{q-1} i! \Big)^{-1} \left| \begin{array}{c|c} l_j^{i-1} & l_j^{i-1} \\ \hline -l_j^{i-p-1} & l_j^{i-p-1} \end{array} \right|, \tag{21}$$

where we have kept only the leading terms for small $\zeta_i$. Dividing $\xi_\lambda$, Eq. (21), by $\xi_0$, Eq. (19), we find

$$S_\lambda = \epsilon_\lambda (-1)^{(1/2)p(p-1)} 2^{-pq} \Big( \prod_{i=1}^{p-1} i! \Big)^{-1} \Big( \prod_{i=1}^{q-1} i! \Big)^{-1} \times \left| \begin{array}{c|c} l_j^{i-1} & l_j^{i-1} \\ \hline -l_j^{i-p-1} & l_j^{i-p-1} \end{array} \right|. \tag{22}$$

Now make a Laplace expansion of the determinant in (22) by its first $s$ columns. In the first $s$ columns the first $q$ rows are the negatives of the last $q$. Therefore in the Laplace expansion one must take one from each of the $q$ pairs in the $s$ determinant and the other in the $t$ determinant. It follows that $s$ and $t$ must lie between $p$ and $q$: $p \geqslant s \geqslant q$ and $p \geqslant t \geqslant q$; otherwise $S_\lambda$ vanishes. There are $2^q$ ways of choosing one of each of the $q$ pairs and it may be shown straightforwardly that each choice contributes equally. We therefore make a conventional choice, the first $q$ rows in the $s$-determinant, the last $q$ rows in the $t$-determinant, and multiply by $2^q$. We find

$$\left| \begin{array}{c|c} l_j^{i-1} & l_j^{i-1} \\ \hline -l_j^{i-p-1} & l_j^{i-p-1} \end{array} \right| = 2^q (-1)^{s(p-1) + (1/2)q(q-1) + (1/2)p(p-1)} \times \Big( \prod_{1 \leqslant i < j \leqslant s} (l_i - l_j) \Big) \Big( \prod_{s+1 \leqslant i < j \leqslant s+t} (l_i - l_j) \Big) \times \sum_{(\beta_i)} (-1)^{\Sigma \beta_i - (1/2)(s-q)(s+q-1)} |p_{\beta_i - s + j}(l^o)| \, |p_{\gamma_i - t + j}(l^e)|. \tag{23}$$

Inserting (23) into (22) yields the desired result, Eq. (4).

We conclude this section by noting the symmetry relation satisfied by $F_{a,b}(l^o, l^e)$, namely

$$F_{a,b}(l^o, l^e) = (-1)^{(a-1)b} F_{a,a-b}(l^e, l^o). \tag{24}$$

Equation (24) follows straightforwardly from the definition (5).

## IV. EXAMPLES AND REMARKS

First we evaluate the signature for the groups SU(1,1), SU(2,1), SU(2,2), and SU(3,1). Table I summarizes the results. As in Sec. II, the odd valued $l$'s are labeled $l_1^o, l_2^o, \dots l_s^o$ in decreasing order and the even $l$'s are $l_1^e, l_e^e, \dots, l_t^e$ in decreasing order. In the conventional labeling the $l$'s are $l_1, l_2, \dots, l_{p+q}$ in decreasing order. $\epsilon_\lambda$ is $\pm 1$ according to whether the permutation from $l_1, l_2, \dots, l_{p+q}$ to $l_1^o, l_2^o, \dots, l_s^o, l_1^e, l_2^e, \dots, l_t^e$ is even or odd.

SU(1,1). In this case $p = q = s = t = 1$ for $S_\lambda \neq 0$. Then,

TABLE I. The signatures $S_\lambda$ of irreducible representations $\lambda = (\lambda_1), (\lambda_1, \lambda_2)$, and $(\lambda_1, \lambda_2, \lambda_3)$ of, respectively SU(1,1), SU(2,1), and SU(3,1), SU(2,2). Symbol $e$ ($o$) in the column $\lambda_i$ denotes an even (odd) $\lambda_i$.

| Group | $\lambda_1$ | Parity of $\lambda_2$ | $\lambda_3$ | $S_\lambda$ |
|---|---|---|---|---|
| SU(1,1) | $e$ | - | - | 1 |
| | $o$ | - | - | 0 |
| SU(2,1) | $e$ | $e$ | - | $\frac{1}{2}(\lambda_1 + \lambda_2 + 2)$ |
| | $o$ | $o$ | - | 0 |
| | $o$ | $e$ | - | $\frac{1}{2}(\lambda_1 + 1)$ |
| | $e$ | $o$ | - | $\frac{1}{2}(\lambda_2 + 1)$ |
| SU(2,2) | $e$ | $e$ | $e$ | $\frac{1}{4}(\lambda_1 + \lambda_2 + 2)(\lambda_2 + \lambda_3 + 2)$ |
| | $e$ | $o$ | $e$ | $-\frac{1}{4}(\lambda_2 + 1)(\lambda_1 + \lambda_2 + \lambda_3 + 3)$ |
| | $o$ | $e$ | $o$ | $-\frac{1}{4}(\lambda_1 + 1)(\lambda_3 + 1)$ |
| | | otherwise | | 0 |
| SU(3,1) | $e$ | $e$ | $e$ | $\frac{1}{8}(\lambda_1 + \lambda_2 + 2)(\lambda_1 + \lambda_3 + 2)(\lambda_2 + \lambda_3 + 2)$ |
| | $e$ | $e$ | $o$ | $-\frac{1}{8}(\lambda_3 + 1)(\lambda_1 + \lambda_2 + 2)(\lambda_1 + \lambda_2 + \lambda_3 + 3)$ |
| | $o$ | $e$ | $e$ | $\frac{1}{8}(\lambda_1 + 1)(\lambda_2 + \lambda_3 + 2)(\lambda_1 + \lambda_2 + \lambda_3 + 3)$ |
| | $e$ | $o$ | $e$ | $\frac{1}{8}(\lambda_2 + 1)(\lambda_1 - \lambda_3)(\lambda_1 + \lambda_2 + \lambda_3 + 3)$ |
| | $o$ | $o$ | $e$ | $\frac{1}{8}(\lambda_1 + 1)(\lambda_2 + 1)(\lambda_1 + \lambda_2 + 2)$ |
| | $o$ | $e$ | $o$ | $-\frac{1}{8}(\lambda_1 + 1)(\lambda_3 + 1)(\lambda_1 + 2\lambda_2 + \lambda_3 + 4)$ |
| | $e$ | $o$ | $o$ | $\frac{1}{8}(\lambda_2 + 1)(\lambda_3 + 1)(\lambda_2 + \lambda_3 + 2)$ |
| | $o$ | $o$ | $o$ | 0 |

according to (5) and (8), $S_\lambda = F_{00} = 1$.

SU(2,1). In this case $p = 2, q = 1$, so we must have $s = 2, t = 1$ or $s = 1, t = 2$ for $S_\lambda \neq 0$. According to (5) one has

$$S_\lambda = \tfrac{1}{2}\epsilon_\lambda(-1)^s(l_n^* - l_m^*), \quad n < m, \tag{25}$$

where $l_n^* - l_m^*$ is the difference of the odd $l$'s ($* = o$) for $s = 2$, or of the even $l$'s ($* = e$) for $t = 2$.

SU(2,2). Here $p = q = 2$, so we must have $s = t = 2$ for nonzero $S_\lambda$. According to (5),

$$S_\lambda = -\epsilon_\lambda(l_1^o - l_2^o)(l_1^e - l_2^e). \tag{26}$$

SU(3,1). Here $p = 3, q = 1$, so there are two distinct cases corresponding to $S_\lambda \neq 0$. (i) $s = t = 2$, and (ii) $s = 3, t = 1$ or $s = 1, t = 3$. From (5) we have

$$\text{(i)} \quad S_\lambda = \epsilon_\lambda(l_1^o - l_2^o)(l_1^e - l_2^e)F_{2,1}(l),$$
$$\text{(ii)} \quad S_\lambda = \epsilon_\lambda(l_1^* - l_2^*)(l_2^* - l_3^*)(l_1^* - l_3^*), \tag{27}$$

where $* = o$ or $e$ according to whether $s = 3$ or $t = 3$, respectively. Substitution of the representation labels $\lambda_1,\ldots,\lambda_4$ of (3) and (4) into (25)–(27) gives the expressions for $S_\lambda$ summarized in Table I.

The problem solved in this paper could be viewed as a special case of the evaluation characters of elements of finite order of SU($n$). Indeed, if $q$ is even the U($p+q$) element $M$, whose characters we evaluate, is also an element of SU($p+q$). If $q$ is odd and $p$ even $M \notin$ SU($p+q$) because det $M = -1$; we consider an SU($p+q$) element $M' = -M$, SU($p+q$). Then

$$S_\lambda = \operatorname{tr} M_\lambda = (\operatorname{tr} M'_\lambda)(-1)^{\Sigma_i l_i + (1/2)n(n+1)} \tag{28}$$

in any irreducible representation $\lambda$. Without loss of generality we could have here redefined the bilinear invariant $(x,y) \rightarrow -(x,y)$ so that then $S_\lambda = \operatorname{tr} M'_\lambda$. If $p$ and $q$ are both odd, $M$ is in one-to-one correspondence with SU($p+q$) element $M'' = M \exp(2\pi i/(p+q))$. Then

$$S_\lambda = \operatorname{tr} M_\lambda = (\operatorname{tr} M''_\lambda)\exp\left(-2\pi i\left[\sum_i l_i + \frac{1}{2}n(n+1)\right]\right)/n, \quad n = p + q. \tag{29}$$

An identification of SU($n$) elements $M$ and $M''$ in a general standard notation for elements of finite order is found in Sec. 9.2 of Ref. 2.

It is possible to evaluate characters of other elements of finite order in SU($n$) by a generalization of the methods of this paper.

[1] J. Patera and R. T. Sharp, Kinam **4**, 93–98 (1982).
[2] R. V. Moody and J. Patera, to appear in SIAM J. Algebraic Discrete Methods **5** (2) (1984).

# The representation matrix elements of the group $O^+(2,2)$

Ansaruddin Syed
*Department of Mathematics, University of Karachi, Karachi, Pakistan*

(Received 19 April 1983; accepted for publication 9 December 1983)

Making use of the isomorphism of $O^+(2,2)$ with the direct product $O^+(2,1)\times O^+(2,1)$, the matrix elements of $O^+(2,2)$ in its unitary irreducible representations are explicitly calculated in terms of Euler angles introduced in a previous paper. The expressions so obtained consist of infinite sums of product of Clebsch–Gordan coefficients and Bargmann's $v$ functions, both for the group $O^+(2,1)$.

PACS numbers: 02.20. + b

## 1. INTRODUCTION

The problem of computation of unitary irreducible representation (UIR) matrix elements for the unimodular orthogonal and pseudo-orthogonal (generalized Lorentz) groups has a pretty long history, although quite a large amount of work on it was done in fairly recent past. It was originated in a recognizable form by Wigner in 1931 when he introduced[1] his, by now very well-known and extensively used, $D$ and $d$ functions, which are just the elements for the three-dimensional pure rotation group $O^+(3)$. Next, Bargmann[2] obtained them for the three-dimensional Lorentz group $O^+(2,1)$ as well as for the ordinary Lorentz group $O^+(3,1)$; these latter have also been calculated by several other authors.[3–7] The $d$ functions of $O^+(4)$ have been obtained by Friedman and Wang[8] and Biedenharn,[9] those of $O^+(5)$ by Holman[10] and those of $O^+(4,1)$ by Holman,[11] Ström,[12] and Takahashi.[13] For the general cases $O^+(n)$ and $O^+(n,1)$, the problem has been studied in considerable details by quite a large number of authors.[14–18] However, as far as the author knows, none of the cases $O^+(n,2)$, $n\geqslant 2$, has ever been considered in this connection. Hence, in order to make a beginning, we start with $O^+(2,2)$ and obtain its UIR matrix elements in the present paper. This group turns out to be exceptionally simple due, essentially, to the fact that it is isomorphic to the direct product $O^+(2,1)\times O^+(2,1)$; this enables one to use the trick of Friedman and Wang[8] [introduced in connection with the isomorphism $O^+(4)\approx O^+(3)\times O^+(3)$] and make the calculations almost trivial. One of the main reasons for the lack of interest in the matrix elements of $O^+(n,2)$, $n\geqslant 2$, in spite of the fact that a number of series of UIR's of $O^+(p,q)$, $p,q\geqslant 2$, have been known[19] for some time, has probably been the absence of a suitable set of parameters for these groups, similar to the set of Euler angles for $O^+(n)$ and $O^+(n,1)$. In a previous paper,[20] the author was able to define a set of Euler angles for the general case $O^+(p,q)$; these, and a second similar but slightly different set of angles, are now used to obtain explicit expressions for the matrix elements of $O^+(2,2)$.

## 2. THE GROUP $O^+$ (2,2) AND ITS UIR'S

The group $O^+(2,2)$ consists of all the $4\times 4$ real matrices $\alpha=\{\alpha_{\mu\nu}\}$ which satisfy

$$\alpha^T g\alpha = g,$$

$$\det\alpha = 1,$$

$g$ being the $4\times 4$ diagonal matrix

$$g=\operatorname{diag}(1,1,-1,-1).$$

The first condition ensures that these matrices keep the lengths of vectors

$$x=(x_1,x_2,x_3,x_4)$$

in the four-dimensional real Minkowski space $M(2,2)$, given by

$$x^2=x_1^2+x_2^2-x_3^2-x_4^2,$$

invariant, i.e., are orthogonal linear transformations in this space. It is a six-parameter group and the six generators

$$a_{12},\quad a_{34},\quad b_{13},\quad b_{14},\quad b_{23},\quad b_{24},$$

of the infinitesimal transformations in various $x_\mu - x_\nu$ planes, i.e., the generators of the Lie algebra of $O^+(2,2)$ are given by

$$(a_{12})_{\lambda\kappa} = -\delta_{12}\delta_{2\kappa}+\delta_{1\kappa}\delta_{2\lambda},$$

$$(a_{34})_{\lambda\kappa} = \delta_{12}\delta_{2\kappa}-\delta_{1\kappa}\delta_{2\lambda},$$

$$(b_{\mu\nu})_{\lambda\kappa} = \delta_{\mu\lambda}\delta_{\nu\kappa}+\delta_{\mu\kappa}\delta_{\nu\lambda},\quad \mu=1,2,\ \nu=3,4,$$

$\delta_{ij}$ being the usual Kronecker delta. Setting

$$h_1=ib_{23},\quad h_2=ib_{13},\quad h_3=ia_{12},$$

$$k_1=ib_{14},\quad k_2=ib_{24},\quad k_3=ia_{34},$$

it is easily checked that

$$[h_1,h_2]=-ih_3,\quad [h_2,h_3]=ih_1,\quad [h_3,h_1]=ih_2.$$

Note that $(a_{12},b_{13},b_{23})$, i.e., $(h_1,h_2,h_3)$, are just the generators of the Lie algebra of the subgroup $O^+(2,1)$ of $O^+(2,2)$ consisting of those of its elements which leave $x_4$ invariant.

Introducing now

$$j_i=\tfrac{1}{2}(h_i+k_i),\quad l_i=\tfrac{1}{2}(h_i-k_i),\quad i=1,2,3,$$

we find that

$$[j_1,j_2]=-ij_3,\quad [j_2,j_3]=ij_1,\quad [j_3,j_1]=ij_2,$$

$$[l_1,l_2]=-il_3,\quad [l_2,l_3]=il_1,\quad [l_3,l_1]=il_2,$$

$$[j_i,l_j]=0,\quad i,j=1,2,3.$$

Thus $\{h_i\}$ and $\{k_i\}$ combine together to give, and are themselves determined by, two independent sets $\{j_i\}$ and $\{l_i\}$ of generators of the Lie algebra of $O^+(2,1)$; this leads to the well-known fact that $O^+(2,2)$ is isomorphic with the direct product $O^+(2,1)\times O^+(2,1)$. If $-q_j(q_j+1)$ and $m_j$ are the eigenvalues of

$$\mathbf{j}^2=-(j_1^2+j_2^2-j_3^2)$$

 0022-2488/84/072132-04$02.50  

and $j_3$, we denote by $\mathscr{D}^{q_j}$ any of the UIR's of $O^+(2,1)$ (generated by $\{j_i\}$) labeled by $q_j$ according to the labeling scheme of Holman and Biedenharn.[21] The standard basis for the representation space of $\mathscr{D}^{q_j}$ consists of

$$\{|q_j,m_j\rangle\},$$

the collection of simultaneous eigenvectors of $\mathbf{J}^2$ and $J_3$:

$$\mathbf{J}^2|q_j,m_j\rangle = -q_j(q_j+1)|q_j,m_j\rangle,$$
$$J_3|q_j,m_j\rangle = m_j|q_j,m_j\rangle,$$

(capital letters denote the representatives, in the representation under consideration, of the operators denoted by the corresponding small letters). $q_l$, $m_l$, $q_h$, $m_h$, and $\mathscr{D}^{q_l}$ are similarly defined. The range of values of $m_j$, $m_l$ and $m_h = m_j + m_l$ (as $h_3 = j_3 + l_3$) depend on the particular representations $\mathscr{D}^{q_j}$ and $\mathscr{D}^{q_l}$ chosen; we shall carry out our calculations only for the case when both $\mathscr{D}^{q_j}$ and $\mathscr{D}^{q_l}$ belong to the integral variety of the principal series of continuous representations[21] (i.e., they are of the type $c_q^0$, $q > \frac{1}{4}$, in the notation of Bargmann[2]) as results for other choices can be obtained in a similar manner. The range of values of the three eigenvalues $m_j$, $m_l$, and $m_h$ will therefore be

$$0, \pm 1, \pm 2, \cdots .$$

As $O^+(2,2)$ is generated by the union $\{j_i\}\cup\{l_i\}$, its UIR's will be labeled by the pair $(q_j,q_l)$. We shall denote them by $\mathscr{D}^{q_j,q_l}$; these are, in fact, the direct product[22] of $\mathscr{D}^{q_j}$ and $\mathscr{D}^{q_l}$:

$$\mathscr{D}^{q_j,q_l} = \mathscr{D}^{q_j} \otimes \mathscr{D}^{q_l}.$$

Obviously, one basis for the representation space of $\mathscr{D}^{q_j,q_l}$ will be the set of vectors

$$|q_j,m_j;q_l,m_l\rangle \equiv |q_j,m_j\rangle|q_l,m_l\rangle,$$
$$m_j,m_l = 0, \pm 1, \pm 2, \cdots .$$

However, as

$$\mathbf{J}^2,\mathbf{L}^2,\mathbf{H}^2,H_3$$

also form a set of four mutually commuting independent Hermitian operators, another basis for it will consist of their simultaneous eigenvectors, i.e., the set of vectors

$$|q_j,q_l;q_h,m_h\rangle.$$

As $h_i = j_i + l_i$, the range of $q_h$ will consist of those values which label those UIR's of $O^+(2,1)$, which appear in the reduction of the product of $\mathscr{D}^{q_j}$ and $\mathscr{D}^{q_l}$. Looking at the analysis of this reduction given by Holman and Biedenharn,[21] we see that the possible values of $q_h$ are such that it labels either a continuous representation of principal series and integral variety or a discrete representation, again of integral variety. In the former case, the range of values of $m_h$ is

$$0, \pm 1, \pm 2, \cdots ,$$

while it is

$$-q_h, -q_h+1, -q_h+2, \cdots$$

if $q_h$ labels a positive discrete representation and

$$q_h, q_h-1, q_h-2, \cdots$$

if it labels a negative discrete representation.

As $h_i$ are the usual generators of the subgroup $O^+(2,1)$ of $O^+(2,2)$ which keeps $x_4$ invariant, we shall have, for $a \in O^+(2,1)$,

$$T(a)|q_j,q_l;q_h,m_h\rangle = \sum_{m_h'} v^{q_h}_{m_h m_h'}(a)|q_j,q_l;q_h,m_h'\rangle,$$

where $T(a)$ is the operator representing $a$ in $\mathscr{D}^{q_j,q_l}$, and the $v^q_{mn}(a)$ are Bargmann's[2] $v$ functions for $O^+(2,1)$. This leads to

$$\langle q_j,q_l;q_h,m_h|T(a)|q_j,q_l;q_h',m_h'\rangle$$

$$= \begin{cases} \delta(q_h - q_h')v^{q_h}_{m_h m_h'}(a) & \text{if (A) is satisfied} \\ 0 & \text{if (B) is satisfied} \\ \delta_{q_h q_h'} v^{q_h}_{m_h m_h'}(a) & \text{if (C) is satisfied,} \end{cases} \quad (1)$$

where (A), (B), and (C) are the following conditions:

(A) $q_h, q_h'$ both label continuous representations of principal series;

(B) one of $q_h, q_h'$ labels a continuous representation of principal series and the other a discrete representation;

(C) both $q_h$ and $q_h'$ label discrete representations.

This equation will be used in the next section.

## 3. THE MATRIX ELEMENT

Let

$$\alpha \equiv \{\alpha_{\mu\nu}\} \in O^+(2,2).$$

We shall calculate

$$v^{q_j,q_l}_{q_h,m_h;q_h',m_h'}(\alpha) = \langle q_j,q_l;q_h,m_h|T(\alpha)|q_j,q_l;q_h',m_h'\rangle,$$

the matrix elements of $\alpha$ in the representation $\mathscr{D}^{q_j,q_l}$. It turns out that the two cases

$$\alpha_{44} > 1, \quad \alpha_{44} < 1$$

have to be considered separately.

### Case I: $\alpha_{44} > 1$

Here the suitable Euler angles are the ones given by Syed.[20] These are

$$\phi_{44},\phi_{43},\theta_{42},\phi_{33},\theta_{32},\theta_{22},$$

with $\alpha$ given in terms of them by

$$\alpha = r_{12}(\theta_{42})l_{13}(\phi_{43})l_{14}(-\phi_{44})r_{12}(\theta_{32})l_{13}(-\phi_{33})r_{12}(-\theta_{22}),$$

where $r_{\mu\nu}(\theta)$ = a rotation by an angle $\theta$ in the $\mu$–$\nu$ plane and $l_{\mu\nu}(\phi)$ = a Lorentz transformation by an angle $\phi$ in the $\mu$–$\nu$ plane. We write

$$\alpha = bl_{14}(-\phi_{44})a$$

so that

$$a = r_{12}(\theta_{32})l_{13}(-\phi_{33})r_{12}(-\theta_{22}) \in O^+(2,1),$$
$$b = r_{12}(\theta_{42})l_{13}(\phi_{43}) \in O^+(2,1).$$

Then

$$v^{q_j q_l}_{q_h,m_h;q'_h,m'_h}(\alpha)$$
$$= \langle q_j,q_l;q_h,m_h|T(b)L_{14}(-\phi_{44})T(a)|q_j,q_l;q'_h,m'_h\rangle$$
(by Ref. 23)
$$= \int_{q''_h}\int_{q'''_h}\sum_{m''_h,m'''_h}\langle q_j,q_l;q_h,m_h|T(b)|q_j,q_l;q''_h,m''_h\rangle$$
$$\times\langle q_j,q_l;q''_h,m''_h|L_{14}(-\phi_{44})|q_j,q_l;q'''_h,m'''_h\rangle$$
$$\times\langle q_j,q_l;q'''_h,m'''_h|T(a)|q_j,q_l;q'_h,m'_h\rangle,$$
where $\int_q$ stands for summation over discrete range and integration over the continuous range of values of $q$. Using now (1), we get
$$v^{q_j q_l}_{q_h,m_h;q'_h,m'_h}(\alpha) = \sum_{m''_h,m'''_h} v^{q_j}_{m_h m''_h}(b)\, v^{q_l}_{m'''_h m'_h}(a) V^{q_j q_l}_{q_h,m''_h;q'_h,m'''_h}(-\phi_{44}), \qquad (2)$$
where
$$V^{q_j q_l}_{q_h,m''_h;q'_h,m'''_h}(\phi) = \langle q_j,q_l;q_h,m''_h|e^{iK_1\phi}|q_j,q_l;q'_h,m'''_h\rangle,$$
and, of course, $e^{-iK_1\phi} = L_{14}(\phi)$. To evaluate the last matrix element, we use the expansion of $|q_j,q_l;q_h,m_h\rangle$ in a series of
$$|q_j,m_j;q_l,m_l\rangle \equiv |q_j,m_j\rangle|q_l,m_l\rangle.$$
Taking this expansion in terms of Clebsch–Gordan coefficients of $O^+(2,1)$[24] as
$$|q_j,q_l;q_h,m_h\rangle = \sum_{m_j} C(q_j,q_l,q_h;m_j,m_h-m_j,m_h)$$
$$\times|q_j,m_j\rangle|q_l,m_h-m_j\rangle,$$
and using $K_1 = J_1 - L_1$, we get
$$V^{q_j q_l}_{q_h,m''_h,q'_h,m'''_h}(\phi)$$
$$= \sum_{m''_j,m'''_j} C^*(q_j,q_l,q_h;m''_j,m''_h-m''_j,m''_h)$$
$$\times C(q_j,q_l,q'_h;m'''_j,m'''_h-m'_j,m'''_h)$$
$$\times\langle q_j,m''_j|e^{-iJ_1\phi}|q_j,m'''_j\rangle$$
$$\times\langle q_j,m''_h-m''_j|e^{iL_1\phi}|q_l,m'''_h-m'''_j\rangle.$$
Now, by actually carrying out the matrix multiplications, it is easy to check that
$$r_{12}(\pi/2)l_{13}(-\phi)r_{12}(-\pi/2) = l_{23}(-\phi);$$
this leads to
$$e^{iJ_1\phi} = e^{i\pi J_3/2}e^{iJ_2\phi}e^{i\pi J_3/2}.$$
Hence
$$\langle q_j,m''_j|e^{iJ_1\phi}|q_j,m'''_j\rangle$$
$$= \langle q_j,m''_j|e^{i\pi J_3/2}e^{iJ_2\phi}e^{i\pi J_3}|q_j,m'''_j\rangle$$
$$= e^{i\pi(m'''_j-m''_j)/2}\langle q_j,m''_j|e^{iJ_2\phi}|q_j,m'''_j\rangle$$
$$= (i)^{m'''_j-m''_j}V^{q_j}_{m''_j m'''_j}(-\phi/2),$$
where
$$V^q_{mn}(y) = \langle q,m|-e^{iJ_2\cdot 2y}|q,n\rangle$$
is Bargmann's $V$ function.[2] Similarly,
$$\langle q_l,m''_h-m''_j|e^{iL_1\phi}|q_l,m'''_h-m'''_j\rangle$$
$$= (i)^{m'''_h-m''_h-(m'''_j-m''_j)}V^{q_l}_{m''_h-m''_j m'''_h-m'''_j}(\phi/2),$$
so that
$$V^{q_j q_l}_{q_h,m''_h;q'_h,m'''_h}(\phi) = \sum_{m''_j,m'''_j} C^*(q_j,q_l,q_h;m''_j,m''_h-m''_j,m''_h)$$
$$\times C(q_j,q_l,q'_h;m'''_j,m'''_h-m'''_j,m'''_h)$$
$$\times(i)^{m'''_h-m''_h}V^{q_j}_{m''_j m'''_j}(-\phi/2)$$
$$\times V^{q_l}_{m''_h-m''_j m'''_h-m'''_j}(\phi/2). \qquad (3)$$
(2) and (3) completely determine the matrix element of $\alpha$ in $\mathscr{D}^{q_j q_l}$.

**Case II: $\alpha_{44} < 1$**

We need, for this case, a slightly different set of Euler angles which are obtained by a variation in the definition of "polar angles in $\mathbb{C}^n$" given by Syed.[20] We define the "new" polar angles in $C^4$ by
$$\begin{aligned}
z_4 &= t\cos\chi_4, & 0\leqslant\theta_4\leqslant\pi,\\
z_3 &= t\sin\chi_4\cos x_3, & 0\leqslant\theta_3\leqslant\pi,\\
z_2 &= t\sin\chi_4\sin x_3\cos x_2, & 0\leqslant\theta_2\leqslant 2\pi,\\
z_1 &= t\sin\chi_4\sin x_3\sin x_2, &\\
t &= t_1 + it_2,\quad t_1\geqslant 0, &\\
\chi_m &= \theta_m + i\phi_m,\quad 1\leqslant m\leqslant 4. &
\end{aligned}$$
These give
$$t = \pm(z_1^2+z_2^2+z_3^2+z_4^2)^{1/2},$$
$$\cos\chi_4 = z_4/t,$$
$$\cos\chi_3 = \pm z_3/(z_1^2+z_2^2+z_3^2)^{1/2},$$
$$\cos\chi_2 = \pm z_2/(z_1^2+z_2^2)^{1/2},$$
$$\sin\chi_2 = \pm z_1/(z_1^2+z_2^2)^{1/2},$$
Let now $\alpha\in o^+(2,2)$ with $\alpha_{44}\leqslant 1$, and set
$$\hat\alpha = f\alpha f^{-1},$$
where $f$ is the $4\times 4$ diagonal matrix
$$f = \text{diag}(1,1,i,i).$$
Thus
$$\hat\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & i\alpha_{13} & -i\alpha_{14}\\ \alpha_{21} & \alpha_{22} & -i\alpha_{23} & -i\alpha_{24}\\ i\alpha_{31} & i\alpha_{32} & \alpha_{33} & \alpha_{34}\\ i\alpha_{41} & i\alpha_{42} & \alpha_{43} & \alpha_{44}\end{pmatrix}.$$
Let
$$-\chi_{44}, -\chi_{43}, -\chi_{42},$$
be the new polar angles of the fourth column
$$[-i\alpha_{14}, -i\alpha_{24}, \alpha_{34}, \alpha_{44}]^T$$
of $\hat\alpha$. Then it is easy to check (using $\alpha_{44}\leqslant 1$) that
$$\chi_{44} = \theta_{44},\quad \chi_{43} = i\phi_{43},\quad \chi_{42} = \theta_{42},$$
and that
$$\hat\alpha^3 = r^T_{34}(\theta_{44})r^T_{23}(i\phi_{43})r^T_{12}(\theta_{42})\alpha$$

has the last row and column as those of the $4\times 4$ unit matrix. Hence, if $\hat{\alpha}^{(3)}$ is the matrix obtained from $\hat{\alpha}^3$ by deleting its last row and column, we will have

$$\hat{\alpha}^{(3)}\in O^+(2,1).$$

We now take

$$\chi_{33},\chi_{32},\chi_{22}$$

as the old Euler angles of $\hat{\alpha}^{(3)}$ defined by Syed,[20] who shows that

$$\chi_{33} = i\phi_{33}, \quad \chi_{32} = \theta_{32}, \quad \chi_{22} = \theta_{22}.$$

The collection

$$\{\theta_{44},\phi_{43},\theta_{42},\phi_{33},\theta_{32},\theta_{22}\}$$

of six angles is now taken as the set of new Euler angles of $\alpha$.

Now, from Syed,[20]

$$\hat{\alpha}^3 = r_{12}(\theta_{32})r_{13}(-i\phi_{33})r_{12}(-\theta_{22}),$$

so that

$$\hat{\alpha} = r_{12}(\theta_{42})r_{23}(i\phi_{43})r_{34}(\theta_{44})r_{12}(\theta_{32}) \times r_{13}(-i\phi_{33})r_{12}(-\theta_{22})$$

$$\Rightarrow \alpha = f^{-1}\hat{\alpha} f = r_{12}(\theta_{42})l_{23}(\phi_{43})r_{34}(\theta_{44})r_{12}(\theta_{32}) \times l_{13}(-\phi_{33})r_{12}(-\theta_{22})$$

i.e.,

$$\alpha = br_{34}(\theta_{44})a,$$

where

$$b = r_{12}(\theta_{42})l_{23}(\phi_{43})\in O^+(2,1),$$
$$a = r_{12}(\theta_{32})l_{13}(-\phi_{33})r_{12}(-\theta_{22})\in O^+(2,1).$$

Thus

$$v^{q_j q_l}_{q_h,m_h;q_h',m_h'}(\alpha) = \langle q_j,q_l;q_h,m_h|T(b)R_{34}(\theta_{44})T(a)|q_j,q_l;q_h',m_h'\rangle$$
$$= \sum_{m_h''}\sum_{m_h'''} v^{q_h}_{m_h,m_h''}(b)v^{q_h'}_{m_h''',m_h'}(a) \times \langle q_j,q_l;q_h,m_h''|e^{-iK_3\theta_{44}}|q_j,q_l;q_h',m_h'''\rangle$$
$$= \sum_{m_h''}\sum_{m_h'''} v^{q_h}_{m_h,m_h''}(b)v^{q_h'}_{m_h''',m_h'}(a)V^{q_j q_l}_{q_h,q_h';m_h'',m_h'''}(\theta_{44}),$$

where

$$V^{q_j q_l}_{q_h,q_h';m_h'',m_h'''}(\theta) = \langle q_j,q_l;q_h,m_h''|e^{iK_3\theta}|q_j,q_l;q_h',m_h'''\rangle$$
$$= \sum_{m_j''}\sum_{m_j'''} C^*(q_j,q_l,q_h;m_j'',m_h''-m_j'',m_h'') \times C(q_j,q_l,q_h';m_j''',m_h'''-m_j''',m_h''') \times \langle q_j,m_j''|e^{iJ_3\theta}|q_j,m_j'''\rangle \times \langle q_l,m_h''-m_j''|e^{iL_3\theta}|q_l,m_h'''-m_j'''\rangle$$
$$= \sum_{m_j''} C^*(q_j,q_l,q_h;m_j'',m_h''-m_j'',m_h'') \times C(q_j,q_l,q_h';m_j'',m_h''-m_j'',m_h'')e^{i(m_h''-2m_j'')}\delta_{m_h''m_h'''}.$$

Thus we finally get the matrix element of $\alpha$ in $\mathscr{D}^{q_j q_l}$ as

$$v^{q_j q_l}_{q_h,m_h;q_h',m_h'}(\alpha) = \sum_{m_h''} v^{q_h}_{m_h m_h''}(b)v^{q_h'}_{m_h'' m_h'}(a)V^{q_j q_l}_{q_h,q_h';m_h''}(\theta_{44})$$

with

$$V^{q_j q_l}_{q_h,q_h';m_h''}(\theta) = \sum_{m_j''} C^*(q_j,q_l,q_h;m_j'',m_h''-m_j'',m_h'') \times C(q_j,q_l,q_h';m_j'',m_h''-m_j'',m_h'')e^{i(m_h''-2m_j'')\theta}.$$

Note the close similarity of this last expression with the expression for the "boost" matrix of $O^+(4)$ given by Friedman and Wang.[8]


[1] E. P. Wigner, *Gruppentheorie* (Vieweg, Braunschweig, 1931), pp. 180,233.
[2] V. Bargmann, Ann. Math. **48**, 568 (1947).
[3] S. Ström, Ark. Fys. **34**, 215 (1967).
[4] Dao Vong Due and Nguyen Van Hieu, Dokl. Acad. Nauk. SSSR **137**, 1281 (1967); Acad. Inst. H. Poincaré **6**, 17 (1967).
[5] G. I. Kuznetzov, M. A. Liberman, A. A. Markov, and Ya. A. Smorodinski, Sov. J. Nucl. Phys. **10**, 370 (1970).
[6] K. Toth, Trieste preprint I. C. 69, 1969, p. 128.
[7] A. Markov and G. I. Shepelev, Sov. J. Nucl. Phys. **12**, 596 (1971).
[8] D. Z. Friedman and J. M. Wang, Phys. Rev. **160**, 1560 (1967).
[9] L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).
[10] J. W. Holman, J. Math. Phys. **10**, 1710 (1969).
[11] J. W. Holman, J. Math. Phys. **10**, 888 (1969).
[12] S. Ström, Ark. Fys. **30**, 455 (1965); A. Kihlberg and S. Ström, Ark. Fys. **31**, 491 (1966).
[13] R. Takahashi, Bull. Soc. Math. France **91**, 289 (1963).
[14] S. C. Peng and K. T. Hecht, J. Math. Phys. **8**, 1233 (1967).
[15] M. K. F. Wong, J. Math. Phys. **8**, 1899 (1967).
[16] K. B. Wolf, J. Math. Phys. **12**, 197 (1971).
[17] M. K. F. Wong, J. Math. Phys. **15**, 25 (1975).
[18] T. Mackawa, J. Math. Phys. **16**, 334 (1975).
[19] J. Fischer, J. Niederle, and R. Raczka, J. Math. Phys. **7**, 816 (1966); R. Raczka, N. Limic, and J. Niederle, *ibid.* **7**, 1861 (1966); B. Limic, J. Niederle and R. Raczka, *ibid.* **7**, 2026 (1966); **8**, 1079 (1967); A. V. Nikolov, Funk. Anal. **2**, 99 (1968); **3**, 91 (1969).
[20] A. Syed, J. Math. Phys. **24**, 463 (1983).
[21] W. J. Holman and L. C. Biedenharn, Ann. Phys. **39**, 1 (1966); **47**, 205 (1968).
[22] E. P. Wigner, *Group Theory* (Academic, New York, 1959), p. 172.
[23] Recall that $T(a)$ and $L_{14}(-\phi_{44})$, etc., are the operators representing $a$ and $l_{14}(-\phi_{44})$ etc. in $\mathscr{D}^{q_j q_l}$.
[24] I. Ferreti and M. Verde, Nuovo Cimento **55**, 110 (1968).

# Some geometrical consequences of physical symmetries

Jan Rzewuski
*Institute of Theoretical Physics, Wrocław University, 50-205 Wrocław, Cybulskiego 36, Poland*

(Received 23 August 1983; accepted for publication 18 November 1983)

Invariant submanifolds of the linear representation space $\mathbb{C}^{4m}$ of the physical symmetry group $SU(2,2)\times SU(m)$ and its subgroup $\mathscr{P}\times SU(m)$ are studied in some detail. It is shown that there exists only one such manifold admitting unique projection onto Minkowski space. The structure of this manifold is investigated by using proper local coordinate systems.

PACS numbers: 02.20. + b, 02.40. + m, 11.30. − j

## INTRODUCTION

I describe in this paper some of the geometrical consequences of the physical symmetry $SU(2,2)\times SU(m)$ and $\mathscr{P}\times SU(m)$ and the assumption that the physical space is the minimal invariant submanifold, containing Minkowski space, of the complex linear representation space $\mathbb{C}^{4m}$ of this symmetry.

The assumption is, of course, quite arbitrary. It can be justified, to some extent, by the basic role played by spinors in the description of elementary particles. The idea to use spinor spaces as the geometrical basis is quite old[1] and has already many applications.[1–4] It has, so far, been limited to the direct product $SU(2,2)\times SU(2)$ for the full physical symmetry (cf., e.g., Ref. 4). Therefore, this paper may also be considered as an extension of the idea to $SU(2,2)\times SU(m)$ with arbitrary $m$ (cf. however also Ref. 5 where such an extension was considered for the first time in a different setting).

Another justification may be found in the desire to provide a common geometrical background for both the internal and external symmetries. A common background is quite natural if we consider the direct product $SU(2,2)\times SU(m)$ as a subgroup of some larger symmetry, say $GL(nm,\mathbb{C})$.

In the two sections to follow, I describe in some detail the invariant manifold mentioned above. First, some general properties of matrix manifolds, local coordinate system in these manifolds and their transformation character are derived in the general case of the direct product $GL(n,\mathbb{C})\times GL(m,\mathbb{C})$ (Sec. 1). These properties are then specialized to the physically interesting case of $SU(2,2)\times SU(m)$ and its subgroup $\mathscr{P}\times SU(m)$ and applied to the description of the invariant manifold in question (Sec. 2). It is shown that there exists only one such invariant submanifold of $\mathbb{C}^{4m}$ which admits a unique projection onto Minkowski space consistent with the physical symmetry group under consideration.

## 1. SOME PROPERTIES OF MATRIX MANIFOLDS

Consider the linear representation space $\mathbb{C}^{nm}$ of $GL(nm,\mathbb{C})$. With respect to the subgroup $GL(n,\mathbb{C})\times GL(m,\mathbb{C})$ this representation space decomposes into invariant submanifolds

$$\mathscr{O}_k := \{\xi\in\mathbb{C}^{nm}:\text{rank}\,\xi=k\},\quad k=0,1,\ldots,\min(n,m), \tag{1.1}$$

in such a way that

$$\mathscr{O}_k\cap\mathscr{O}_1=0\quad\text{for } 1\neq k, \tag{1.2}$$

and

$$\bigcup_{k=0}^{\min(n,m)}\mathscr{O}_k=\mathbb{C}^{nm}. \tag{1.3}$$

We introduce an atlas on $\mathscr{O}_k$ consisting of the $\binom{n}{k}\binom{m}{k}$ neighborhoods

$$\xi\begin{pmatrix}\alpha_1,\ldots,\alpha_k\\ a_1,\ldots,a_k\end{pmatrix}:=\det\begin{pmatrix}\xi_{a_1;\alpha_1} & ,\ldots, & \xi_{a_1;\alpha_k}\\ \vdots & & \vdots\\ \xi_{a_k;\alpha_1} & ,\ldots, & \xi_{a_k;\alpha_k}\end{pmatrix}\neq 0, \tag{1.4}$$

where $\{a_1,\ldots,a_k\}$ and $\{\alpha_1,\ldots,\alpha_k\}$ run over all possible selections of $k$ numbers out of $n$ or $m$ numbers, resp. In particular, in the neighborhood $\xi\binom{1,\ldots,k}{1,\ldots,k}\neq 0$ we can decompose the matrix

$$\xi=\{\xi_{a;\alpha}\}_{\substack{a=1,\ldots,n\\ \alpha=1,\ldots,m}} \tag{1.5}$$

in the following way,

$$\xi=\begin{pmatrix}K & B\\ A & Y\end{pmatrix}, \tag{1.6}$$

where

$$K:=\{\xi_{a';\alpha'}\}_{\substack{a'=1,\ldots,k\\ \alpha'=1,\ldots,k}},\quad B:=\{\xi_{a';\alpha''}\}_{\substack{a'=1,\ldots,k\\ \alpha''=k+1,\ldots,m}},$$
$$A:=\{\xi_{a'';\alpha'}\}_{\substack{a''=k+1,\ldots,n\\ \alpha'=1,\ldots,k}},\quad Y:=\{\xi_{a'';\alpha'}\}_{\substack{a''=k+1,\ldots,n\\ \alpha''=k+1,\ldots,m}}. \tag{1.7}$$

Four different local coordinate systems can be introduced in the neighborhood $\det K\neq 0$ by means of the formulas

$$Y=AK^{-1}B=aB=Ab=aKb, \tag{1.8}$$

where

$$a=AK^{-1},\quad b=K^{-1}b. \tag{1.9}$$

It is seen from (1.6) and (1.8) that the complex dimension of $\mathscr{O}_k$ is

$$\dim\mathscr{O}_k=k(n+m-k). \tag{1.10}$$

In a similar way local coordinate systems are introduced in the other neighborhoods (1.4).

We can consider $\mathscr{O}_k$ as the set of independent coordinates of $n$ complex $m$-vectors or $m$ complex $n$-vectors of which only $k$ are linearly independent.

We shall need the following statements concerning the relation between some of these coordinate systems (1.8) and their transformation properties.

*Statement 1:* On the common part of the respective neighborhoods, the following relations hold:

  

$$a_{a''a'} = \xi\begin{pmatrix}1,...,k\\1,...,k\end{pmatrix}^{-1}\xi\begin{pmatrix}1, & ...,k\\1,...,a'-1,a'',a'+1, & ...,k\end{pmatrix}$$
$$= \xi\begin{pmatrix}\alpha_1,...,\alpha_k\\1,...,k\end{pmatrix}^{-1}\xi\begin{pmatrix}\alpha_1,..., & ...,\alpha_k\\1,...,a'-1,a'',a'+1, & ...,k\end{pmatrix}, \quad \text{for } a'=1,...,k, \quad a''=k+1,...,n, \tag{1.11}$$

$$b_{\alpha'\alpha''} = \xi\begin{pmatrix}1,...,k\\1,...,k\end{pmatrix}^{-1}\xi\begin{pmatrix}1,...,\alpha'-1,\alpha'',\alpha'+1, & ...,k\\1,... & ...,k\end{pmatrix}$$
$$= \xi\begin{pmatrix}1,...,k\\a_1,...,a_k\end{pmatrix}^{-1}\xi\begin{pmatrix}1,...,\alpha'-1,\alpha'',\alpha'+1, & ...,k\\a_1,... & ...,a_k\end{pmatrix}, \quad \text{for } \alpha'=1,...,k, \quad \alpha''=k+1,...,m. \tag{1.12}$$

The proof follows from general properties of matrices and can be found, e.g., in Ref. 6. One can easily verify that formulas (1.11) and (1.12) admit extension to $a_{aa'}$ and $b_{\alpha'\alpha}$ with $a = 1,...,n$; $\alpha = 1,...,m$; and $a',\alpha' = 1,...,k$, and that

$$a_{a'b'} = \delta_{a'b'}, \quad \text{for } a',b' = 1,...,k,$$
$$b_{\alpha'\beta'} = \delta_{\alpha'\beta'}, \quad \text{for } \alpha',\beta' = 1,...,k. \tag{1.13}$$

Consider now the transformation properties of the coordinates. With respect to GL($n$,ℂ)×GL($m$,ℂ) the matrix $\xi$ transforms according to

$$\xi \rightarrow \xi' = g\xi\tilde{h}, \tag{1.14}$$

where $g \in \text{GL}(n,\mathbb{C})\times \mathbb{1}$ and $h \in \mathbb{1}\times\text{GL}(m,\mathbb{C})$. The corresponding transformation properties of the matrices $a$ and $b$ are (cf. Ref. 6)

$$a \rightarrow a' = a(\xi') = a(g\xi\tilde{h}) = a(g\xi),$$
$$b \rightarrow b' = b(\xi') = b(g\xi\tilde{h}) = b(\xi\tilde{h}). \tag{1.15}$$

We have, therefore, the statement

*Statement 2:* $a$ is $\mathbb{1}\times h$ and $b$ is $g\times\mathbb{1}$ invariant.

The explicit form of (1.15) is

$$a'_{a''b'} = \frac{\sum_c g\begin{pmatrix}c_1, & ...,c_k\\1,...,b'-1,a'',b'+1, & ...,k\end{pmatrix} a\begin{pmatrix}1,...,k\\c_1,...,c_k\end{pmatrix}}{\sum_d g\begin{pmatrix}d_1,...,d_k\\1,...,k\end{pmatrix} a\begin{pmatrix}1,...,k\\d_1...d_k\end{pmatrix}}, \tag{1.16}$$

$$b'_{\alpha'\beta''} = \frac{\sum_\gamma b\begin{pmatrix}\gamma_1,...,\gamma_k\\1,...,k\end{pmatrix} h\begin{pmatrix}1,...,\alpha'-1,\beta'',\alpha'+1,...,k\\\gamma_1,... & ...,\gamma_k\end{pmatrix}}{\sum_\delta b\begin{pmatrix}\delta_1,...,\delta_k\\1,...k\end{pmatrix} h\begin{pmatrix}1,...,k\\\delta_1,...,\delta_k\end{pmatrix}}. \tag{1.17}$$

The various factors in (1.16) and (1.17) are subdeterminants of the matrices $a,b,g,h$ taken according to the general rule

$$m\begin{pmatrix}s_1,...,s_k\\r_1,...,r_k\end{pmatrix} := \det\begin{pmatrix}m_{s_1r_1} & \cdots & m_{s_1r_k}\\ \vdots & & \vdots\\ m_{s_kr_1} & \cdots & m_{s_kr_k}\end{pmatrix}. \tag{1.18}$$

The sums in (1.9) and (1.17) are over all the $\binom{n}{k}$ or $\binom{m}{k}$ possibilities to choose $k$ different numbers out of $n$ or $m$ numbers, resp. These formulas show that Statement 2 can be completed by the following:

*Statement 3:* The elements of the matrix $a$ transform among themselves, and similarly the elements of $b$, under the transformations of GL($n$,ℂ)×GL($m$,ℂ).

We are interested eventually in the physically important case of SU(2,2)×SU($m$) or its subgroups $\mathscr{P}$×SU($m$). Therefore, we are going now to specialize the above results to GL(4,ℂ)×GL($m$,ℂ) and to derive the consequences of further restriction of the symmetry to SU(2,2)×SU($m$) or $\mathscr{P}$×SU($m$).

## 2. THE MODEL

Let us consider first the, still too general, case GL(4,ℂ)×GL($m$,ℂ) with $0\leqslant k\leqslant \min(4,m)$. With each pair $\xi_\alpha := \{\xi_{a;\beta}\}_{a=1,...,4}$ and $\xi_\beta := \{\xi_{a;\beta}\}_{a=1,...,4}$ of the $m$ complex four-vectors represented by the $4\times m$ matrix $\xi = \{\xi_{a;\alpha}\}_{a=1,...,4;\ \alpha=1,...,m}$ one can associate the $2\times 2$ matrix $a^{(\alpha,\beta)}_{a''a'}$ defined by the formulas [cf. (1.11)]

$$a^{(\alpha,\beta)}_{a''1} = \xi\begin{pmatrix}\alpha,\beta\\1,2\end{pmatrix}^{-1}\xi\begin{pmatrix}\alpha, & \beta\\a'', & 2\end{pmatrix},$$
$$a^{(\alpha,\beta)}_{a''2} = \xi\begin{pmatrix}\alpha,\beta\\1,2\end{pmatrix}^{-1}\xi\begin{pmatrix}\alpha, & \beta\\1, & a''\end{pmatrix}, \quad a''=3,4. \tag{2.1}$$

Again with each such $2\times 2$ complex matrix one can associate a complex four-vector by means of the Pauli relation

$$z^{(\alpha,\beta)}_\mu = -(\lambda/2)(\sigma_\mu)^{a''b'}a^{(\alpha,\beta)}_{a''b'}, \tag{2.2}$$

where $\lambda$ is a constant with dimension of length ($a$ is dimensionless).

It can be shown (cf., e.g., Refs. 2,4,7,8) that conformal linear transformations SU(2,2)×$\mathbb{1}$ of the matrix $\xi$ induce, via $a_{a''b'}$, conformal nonlinear transformations of each of the complex vectors $z^{(\alpha,\beta)}_\mu$. In the infinitesimal version they have the form

$$z_\lambda \rightarrow z'_\lambda = z_\lambda - \epsilon z_\lambda - \epsilon_\lambda + \bar{\epsilon}^\mu(g_{\mu\lambda}z^2 - 2z_\mu z_\lambda) + \epsilon^{\mu\nu}(g_{\mu\lambda}z_\nu - g_{\nu\lambda}z_\mu), \tag{2.3}$$

representing dilatations, translations, special conformal transformations, and Lorentz rotations. It is seen from (2.3) that the coordinates $x_\mu = \frac{1}{2}(z_\mu + z^*_\mu)$ of the real part of $z_\mu$ transform among themselves like a real Minkowski vector with respect to dilatations, translations, and rotations. The coordinates $y_\mu = \frac{1}{2}(z_\mu - z^*_\mu)$ of the imaginary part of $z_\mu$ transform similarly, the only difference being their invariance with respect to translations. The coordinates of the real and imaginary part of $z_\mu$ are transformed into each other by the special conformal transformation only. A consequence of these facts is that in the case of Poincaré symmetry (extended possibly by dilatations) one can consider the coordinates $x_\mu = (x^{(1)}_\mu + x^{(2)}_\mu)$ and $y_\mu = x^{(1)}_\mu - x^{(2)}_\mu$ as the proper linear combinations of two vectors $x^{(1)}_\mu$ and $x^{(2)}_\mu$ of the same Minkowski space $M_4$. This interpretation corresponds to the idea of Yukawa's bilocal theory[9] in which the coordinates $x_\mu = \frac{1}{2}(x_\mu^{(1)} + x_\mu^{(2)})$ of the center of mass of the elementary particle and the relative coordinates $y_\mu = x_\mu^{(1)} - x_\mu^{(2)}$ were introduced *a priori*. It breaks down if we extend the symmetry to the full conformal group due to the mixing of $x_\mu$ and $y_\mu$ caused by special conformal transformation.

So far we have considered only conformal transformations of the external group $SU(2,2)\times \mathbb{1}$. What are the transformation properties of $z_\mu^{(\alpha,\beta)}$ with respect to transformations of the internal group $\mathbb{1}\times SU(m)$? The second order determinants $\xi\binom{\alpha,\beta}{a,b}$ appearing in the numerator and denominator of $a_{a''a'}^{(\alpha,\beta)}$ in (2.1) transform with respect to the Greek indices as an $\binom{m}{2}$-dimensional representation of $\mathbb{1}\times SU(m)$.

This situation is highly unsatisfactory for all $k>2$ because of two reasons: First of all, for $m>2$ we have $\binom{m}{2}$ different complex Minkowski spaces $M_4^{(\alpha,\beta)}$ with coordinates $z_\mu^{(\alpha,\beta)}$ and, therefore there is no unique projection from $\mathscr{O}_k$ onto $M_4$. Secondly, the $z_\mu^{(\alpha,\beta)}$ and, therefore, also the real parts $x_\mu^{(\alpha,\beta)}$ are not invariant with respect to internal symmetries which contradicts experimental evidence. Thus all invariant submanifolds $\mathscr{O}_k\subset\mathbb{C}^{nm}$ with $k>2$ must be discarded. Also the manifolds $\mathscr{O}_0,\mathscr{O}_1$ are out of question because $\mathscr{O}_0$ is the point $\xi=0$ and $\mathscr{O}_1$ does not admit an imbedding of $M_4$ according to (2.2) because all second-order determinants vanish.

Thus we are left with $\mathscr{O}_2$ and we shall show now that in this case the projection $\mathscr{O}_2\to M_4$ is unique and invariant with respect to the internal symmetry group $\mathbb{1}\times GL(m,\mathbb{C})$.

Indeed, from Statement 1 it follows that on the common part of the respective neighborhoods

$$a_{a''a'}^{(\alpha,\beta)} = a_{a''a'}^{(1,2)} = a_{a''a'}, \tag{2.4}$$

and, consequently,

$$z_\mu^{(\alpha,\beta)} = z_\mu^{(1,2)} = z_\mu. \tag{2.5}$$

The projection is unique. Moreover from Statement 2 it follows that the (unique) $a_{a''a'}$ and, therefore, also $z_\mu$ are invariant with respect to the internal symmetry group $\mathbb{1}\times GL(m,\mathbb{C})$. Finally from Statement 3, we infer that the matrix elements of the matrix $a$ and, therefore, also the $z_\mu$ transform among themselves with respect to the external symmetry $GL(n,\mathbb{C})\times\mathbb{1}$.

It is seen that $\mathscr{O}_2$ is the only invariant submanifold of $\mathbb{C}^{nm}$ which admits a unique projection on $M_4$ consistent with the group.

If we now restrict the symmetry to the physically interesting case of $SU(2,2)\times SU(m)$ the invariant manifold $\mathscr{O}_2$ will decompose into submanifolds according to the existence of two independent $SU(2,2)\times SU(m)$-invariants

$$r_{;\dot\alpha\alpha} \quad\text{and}\quad r_{;\dot\alpha\beta}r_{;\dot\beta\alpha}, \tag{2.6}$$

where

$$r_{;\dot\alpha\beta} = \xi^*_{a;\alpha} f^{\dot a b}\xi_{b;\beta} \tag{2.7}$$

is the SU(2,2) invariant Hermitian SU($m$)-tensor and $f^{\dot a b}$ the Hermitian matrix with eigenvalues $1,1,-1,-1$, determining the transformations of SU(2,2).

It is convenient to use a representation in which

$$f = \begin{pmatrix} 0 & i\sigma_0 \\ -i\sigma_0 & 0\end{pmatrix}. \tag{2.8}$$

In this representation

$$r_{;\dot\alpha\beta} = 2\lambda^{-1} y^\mu r_{\mu;\dot\alpha\beta}, \tag{2.9}$$

where

$$r_{\mu;\dot\alpha\beta} = \xi^*_{a';\alpha}(\sigma_\mu)^{\dot a' b'}\xi_{b';\beta}. \tag{2.10}$$

Equation (2.9) is a consequence of the second equation (1.8) specialized to the case $n=4$, $k=2$.

By virtue of (2.9) the two invariants (2.6) can be written in the form

$$r_{;\dot\alpha\alpha} = 2\lambda^{-1} y^\mu r_\mu,\quad r_{;\dot\alpha\beta}r_{;\dot\beta\alpha} = 4\lambda^{-2} y^\mu y^\nu r_{\mu\nu}, \tag{2.11}$$

where

$$r_\mu := r_{\mu;\dot\alpha\alpha},$$
$$r_{\mu\nu} := r_{\mu;\dot\alpha\beta} r_{\nu;\dot\beta\alpha} = -\tfrac{1}{2} g_{\mu\nu} r_\lambda r^\lambda + r_\mu r_\nu. \tag{2.12}$$

Introducing the second equation (2.12) into the second relation (2.11) one obtains

$$r_{;\dot\alpha\beta}r_{;\dot\beta\alpha} = 4\lambda^{-2}\{(r_\mu y^\mu)^2 - \tfrac{1}{2} r_\lambda r^\lambda y_\rho y^\rho\}. \tag{2.13}$$

We can use, therefore, instead of (2.6) the two invariants

$$r_\mu y^\mu \quad\text{and}\quad r_\mu r^\mu y_\nu y^\nu. \tag{2.14}$$

The submanifolds of $\mathscr{O}_2$ can now be described by the two equations

$$r_\mu y^\mu = -c_1,\quad y_\mu y^\mu r_\nu r^\nu = c_2. \tag{2.15}$$

To describe these manifolds in more detail, let us note that $r_\nu$ is a "time-like" vector pointing towards the "future"

$$\begin{aligned} r_\nu r^\nu &= -2\sum_{\alpha,\beta}|\xi\tbinom{\alpha,\beta}{1,2}|^2 \\ &= -4\{\|\xi_1\|^2\|\xi_2\|^2 - |\langle\xi_1,\xi_2\rangle|^2\} = -\kappa^2, \end{aligned} \tag{2.16}$$
$$r_0 = \sum_{a',\alpha}|\xi_{a';\alpha}|^2 = \|\xi_1\|^2 + \|\xi_2\|^2 > 0,$$

where

$$\langle\xi_1,\xi_2\rangle := \sum_{\alpha=1}^m \xi^*_{1;\alpha}\xi_{2;\alpha},\quad \|\xi_{a'}\|^2 = \sum_{\alpha=1}^m |\xi_{a';\alpha}|^2. \tag{2.17}$$

It is seen that the first equation (2.15) describes a hyperplane in the space of the variables $\{y_\mu\}$ perpendicular to the vector $r_\mu$. The second equation (2.15) describes a rotational ellipsoid with $y_0$-axis as symmetry axis. In the case when $c_2>0$ ($c_2<0$) $y_\mu$ is timelike (spacelike). In the first case these surfaces intersect for a proper choice of $c_1$ and $c_2$ [cf. (2.21)]. In the second case they intersect for all $c_1$ and $c_2$. Their union is an (dim $\mathscr{O}_2 - 1$)-dimensional invariant submanifold. Their intersection has one dimension less and is of particular interest in view of the assumption of minimality mentioned in the Introduction.

From the first equation (2.16) and second equation (2.15) we have

$$y_\mu y^\mu = -c_2/\kappa^2,\ y_0^2 = \mathbf{y}^2 + c_2/\kappa^2. \tag{2.18}$$

From the first equation (2.15), together with (2.18) we obtain

$$y_0 = (\mathbf{yr} + c_1)/r_0, \tag{2.19}$$

and

$$\mathbf{yy} - (\mathbf{yr} + c_1)^2/r_0^2 + c_2/\kappa^2 = 0. \tag{2.20}$$

(2.20) is a second-order equation for the three-vector $\mathbf{y}$ with coefficients depending on $\xi_{a';\alpha}$, $a' = 1,2$, $\alpha = 1,...,m$ by the intermediary of the vector $r_{\mu;}$ [cf. (2.12)]. It is symmetric with respect to rotations around $\mathbf{r}$. In a coordinate system in which $r_1 = r_2 = 0$, it has the form

$$\frac{y_1^2 + y_2^2}{(c_1^2 - c_2)/\kappa^2} + \frac{(y_3 - c_1 r_3/\kappa^2)^2}{((c_1^2 - c_2)/\kappa^2)\cdot r_0^2/\kappa^2} = 1. \tag{2.21}$$

For $c_1^2 > c_2$, (2.20) represents a rotational ellipsoid, for $c_1^2 < c_2$ (2.21) has no real solutions and we have to do with two three-dimensional disjoint surfaces [cf. (2.15),(2.18)]: one plane and one hyperboloid. Note that for $c_2 > 0$, $y_\mu$ is a timelike vector [cf. (2.18)]. In a coordinate system in which also $r_3$ vanishes $r_0 = \kappa$ and (2.21) becomes a sphere

$$\mathbf{y}^2 = (c_1^2 - c_2)/\kappa^2. \tag{2.22}$$

If we further restrict the symmetry to $\mathscr{P} \times SU(m)$ another invariant appears, namely $\kappa^2 = -r_\mu r^\mu$ [cf. first relation (2.16)]. The equations $\kappa^2 = \text{const}$ describe a one-parameter set of $(4m - 1)$-dimensional real submanifolds of $\mathbb{C}^{2m}$ given by the equations [cf. first relation (2.16)]

$$\|\xi_1\|^2\|\xi_2\|^2 - |\langle\xi_1,\xi_2\rangle|^2 = (\kappa/2)^2. \tag{2.23}$$

We have mentioned already that the three sets of variables $\{x_\mu\}$, $\{y_\mu\}$, and $\{\xi_{a';\alpha}\}$, $\mu = 1,...,4$; $a' = 1,2$; $\alpha = 1,...,m$, do not mix under transformations from $\mathscr{P} \times SU(m)$ and, therefore, we can consider $x_\mu = \frac{1}{2}(x_\mu^{(1)} + x_\mu^{(2)})$ and $y_\mu = x_\mu^{(1)} - x_\mu^{(2)}$ as proper linear combinations of the coordinates of two points in the same Minkowski space. It is seen first of all that in the case $c_2 < c_1^2$ the relative coordinates are restricted to the surface of the ellipsoid (2.20). There appears, moreover, another Minkowski timelike four-vector $r_\mu$ of constant length ($r_\mu r^\mu = -\kappa^2$) and pointing towards the future. This vector determines the direction of the symmetry axis and the ratio of the axes of the ellipsoid and is itself determined by the position of the point $\{\xi_{a';\alpha}\}$ of $\mathbb{C}^{2m}$ on the surface (2.23). The coordinates $x_\mu$ of the center of mass of the particle are not restricted. The invariant submanifolds of $\mathscr{O}_2$ have in this case $4(4 + m - 2) - 3 = 4m + 5$ real dimensions and consist each of the Minkowski space of the four real variables $x_\mu$, $\mu = 0,1,2,3$, the two-dimensional ellipsoid (2.20) determined by the values of the coordinates of the four-vector $r_\mu$ which are functions of $\xi_{a';\alpha}$, $a' = 1,2$; $\alpha = 1,...,m$, and of the $4m - 1$ real variables on the surface (2.23).

There exist also $SU(2,2) \times GL(m,\mathbb{C})$-invariant $(4m + 4)$-dimensional submanifolds of $\mathscr{O}_2$ determined by the conditions

$$r_{;\dot{\alpha}\beta} = 0. \tag{2.24}$$

Due to the fact that Eqs. (2.29) can be solved with respect to $y_\mu$ for any pair of indices $\alpha$, $\beta$, of the set $1,...,m$, condition (2.24) implies $y_\mu = 0$. From the space-time structure only the timelike direction $r_\mu$ remains. According to (2.7),(2.6) both invariant forms vanish in this case and we have to do with the isotropic submanifold.

Another kind of invariant condition would be

$$r_{;\dot{\alpha}\beta} = \delta_{\dot{\alpha}\beta}. \tag{2.25}$$

However, one easily persuades oneself that this condition is consistent on $\mathscr{O}_2$ only in the case $m = 2$ and, therefore, has a rather limited application.

One may note that all considerations concerning the relations between the three Minkowski vectors $x_\mu$, $y_\mu$, $r_\mu$ are independent on $m$.

The generators and Casimirs of the symmetry group in the Hilbert space of functions over the minimal manifold were derived in Ref. 8 in the case of $SU(2,2) \times SU(2)$ or $\mathscr{P} \times SU(2)$ in terms of the local coordinates $\{x_\mu, y_\mu, \xi_{a';\alpha}\}$ and $\{x_\mu, r_{;\dot{\alpha}\beta}, \xi_{a';\alpha}\}$ [in the case of $SU(2,2) \times SU(2)$, $y_\mu$ is a linear invertible form in $r_{;\dot{\alpha}\beta}$ with coefficients depending on the variables $\xi_{a';\alpha}$]. In the general case of $SU(2,2) \times SU(m)$ the results obtained for the first set of coordinates can be taken over from the particular case $SU(2,2) \times SU(2)$ by extending the summation over the index $\alpha$ from 2 to $m$.


[1] J. Rzewuski, Bull. Acad. Pol. Sci., Ser. Sci.Math. Astr. Phys. **6**, 261, 335 (1958); 7; 571 (1959); **8**, 777,783 (1960); Nuovo Cimento **5**, 94 (1958); Acta Physica Pol. **17**, 417 (1958); **18**, 549 (1959).

[2] R. Penrose, Ann. Phys. **10**, 171 (1960); J. Math. Phys. **8**, 345 (1967); Quantum Theory and Structure of Space Time, Munich, 1975; R. Penrose and M. A. H. MacCallum, Phys. Rep. C **6**, 241 (1972).

[3] R. Penrose, in *General Theory of Relativity, Selected Reading in Physics* (Pergamon, New York, 1973).

[4] L. P. Hugston, "Twistors and Particles," in *Lecture Notes in Physics* (Springer, New York, 1979), Vol. 97.

[5] J. Rzewuski, J.Math. Phys. **23**, 1573 (1982).

[6] J. Rzewuski, in *Quantum Theory of Fields and Particles. Birthday Volume to Jan Lopuszanski*, edited by B. Jancewicz and J. Lukierski (World Sci., Singapore, 1983).

[7] R. Ablamowicz, J. Mozrzymas, Z. Oziewicz, and J. Rzewuski, Rep. Math. Phys. **14**, 89 (1978); R. Ablamowicz, Z. Oziewicz, and J. Rzewuski, J. Math. Phys. **23**, 231 (1982).

[8] J. Rzewuski, Bull. Acad. Pol. Sci. **28**, 63 (1980); *Description of Physical Laws in the Spinor Space* (Acta Universitatis Wratislaviensis, Wroclaw, in press).

[9] H. Yukawa, Phys. Rev. **77**, 219 (1950); **80**, 1047 (1950); **91**, 415 (1953); J. Rayski, Nuovo Cimento **2**, 255 (1955).

# Proof of an algorithm for the evaluation of the branching multiplicity SO(2n)→SO(2n − 2) ⊗ U(1)

V. Amar, U. Dozzio, and C. Oleari
*Instituto di Fisica dell'Università, 43100 Parma, Italy and Istituto di Fisica Nucleare, Sezione di Milano, Milan, Italy*

(Received 28 April 1981; accepted for publication 9 December 1983)

The proof of an algorithm, previously proposed by us, for the evaluation of the branching multiplicity SO(2$n$)→SO(2$n$ − 2) ⊗ U(1) is given. This proof is based on explicit construction of lowering shift operators for the class $D_n$ of Cartan.

PACS numbers: 02.20. + b, 02.60.Gf

## INTRODUCTION

In a series of papers[1-3] we proposed an algorithm for computing the branching multiplicity in the reduction SO(2$n$)→SO(2$n$ − 2) ⊗ U(1). Using this algorithm, we made also a very efficient computer program[3] for the evaluation of the inner multiplicity of SO(2$n$) and SO(2$n$ − 1). The validity of the proposed algorithm has been verified by a large number of numerical tests by computer; however, this algorithm was so far without proof.

In this paper we give a proof of our algorithm. This proof enables us to better understand the surprising fact that by some constraints on the Gel'fand triangle[4] we can evaluate the degeneracy of the eigenvalues of the elements of the Cartan's subalgebra, in spite of the fact that the Gel'fand triangle is an orthogonal basis for the irreducible representation (IR) of SO(2$n$), in which, however, the elements of the Cartan's subalgebra are not diagonal.

## NOMENCLATURE

We use the tensorial notation introduced by Louck and Biedenharn[5] for the unitary groups and recently by Bincer[6] for the orthogonal groups. We denote the generators of SO(2$n$) by $C^a_b$ with the indices ranging from $-n$ to $+n$, zero excluded. Their commutation relations are

$$[C^a_b,C^c_d] = \delta^c_b C^a_d - \delta^a_d C^c_b + \delta^{\bar b}_d C^c_{\bar a} - \delta^c_{\bar a} C^{\bar b}_d, \qquad (1)$$

where

$$\bar a = -a. \qquad (2)$$

These $C$'s obey

$$C^a_b = -C^{\bar b}_{\bar a}; \qquad (3)$$

moreover, in the unitary representations we demand that $C^{a^+}_b = C^b_a$. The generators $C^a_a$, $1 \leqslant a \leqslant n$, which are the elements of the Cartan subalgebra of SO(2$n$), may be taken simultaneously diagonal. Let $|m\rangle$ denote a simultaneous eigenvector of any $C^a_a$:

$$C^a_a|m\rangle = m_a|m\rangle, \quad \bar n \leqslant a \leqslant n,$$

where

$$m = (m_n, m_{n-1}, \ldots, m_1, m_{\bar 1}, \ldots, m_{\bar n}) \qquad (4)$$

is called the weight of the vector $|m\rangle$ and the $m_a$'s are the components of the weight. From (3) it follows that $m_{\bar a} = -m_a$, and therefore the last $n$ entities in relation (4) are redundant and can be omitted. The usual ordering between the weights is $m \gtrless m'$ if $m_a - m'_a \gtrless 0$ for the highest $a$ such that $m_a - m'_a$ is nonzero.

It follows from Eq. (1) that

$$C^c_c\{C^a_b|m\rangle\} = (m_c + \delta^a_c - \delta^c_b + \delta^{\bar b}_c - \delta^c_{\bar a})C^a_b|m\rangle,$$

so that we may write

$$C^a_b|m\rangle \propto |m'\rangle, m'_c = m_c + \delta^a_c - \delta^c_b + \delta^{\bar b}_c - \delta^c_{\bar a}.$$

Consequently,

$$m' \gtrless m \quad \text{if } a \gtrless b,$$

and we may classify generators as *raising*, *weight*, and *lowering* generators. The IR's of SO(2$n$) will be classified by their heighest weight $M = (M_n, M_{n-1}, \ldots, M_1)$. For the dominant weights it must hold

$$m_n \geqslant m_{n-1} \geqslant \cdots \geqslant m_2 \geqslant |m_1|. \qquad (5)$$

*Tensor* $T^c_d$ and *vector* $V_d$ operators are defined as follows:

$$\begin{aligned}[][C^a_b, T^c_d] &= \delta^c_b T^a_d - \delta^a_d T^c_b + \delta^{\bar b}_d T^c_{\bar a} - \delta^c_{\bar a} T^{\bar b}_d,\\ [C^a_b, V_d] &= -\delta^a_d V_b + \delta^{\bar b}_d V_{\bar a}, \quad 1 \leqslant |a|,|b|,|d| \leqslant n.\end{aligned} \qquad (6)$$

Similarly to the generators, the tensors can also be defined as raising, weight, and lowering operators; moreover, the vectors $V_d$ are lowering if $d \geqslant 1$ or rising if $d \leqslant -1$.

## SEMIMAXIMAL STATES AND SHIFT OPERATORS

Let us define *semimaximal vectors* $|sm\rangle$ the vectors satisfying the conditions

$$\begin{cases} C^a_b|sm\rangle = 0 \quad \text{if } a > b, \quad 2 \leqslant |a|,|b| \leqslant n, \\ C^a_a|sm\rangle = m_a|sm\rangle, \quad 1 \leqslant |a| \leqslant n. \end{cases} \qquad (7)$$

Evidently, the vectors $|sm\rangle$ satisfying Eqs. (7) are vectors with heighest weight for SO(2$n$ − 2) and with definite weight for SO(2$n$). Now we define as *shift operators* $S^{\pm 1}_\mu$ the polynomials of generators of SO(2$n$) such that

$$\begin{cases} [C^a_b, S^{\pm 1}_\mu]|sm\rangle = 0 \quad \text{for } a < b, \ 2 \leqslant |a|,|b|,|\mu| \leqslant n, & (8)\\ [C^a_a, S^{\pm 1}_\mu]|sm\rangle = (\delta^{\pm 1}_a - \delta^a_\mu + \delta^{\bar\mu}_a - \delta^{\bar a}_{\pm 1})S^{\pm 1}_\mu|sm\rangle, & (9)\end{cases}$$

$$1 \leqslant a \leqslant n, \ 2 \leqslant |\mu| \leqslant n.$$

From Eqs. (8) and (9) we have that $S^{\pm 1}_\mu|sm\rangle$ is a semimaximal vector whose weight has the $|\mu|$th component lowered or raised by 1 if $\mu$ is positive or negative and the component $m_1$ becomes $(m_1 \pm 1)$. Our aim is to construct explicitly lowering operators $S^{\pm 1}_\mu$ such that

  

$$[S^i_\mu, S^{i'}_{\mu'}]\,|sm\rangle = 0, \quad 2 \leqslant \mu, \mu' \leqslant n, \; i,i' = \pm 1. \tag{10}$$

Following the technique of Bincer,[6] let $V(\mu)^{\pm 1}_d$ be for fixed $\mu$ an SO$(2n-2)$ vector operator, which transforms $m_1$ into $(m_1 \pm 1)$. If we set $S^{\pm 1}_\mu = V(\mu)^{\pm 1}_\mu$, Eq. (9) holds true if Eq. (6) holds; moreover, Eq. (8) becomes

$$[\delta^{\bar b}_\mu V(\mu)^{\pm 1}_{\bar a} - \delta^a_\mu V(\mu)^{\pm 1}_b]\,|sm\rangle = 0, \quad a > b. \tag{11}$$

In order that Eq. (11) is satisfied, it is sufficient that

$$V(\mu)^{\pm 1}_d |sm\rangle = 0 \quad \text{for } \mu > d. \tag{12}$$

A solution of Eq. (12) can be found recursively as follows: let $V(\mu)^{\pm 1}_d$ satisfy Eq. (12) and define

$$V(\mu+1)^{\pm 1}_d \equiv \{V(\mu)(C - C_\mu I)\}^{\pm 1}_d = \sum_{a=\bar n}^{n}{}' V(\mu)^{\pm 1}_a (C - C_\mu I)^a_d,$$

where $C_\mu$ is a number to be evaluated below and the prime on $\Sigma'$ indicates that the range of $a$ is from $\bar n$ to $n$, 0 and $\pm 1$ excluded. We have

$$V(\mu+1)^{\pm 1}_d = \sum_{a=\bar n}^{d}{}' V(\mu)^{\pm 1}_a (C - C_\mu I)^a_d + \sum_{a=d+1}^{n}{}' V(\mu)^{\pm 1}_a C^a_d \doteq \sum_{a=\bar n}^{d}{}' V(\mu)^{\pm 1}_a (C - C_\mu I)^a_d,$$

where $\doteq$ means that the equation holds when both sides are applied to semimaximal states.

Next by using relation (6) we obtain

$$\begin{aligned} V(\mu+1)^{\pm 1}_d &\doteq V(\mu)^{\pm 1}_d (C^d_d - C_\mu) + \sum_{a=\bar n}^{d-1}{}' \{C^a_d V(\mu)^{\pm 1}_a + [V(\mu)^{\pm 1}_a, C^a_d]\} \\ &\doteq V(\mu)^{\pm 1}_d \Big[m_d - C_\mu + \sum_{a=\bar n}^{d-1}{}' (1 - \delta^{\bar d}_a)\Big] + \sum_{a=\bar n}^{d-1}{}' C^a_d V(\mu)^{\pm 1}_a. \end{aligned} \tag{13}$$

For $\mu > d$ the rhs of (13) vanishes. For $\mu = d$ the rhs of (12) vanishes too if $C_\mu$ is chosen to be

$$C_\mu = m_\mu + \sum_{a=\bar n}^{\mu-1}{}' (1 - \delta^{\bar \mu}_a). \tag{14}$$

Hence $V(\mu+1)^{\pm 1}_d |sm\rangle = 0$ for $\mu + 1 > d$. By iteration we find the solution

$$V(\mu)^{\pm 1}_d = \Big\{V(\bar n) \prod_{j=\bar n}^{\mu-1}{}' (C - C_j I)\Big\}^{\pm 1}_d,$$

where $\Pi'$ indicates that $j$ is in the SO$(2n-2)$ range, and

$$V(\bar n)^{\pm 1}_d |sm\rangle = 0 \quad \text{for } \bar n > d. \tag{15}$$

But Eq. (15) is empty because the inequality $\bar n > d$ is never satisfied. It follows that the only requirement on $V(\bar n)^{\pm 1}_d$ is that it must be an SO$(2n-2)$ vector operator. The simplest choice is

$$V(\bar n)^{\pm 1}_d = C^{\pm 1}_d.$$

We conclude that

$$V(\mu)^{\pm 1}_d = \Big\{C \prod_{j=\bar n}^{\mu-1}{}' (C - C_j I)\Big\}^{\pm 1}_d.$$

Hence an operator $S^{\pm 1}_\mu = V(\mu)^{\pm 1}_\mu$ satisfying relations (8) and (9) is obtained; but it is worthwhile noticing that $S^{\pm 1}_\mu$ depends on the components of the weight to which is applied, and this must be kept in mind in the next sections.

## COMMUTATIVITY

Let $|m\rangle$ be a semimaximal vector with weight $m$ according to Eqs. (12), (13), and (14). Then we have

$$V(\mu)^{\pm 1}_d |m\rangle = \Big\{[C_d - C_{\mu-1}]\, V(\mu-1)^{\pm 1}_d + \sum_{a=\mu-1}^{d-1}{}' C^a_d V(\mu-1)^{\pm 1}_a\Big\}|m\rangle. \tag{16}$$

In Eq. (16) $C^a_d$ is a lowering operator. Since in an unitary representation $C^{a+}_b = C^b_a$, a lowering generator working on the left is a raising generator; hence

$$\langle m'|V(\mu)^{\pm 1}_d|m\rangle = \langle m'|\{C_d - C_{\mu-1}\}\, V(\mu-1)^{\pm 1}_d|m\rangle.$$

Consequently, by iteration,

$$\langle m'|V(\mu)^{\pm 1}_d|m\rangle = \prod_{j=\bar n}^{\mu-1}{}' (C_d - C_j)\langle m'|V(\bar n)^{\pm 1}_d|m\rangle$$

or

$$\langle m'|S^{\pm 1}_\mu|m\rangle = \prod_{j=\bar n}^{\mu-1}{}' (C_\mu - C_j)\langle m'|C^{\pm 1}_\mu|m\rangle. \tag{17}$$

We note that numbers $C_\mu$ and $C_j$ depend on the weight $m$. We can directly verify that $\Pi'^{\mu-1}_{j=\bar n}(C_\mu - C_j)$ is equal to zero only if $m_2 = 0$ and $\mu = 2$; therefore, we can define the following operator:

$$\tilde S^{\pm 1}_\mu \equiv S^{\pm 1}_\mu \Big/ \prod_{j=\bar n}^{\mu-1}{}' (C_\mu - C_j)$$

only if this operator is applied on a semimaximal vector with $m_2 \neq 0$. We can prove that $\tilde S^{\pm 1}_\mu$ satisfies (8) and (9), and the following holds:

$$\langle m'|\tilde S^{\pm 1}_\mu|m\rangle = \langle m'|C^{\pm 1}_\mu|m\rangle. \tag{18}$$

Let us consider

$$\langle m'|\tilde S^i_\nu \tilde S^{i'}_\mu|m\rangle \quad \text{with } i,i' = \pm 1,$$

by introducing a completeness in the subspace of the semimaximal vectors $|m_j\rangle$, which can be limited to the vectors $|\bar m\rangle$ with weight equal to the weight of $\tilde S^{i'}_\mu|m\rangle$; we obtain

$$\begin{aligned} \langle m'|\tilde S^i_\nu \tilde S^{i'}_\mu|m\rangle &= \sum_{m_j} \langle m'|\tilde S^i_\nu|m_j\rangle\langle m_j|\tilde S^{i'}_\mu|m\rangle \\ &= \sum_{\bar m} \langle m'|\tilde S^i_\nu|\bar m\rangle\langle \bar m|\tilde S^{i'}_\mu|m\rangle \\ &= \sum_{\bar m} \langle m'|C^i_\nu|\bar m\rangle\langle \bar m|C^{i'}_\mu|m\rangle \\ &= \sum_{m_j} \langle m'|C^i_\nu|m_j\rangle\langle m_j|C^{i'}_\mu|m\rangle \\ &= \langle m'|C^i_\nu C^{i'}_\mu|m\rangle. \end{aligned}$$

Hence we have

$$\begin{aligned} \langle m'|[\tilde S^i_\nu, \tilde S^{i'}_\mu]|m\rangle &= \langle m'|[C^i_\nu, C^{i'}_\mu]|m\rangle \\ &= \langle m'|(\delta^{\bar\nu}_\mu C^{i'}_{\bar i} - \delta^{i'}_{\bar i} C^{\bar\nu}_\mu)|m\rangle, \end{aligned}$$

which is equal to zero if $r$ and $\mu$ are positive, because $\delta_\mu^{\bar r} = 0$ and always $\langle m'|C_\mu^{\bar r}|m\rangle = 0$. Since $|m'\rangle$ is arbitrary relation (10) is true.

From these very important facts it follows that we can apply to a semimaximal vector a succession of $\widetilde S_\mu^{i'}$ in any order without loss of generality. However, it must be remarked that constants $C_\alpha$ that appear in the operators $\widetilde S$'s depend on the weight of the vector $|m\rangle$, on which $\widetilde S$'s act, i.e., in a relation like the following:

$$\widetilde S_r^{\pm 1}\widetilde S_\mu^{\pm 1}|m\rangle = \widetilde S_\mu^{\pm 1}\widetilde S_r^{\pm 1}|m\rangle,$$

the constants appearing in $\widetilde S_\mu^{\pm 1}$ on the lhs depend on the weight of $|m\rangle$ and these on the rhs depend on the weight of $\widetilde S_r^{\pm 1}|m\rangle$.

## THE ALGORITHM

We can obtain the whole set of semimaximal vectors with fixed weight $N = (N_n, N_{n-1},\dots,N_1)$ with $N_n \geqslant N_{n-1} \geqslant \cdots \geqslant N_2 \geqslant 0$ as follows:

$$|N\rangle_i = (\widetilde S_2^{-1})^{q_2 - N_2}(\widetilde S_3^{-1})^{q_3 - N_3}\dots(\widetilde S_n^{-1})^{q_n - N_n} \times (\widetilde S_2^{+1})^{M_2 - q_2}\dots(\widetilde S_n^{+1})^{M_n - q_n}|M\rangle, \quad (19)$$

where $|M\rangle = |M_n,\dots,M_1\rangle$ is the heighest weight of an IR of SO($2n$) and $N_1$ results

$$N_1 = M_1 + \sum_{j=2}^{n}(M_j + N_j) - 2\sum_{j=2}^{n} q_j, \quad (20)$$

with $q_j$ positive integer or positive half-odd integer as the $M_i$'s are.

As was shown before, any ordering of the operators $\widetilde S_k^{\pm 1}$ is equivalent. We choose the ordering (19).

Let us point out that we choose $M$ such that $M_1 \geqslant 0$ and $N$ such that $N_2 \geqslant 0$. However, this is not a limitation since, as we proved in a previous paper,[1] the situation with $M_1 < 0$ and/or $N_2 < 0$ can be reconduced to the above situation.

We remark that from the constraint $N_2 \geqslant 0$ it follows that the operator $\widetilde S_\mu^{\pm 1}$ is always applied on semimaximal vectors $|m\rangle$ with $m_2 > 0$, and, therefore, $\widetilde S_\mu^{\pm 1}$ is always well defined. From relation (19) obviously it holds

$$\left.\begin{matrix} q_i \leqslant M_i \\ q_i \geqslant N_i \end{matrix}\right\}, \quad i = 2,\dots,n. \quad (20a)$$

Since $\widetilde S_k^{\pm 1}$ transforms semimaximal vectors into semimaximal vectors and since dominance condition (5) must hold, we have

$$\left.\begin{matrix} q_i \geqslant M_{i-1} \\ N_i \geqslant q_{i-1} \end{matrix}\right\}, \quad i = 3,\dots,n. \quad (20b)$$

We know from our previous work (see Ref.[2]) that the following conditions:

$$\sum_{i=j}^{n} M_i \geqslant \sum_{i=j}^{n} N_i, \quad j = 1,\dots,n,$$

$$\sum_{i=2}^{n} M_i - M_1 \geqslant \sum_{i=2}^{n} N_i - N_1, \quad (21)$$

$$\sum_{i=1}^{n}(M_i - N_i) = \text{even integer}$$

are necessary and sufficient in order that a dominant vector belong to the weight diagram with highest weight $M$.

We remark that the dominant weight, obtained by the Weyl group from the weight $N = (N_n, N_{n-1},\dots,N_1)$, to which we apply construction (19), is by hypothesis a dominant weight satisfying (21). Relation (19) is meaningful only if the weights of $(\tilde S_\mu^{+1})^{M_\mu - q_\mu}|M\rangle \forall_\mu$ and of $(\tilde S_\mu^{-1})^{q_\mu - N_\mu}|M\rangle \forall_\mu$ belong to the IR defined by $M$. This request for $(\tilde S_2^{+1})^{M_2 - q_2}|M\rangle$ is satisfied if and only if $q_2 \geqslant M_1$ and in the other cases it is satisfied by the relations (20a) and (20b).

Furthermore, we can verify that any semimaximal vector $|m\rangle$ obtained by any partial application of the operators $\tilde S_\mu^{\pm 1}$ in relation (19) has weight belonging to the IR defined by $M$. These checks, even if tedious, are very simple and therefore omitted.

The new condition, in addition to (20a) and (20b), that we obtain is hence $M_1 \leqslant q_2$.

Consequently, a between condition holds, which can be represented in the following triangular form:

$$\begin{matrix} M_n & & M_{n-1} & & M_{n-2} & \cdots & M_2 & & M_1 \\ & q_n & & q_{n-1} & & \cdots & & q_2 & \\ & & N_n & & N_{n-1} & \cdots & & & N_2 \end{matrix} \quad (22)$$

where we have exactly the first three rows of a Gel'fand triangle related to an IR of SO($2n$).

Relation (20) with the constraint (22) for $q_i$'s is exactly the algorithm that we proposed in a previous paper[2,7] for evaluating the branching multiplicity SO($2n$) $\to$SO($2n-2$) $\otimes$ U(1).

Finally we have to show that vectors obtained by (19) are linearly independent. In particular, we will show that if the vectors obtained by (19) are linearly dependent, there exists a contradiction between the branching multiplicity SO($2n$)$\to$SO($2n-2$) obtained by other algorithms and that obtained by enumerating the vectors $|N\rangle_i$ of relation (19) with all the different values of $N_1$ and fixed $N_n,\dots,N_2$, which belong to the IR of SO($2n$) defined by $|M\rangle$. Particularly, given an IR of SO($2n$) defined by $(M_n, M_{n-1},\dots,M_1)$, the constraints on $N_1$, in order that the weight $(N_n, N_{n-1},\dots,N_1)$ with fixed $N_n, N_{n-1},\dots,N_2$ be a weight of the given IR, are given as follows: from condition (21) we have

$$\sum_{i=2}^{n} N_i + M_1 - \sum_{i=2}^{n} M_i \leqslant N_1 \leqslant \sum_{i=1}^{n} M_i - \sum_{i=2}^{n} N_i;$$

by comparison with (20) we obtain

$$\sum_{i=2}^{n} N_i + M_1 - \sum_{i=2}^{n} M_i \leqslant M_1 + \sum_{i=2}^{n}(M_i + N_i) - 2\sum_{i=2}^{n} q_i, \quad (23)$$

$$\sum_{i=1}^{n} M_i - \sum_{i=2}^{n} N_i \geqslant M_1 + \sum_{i=2}^{n}(M_i + N_i) - 2\sum_{i=2}^{n} q_i. \quad (24)$$

From (23) and (24) it follows that

$$\sum_{i=2}^{n} q_i \leqslant \sum_{i=2}^{n} M_i,$$

$$\sum_{i=2}^{n} q_i \geqslant \sum_{i=2}^{n} N_i.$$

These last conditions are always satisfied if the $q_i$'s obey the triangle condition (22); consequently, any choice of $q_i$'s is possible. Then it is shown that the number of vectors given by (20) with all the possible values of $N_1$ is exactly equal to the number of different sets of $q_i$'s, $i = n,n-1,...,2$, satisfying the triangle (22), and this number, as is well known, is equal to the branching multiplicity $SO(2n) \to SO(2n-2)$. Consequently, any different choice of $q_i$'s must give rise to vectors of the form (19) which are linearly independent. Q.E.D.

In conclusion we stress the result that the triangle (22), which is formally equal to the Gel'fand triangle, has, however, a very different group theoretical meaning.


[1]V. Amar, U. Dozzio, and C. Oleari, Physica **A114**, 341 (1982).
[2]V. Amar, U. Dozzio, and C. Oleari, Nuovo Cimento A **70**, 460 (1982).
[3]V. Amar, U. Dozzio, and C. Oleari, Comput. Phys. Commun. **29**, 201 (1983).
[4]I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotations and Lorentz Group and Their Applications* (Pergamon, London, 1963), p. 353.
[5]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **17**, 1611 (1976).
[6]A. M. Bincer, J. Math. Phys. **19**, 1173 (1978).

[7]The reader should be careful: the weight ordering $(M_n \geqslant M_{n-1} \geqslant \cdots \geqslant M_1)$ in this paper corresponds to $(M_1 \geqslant M_2 \geqslant \cdots \geqslant M_n)$ in Refs. 1–3.

# Fixed symmetry and fixed class generating functions

R. W. Gaskell
*Department of Physics, Lafayette College, Easton, Pennsylvania 18042*
R. T. Sharp
*Department of Physics, McGill University, Montréal, Quebec, Canada H3A 2T8*

(Received 23 December 1983; accepted for publication 3 February 1984)

A fixed symmetry (or fixed plethysm) generating function enumerates all representations $R_b$ of a compact Lie group $G$ contained in that part of the direct product of $p$ copies of any irreducible representation $R_a$ of $G$ that has a particular exchange symmetry under the symmetric group $S_p$. Fixed symmetry generating functions are conveniently given as linear combinations of the simpler fixed class generating functions. We give a systematic procedure for their construction and some examples for SU(2), SU(3), and SO(5). For SU(3) the examples include plethysms of up to three boxes; for SO(5) we treat two-box plethysms in general and give the scalar content of three-box plethysms; the SU(2) examples include up to six boxes.

PACS numbers: 02.20.Qs

## I. INTRODUCTION

A problem which occurs frequently in dealing with multiparticle systems is the following: Given a product of $p$ copies of an irreducible representation $R_a$ of a group $G$, what is the multiplicity of the irrep $R_b$ of $G$ contained in the component of the product having a given exchange symmetry under the permutation group $S_p$. Recently, a new type of generating function has been introduced[1] which provides the solution to this problem for all irreps of a compact Lie group $G$ for a fixed exchange symmetry. Examples have been given in Ref. 1 of such fixed plethysm or fixed symmetry generating functions for the group SU(2) as well as for SU(3) with $p=2$.

In this paper we present a systematic procedure for the construction of fixed symmetry generating functions for Lie groups. These are most conveniently given as linear combinations of new "fixed class" generating functions. In Sec. II we describe the construction of fixed class and fixed symmetry generating functions. In Sec. III we give some examples for the groups SU(2), SU(3), and SO(5). Section IV contains a discussion and some concluding remarks.

## II. CONSTRUCTION OF THE GENERATING FUNCTIONS

A fixed symmetry generator enumerates all irreducible representations $R_b$ of a compact Lie group $G$ contained in the part of the direct product of $p$ copies of any irreducible representation $R_a$ of $G$ which has a given exchange symmetry under the symmetric group $S_p$. Specifically, its expansion coefficients are the coefficients $n_{ab}^{(\lambda)}$ which arise in the decomposition

$$R_a \otimes R_a \otimes \cdots \otimes R_a = \bigoplus_{b,\lambda} n_{ab}^{(\lambda)} (\lambda) \times R_b, \tag{1}$$

where there are $p$ factors in the product on the left-hand side and where the $(\lambda)$ are the irreducible representations of $S_p$. The required fixed symmetry generating function has the expansion

$$\phi_{(\lambda)}(A,B) = \sum_{a,b} n_{ab}^{(\lambda)} A^a B^b. \tag{2}$$

We are suppressing subscripts on $A$, $B$, $a$, and $b$. If $G$ has rank $l$, the symbol $A^a$, for example, stands for $A_1^{a_1} A_2^{a_2} \cdots A_l^{a_l}$.

The most straightforward construction of $\phi_{(\lambda)}$ exploits the properties of group characters.[2] For the right-hand side of Eq. (1) we find the character in the class $\rho$ of $S_p$ to be

$$\sum_{b,\lambda} n_{ab}^{(\lambda)} X_\rho^{(\lambda)} \chi_b(\eta), \tag{3}$$

where $X_\rho^{(\lambda)}$ is the character of the irreducible representation $(\lambda)$ of $S_p$ in the class $\rho$ and $\chi_b(\eta)$ is the character of the irreducible representation $b$ of $G$ in the class labeled by $(\eta_1,\ldots,\eta_l)$. The character of the left-hand side of Eq. (1) in the class $\rho = (1^\alpha 2^\beta 3^\gamma \cdots)$ is

$$s_{a\rho}(\eta) = \chi_a^\alpha(\eta)\chi_a^\beta(\eta^2)\chi_a^\gamma(\eta^3)\cdots. \tag{4}$$

Using the orthogonality property of the characters $X_\rho^{(\lambda)}$ we find

$$\sum_b n_{ab}^{(\lambda)} \chi_b(\eta) = \sum_\rho s_{a\rho}(\eta) X_\rho^{(\lambda)} h_\rho (h)^{-1} = \{\lambda\}_a, \tag{5}$$

where $h_\rho$ is the order of the class $\rho$ and $h = p!$ is the order of $S_p$. Here $\{\lambda\}_a$ is known as a Schur function or $S$-function.[2]

In the basis in which the highest weight of the irreducible representation $R_a$ of $G$ is $(a_1,\ldots,a_l)$, the character is given by the Weyl formula[3]

$$\chi_a(\eta) = \xi_a(\eta)/\xi_0(\eta), \tag{6}$$

where $\xi_a(\eta)$ is the characteristic for the irreducible representation $R_a$:

$$\xi_a(\eta) = \sum_S (-1)^{w_S} \prod_{k=1}^{l} (S\eta_k)^{a_k+1}. \tag{7}$$

The sum in Eq. (7) is over the elements $\{S\}$ of the Weyl group of $G$. Each element $S$ can be written[4,5] as the product of $w_S$ generators $S_i$ $(i=1,\ldots,l)$. The action of $S$ on $\eta_k$, $S\eta_k$, is determined from the action of the generators

$$S_k \eta_k = \eta_k \prod_{i=1}^{l} \eta_i^{-A_{ik}}, \quad S_j \eta_k = \eta_k \quad (j \neq k), \tag{8}$$

where $A_{ik}$ is the Cartan matrix of $G$. In this basis it is straightforward[6] to solve equation (5) for $n_{ab}^{(\lambda)}$ to obtain

$$n_{ab}^{(\lambda)} = \sum_\rho X_\rho^{(\lambda)} \left(\frac{h_\rho}{h}\right) \xi_0(\eta) \times \prod_{k=1}^{l} \eta_k^{-1} s_{a\rho}(\eta) \Big|_{\mathrm{EX}(\eta)=b}, \tag{9}$$

  

where the last instruction tells us to keep only those terms whose $\eta_i$ exponents are $b_i$.

In order to obtain the generating function $\phi_{(\lambda)}(A,B)$ it is most convenient to construct first the fixed class generator

$$\psi_\rho(A,\eta) = \xi_0(\eta) \prod_{k=1}^{l} \eta_k^{-1} S_\rho(A,\eta) \Big|_{\mathrm{EX}(\eta)>0}, \tag{10}$$

where the expansion coefficients of $S_\rho(A,\eta)$ are $s_{a\rho}(\eta)$. Then we find

$$\phi_{(\lambda)}(A,B) = \sum_\rho C_\rho^{(\lambda)} \psi_\rho(A,B), \tag{11}$$

where $C_\rho^{(\lambda)} = X_\rho^{(\lambda)} h_\rho / h$. These coefficients are given in Table I for $p \leqslant 6$. If the partition $\rho$ contains $N$ cycles, with the $i$th cycle having length $n_i$, then the compound character generator $S_\rho$ is given by

$$S_\rho(A,\eta) = \Xi_\rho(A,\eta)/\Xi_\rho(0,\eta), \tag{12}$$

with

$$\Xi_\rho(A,\eta) = \sum_{S_1 \cdots S_N} \prod_{i=1}^{N} (-1)^{w_{S_i}} \prod_{k=1}^{l} \prod_{i=1}^{N} (S_i \eta_k^{n_i}) \times \Big[ 1 - A_k \prod_{i=1}^{N} (S_i \eta_k^{n_i}) \Big]^{-1}, \tag{13}$$

where the sum is over $N$ sets $\{S_i\}$ of elements of the Weyl reflection group [the $S_i$ here are not to be confused with the generators appearing in Eq. (8)]. The denominator of Eq. (12) can be written more simply as

$$\Xi_\rho(0,\eta) = \prod_{i=1}^{N} \xi_0(\eta^{n_i}). \tag{14}$$

The procedure outlined above has proven to be the simplest one for the construction of fixed class generators. There is, however, another procedure which makes use of the generating function for the Clebsch–Gordan series to combine two fixed class generators to produce a third. If $\rho_1$ and $\rho_2$ are

TABLE I. Coefficients $C_\rho^{(\lambda)}$ connecting fixed class and fixed symmetry generators.

| $\lambda$ $\rho$: | $(1^2)$ | $(2)$ |
|---|---|---|
| $(2)$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $(1^2)$ | $\frac{1}{2}$ | $-\frac{1}{2}$ |

| $\lambda$ $\rho$: | $(1^3)$ | $(12)$ | $(3)$ |
|---|---|---|---|
| $(3)$ | $\frac{1}{6}$ | $\frac{1}{2}$ | $\frac{1}{3}$ |
| $(2\,1)$ | $\frac{1}{3}$ | $0$ | $-\frac{1}{3}$ |
| $(1^3)$ | $\frac{1}{6}$ | $-\frac{1}{2}$ | $\frac{1}{3}$ |

| $\lambda$ $\rho$: | $(1^4)$ | $(1^2 2)$ | $(13)$ | $(2^2)$ | $(4)$ |
|---|---|---|---|---|---|
| $(4)$ | $\frac{1}{24}$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{1}{8}$ | $\frac{1}{4}$ |
| $(3\,1)$ | $\frac{1}{8}$ | $\frac{1}{4}$ | $0$ | $-\frac{1}{8}$ | $-\frac{1}{4}$ |
| $(2^2)$ | $\frac{1}{12}$ | $0$ | $-\frac{1}{3}$ | $\frac{1}{4}$ | $0$ |
| $(2\,1^2)$ | $\frac{1}{8}$ | $-\frac{1}{4}$ | $0$ | $-\frac{1}{8}$ | $\frac{1}{4}$ |
| $(1^4)$ | $\frac{1}{24}$ | $-\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{1}{8}$ | $-\frac{1}{4}$ |

| $\lambda\rho$: | $(1^5)$ | $(1^3 2)$ | $(1^2 3)$ | $(12^2)$ | $(14)$ | $(23)$ | $(5)$ |
|---|---|---|---|---|---|---|---|
| $(5)$ | $\frac{1}{120}$ | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{8}$ | $\frac{1}{4}$ | $\frac{1}{6}$ | $\frac{1}{5}$ |
| $(4\,1)$ | $\frac{1}{30}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $0$ | $0$ | $-\frac{1}{6}$ | $-\frac{1}{5}$ |
| $(3\,2)$ | $\frac{1}{24}$ | $\frac{1}{12}$ | $-\frac{1}{6}$ | $\frac{1}{8}$ | $-\frac{1}{4}$ | $\frac{1}{6}$ | $0$ |
| $(3\,1^2)$ | $\frac{1}{20}$ | $0$ | $0$ | $-\frac{1}{4}$ | $0$ | $0$ | $\frac{1}{5}$ |
| $(2^2 1)$ | $\frac{1}{24}$ | $-\frac{1}{12}$ | $-\frac{1}{6}$ | $\frac{1}{8}$ | $\frac{1}{4}$ | $-\frac{1}{6}$ | $0$ |
| $(2\,1^3)$ | $\frac{1}{30}$ | $-\frac{1}{6}$ | $\frac{1}{6}$ | $0$ | $0$ | $\frac{1}{6}$ | $-\frac{1}{5}$ |
| $(1^5)$ | $\frac{1}{120}$ | $-\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{8}$ | $-\frac{1}{4}$ | $-\frac{1}{6}$ | $\frac{1}{5}$ |

| $\lambda$ $\rho$: | $(1^6)$ | $(1^4 2)$ | $(1^3 3)$ | $(1^2 2^2)$ | $(1^2 4)$ | $(123)$ | $(15)$ | $(2^3)$ | $(24)$ | $(3^2)$ | $(6)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $(6)$ | $\frac{1}{720}$ | $\frac{1}{48}$ | $\frac{1}{18}$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $\frac{1}{6}$ | $\frac{1}{5}$ | $\frac{1}{48}$ | $\frac{1}{8}$ | $\frac{1}{18}$ | $\frac{1}{6}$ |
| $(5\,1)$ | $\frac{1}{144}$ | $\frac{1}{16}$ | $\frac{1}{9}$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $0$ | $0$ | $-\frac{1}{48}$ | $-\frac{1}{8}$ | $-\frac{1}{18}$ | $-\frac{1}{6}$ |
| $(4\,2)$ | $\frac{1}{80}$ | $\frac{1}{16}$ | $0$ | $\frac{1}{16}$ | $-\frac{1}{8}$ | $0$ | $-\frac{1}{5}$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $0$ | $0$ |
| $(4\,1^2)$ | $\frac{1}{72}$ | $\frac{1}{24}$ | $\frac{1}{18}$ | $-\frac{1}{8}$ | $0$ | $-\frac{1}{6}$ | $0$ | $-\frac{1}{24}$ | $0$ | $\frac{1}{18}$ | $\frac{1}{6}$ |
| $(3^2)$ | $\frac{1}{144}$ | $\frac{1}{48}$ | $-\frac{1}{18}$ | $\frac{1}{16}$ | $-\frac{1}{8}$ | $\frac{1}{6}$ | $0$ | $-\frac{1}{16}$ | $-\frac{1}{8}$ | $\frac{1}{9}$ | $0$ |
| $(3\,2\,1)$ | $\frac{1}{45}$ | $0$ | $-\frac{1}{9}$ | $0$ | $0$ | $0$ | $\frac{1}{5}$ | $0$ | $0$ | $-\frac{1}{9}$ | $0$ |
| $(3\,1^3)$ | $\frac{1}{72}$ | $-\frac{1}{24}$ | $\frac{1}{18}$ | $-\frac{1}{8}$ | $0$ | $\frac{1}{6}$ | $0$ | $\frac{1}{24}$ | $0$ | $\frac{1}{18}$ | $-\frac{1}{16}$ |
| $(2^3)$ | $\frac{1}{144}$ | $-\frac{1}{48}$ | $-\frac{1}{18}$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $-\frac{1}{6}$ | $0$ | $\frac{1}{16}$ | $-\frac{1}{8}$ | $\frac{1}{9}$ | $0$ |
| $(2^2\,1^2)$ | $\frac{1}{80}$ | $-\frac{1}{16}$ | $0$ | $\frac{1}{16}$ | $\frac{1}{8}$ | $0$ | $-\frac{1}{5}$ | $-\frac{1}{16}$ | $\frac{1}{8}$ | $0$ | $0$ |
| $(2\,1^4)$ | $\frac{1}{144}$ | $-\frac{1}{16}$ | $\frac{1}{9}$ | $\frac{1}{16}$ | $-\frac{1}{8}$ | $0$ | $0$ | $\frac{1}{48}$ | $-\frac{1}{8}$ | $-\frac{1}{18}$ | $\frac{1}{6}$ |
| $(1^6)$ | $\frac{1}{720}$ | $-\frac{1}{48}$ | $\frac{1}{18}$ | $\frac{1}{16}$ | $-\frac{1}{8}$ | $-\frac{1}{6}$ | $\frac{1}{5}$ | $-\frac{1}{48}$ | $\frac{1}{8}$ | $\frac{1}{18}$ | $-\frac{1}{6}$ |

two partitions containing, respectively, $a_i$ and $b_i$ cycles of length $i$, then their product $\rho = \rho_1\rho_2$ will contain $a_i + b_i$ cycles of length $i$. We can compute the resulting generating function for the class $\rho$ from

$$\psi_\rho(A,B) = \psi_{\rho_1}(Ap,q)\psi_{\rho_2}(p^{-1},r) \times G(q^{-1},r^{-1};B)|_{\mathrm{EX}(p,q,r)=0}, \quad (15)$$

where, as usual, each letter stands for $l$ variables. The Clebsch–Gordan generator $G$ plays the role of a metric, while the $\mathrm{EX}(p) = 0$ operation ensures that the representations in the product (1) are all the same. By using Eq. (15) we can build up fixed class generating functions from basic units which correspond to partitions $(n)$ containing a single cycle of length $n$. The basic units have fixed class generating functions

$$\psi_n(A,\eta) = \xi_0(\eta)\prod_{k=1}^{l}\eta_k^{-1}X(A,\eta^n)\Big|_{\mathrm{EX}(\eta)>0}, \quad (16)$$

where $X(A,\eta)$ is the group character generator.

It is easy to see that $\psi_n(A,B)$ is also the branching rule generator for the subjoining to the group representations of representations of the group with weights stretched by a factor of $n$.[7] For $n$ sufficiently large we can construct general formulas for $\psi_n(A,B)$. For $G = \mathrm{SU}(2)$ we find, for $n \geqslant 2$,

$$\psi_n(A,B) = (1 - AB^{n-2})/[(1-A^2)(1-AB^n)]. \quad (17)$$

For $G = \mathrm{SU}(3)$ we find, for $n \geqslant 3$,

$$\begin{aligned}\psi_n(A,B;a,b) &= [(1-Aa^n)(1-Bb^n)(1-AB)]^{-1}\\ &\times\{[A^2b^n - A^2ab^{n-2} + A^2b^{n-3} - A^3a^{n-2}b^{n+1} + A^3a^{n-3}b^n - A^3(ab)^{n-2}]/[(1-A^3)(1-A^2b^n)]\\ &+ [B^2a^n - B^2a^{n-2}b + B^2a^{n-3} - B^3a^{n+1}b^{n-2} + B^3a^nb^{n-3} - B^3(ab)^{n-2}]/[(1-B^3)(1-B^3a^n)]\\ &+ (1+AB+A^2B^2)[1 - Bab^{n-2} + Bb^{n-3} - Aa^{n-2}b + Aa^{n-3} - AB(ab)^{n-2}]/[(1-A^3)(1-B^3)]\},\end{aligned} \quad (18)$$

where we have used the variables $(A,B,a,b)$ instead of $(A_1,A_2,B_1,B_2)$ in order to simplify the notation. For $G = \mathrm{SO}(5)$ we find, for $n \geqslant 4$,

$$\begin{aligned}\psi_n(A,B;a,b) &= [(1-A^2)(1-B)(1-Aa^n)(1-Bb^n)]^{-1}\\ &\times\{(1+A^2B)[1 - Aa^{n-2}b + Aa^{n-4}b - Aa^{n-4} - Ba^2b^{n-2}\\ &+ Ba^2b^{n-3} - Bb^{n-3} + AB(ab)^{n-2}]/[(1-A^2)(1-B^2)]\\ &+ A^2[-b^{n-3}(1-b^3) + a^2b^{n-3}(1-b) + A(ab)^{n-2}(1-b^3) - Aa^{n-4}b^n(1-b)]/[(1-A^2)(1-A^2b^n)]\\ &+ (AB + B^2a^n)[ba^{n-4}(1-a^2) - a^{n-4}(1-a^4) + B(ab)^{n-2}(1-a^4) - Ba^nb^{n-3}(1-a^2)]/[(1-B^2)(1-B^2a^{2n}]\},\end{aligned} \quad (19)$$

where (10) and (01) are, respectively, the four- and five-dimensional irreducible representations of SO(5).

## III. EXAMPLES OF FIXED CLASS GENERATORS

In this section we collect examples of fixed class generating functions which, with Eq. (11), can be used to obtain fixed symmetry generators. The results presented in Table II for SU(2) can be used to construct the fixed symmetry generating functions for $p \leqslant 6$. For $G = \mathrm{SU}(3)$ we have the following results which can be used to construct the $p = 2$ fixed symmetry generators given in Ref. 1:

$$\psi_{(1^2)}(A,B;a,b) = (1 + ABab)[(1-Aa^2)\times(1-Bb^2)(1-AB)(1-Ab)(1-Ba)]^{-1}, \quad (20)$$

$$\psi_{(2)}(A,B;a,b) = (1 - ABab)[(1-Aa^2)\times(1-Bb^2)(1-AB)(1+Ab)(1+Ba)]^{-1}. \quad (21)$$

The generating functions $\phi_{(2)}$ and $\phi_{(1^2)}$ for symmetric and antisymmetric combinations of SU(3) irreducible representations are, respectively, $\frac{1}{2}(\psi_{(1^2)} + \psi_{(2)})$ and $\frac{1}{2}(\psi_{(1^2)} - \psi_{(2)})$.

For SU(3) the fixed class generating functions for the product of three copies of an irreducible representation are

$$\begin{aligned}\psi_{(1^3)}(A,B;a,b) &= [(1-Aa^3)(1-Bb^3)(1-AB)(1-A)(1-B)]^{-1}\\ &\times[(1 + ABa^2b^2 + 2ABb^3 + 2ABa^2b^2 + 3AB^2b^3 + 2AB^2a^2b^2 + A^2B^2a^3b^3)/(1-Aab)(1-Bab)(1-ABb^3)\\ &+ (ABa^3 + A^2B^2a^5b^2 + 2ABa^3 + 2ABa^2b^2 + 3A^2Ba^3 + 2A^2Ba^2b^2 + A^2B^2a^3b^3)/(1-Aab)(1-Bab)(1-ABa^3)\\ &+ (3A^2B^2a^2b^2 + 3A^3B^3a^3b^3 + A^4B^4a^4b^4 + A^2B^3ab^4 + A^3B^3a^5b^2)/(1-Bab)(1-A^2Ba^3)(1-AB^2b^3)\\ &+ (3A^3B^2a^3b^3 + 3A^4B^3a^4b^4 + A^5B^4a^5b^5 + A^3B^2a^4b + A^3B^3a^2b^5)/(1-Aab)(1-A^2Ba^3)(1-AB^2b^3)\\ &+ (B^2a^3 + Bab + B^2a^2b^2)(1 + 2A + 2ABa^3 + A^2Ba^3)/(1-Bab)(1-ABa^3)(1-B^2a^3)\\ &+ (A^2b^3 + Aab + A^2a^2b^2)(1 + 2B + 2ABb^3 + AB^2b^3)/(1-Aab)(1-ABb^3)(1-A^2b^3)\\ &+ (3A^2B^4b^6 + A^2B^3ab^4 + A^2B^3a^2b^5 + A^2B^4a^2b^5)/(1-Bab)(1-ABb^3)(1-AB^2b^3)\\ &+ (3A^4B^2a^6 + A^3B^2a^4b + A^3B^2a^5b^2 + A^4B^2a^5b^2)/(1-Aab)(1-ABa^3)(1-A^2Ba^3)\\ &+ (A^3B^2a^4b + A^3B^3a^5b^2 + A^4B^3a^7b)/(1-Bab)(1-ABa^3)(1-A^2Ba^3)\\ &+ (A^2B^3ab^4 + A^3B^3a^2b^5 + A^3B^4ab^7)/(1-Aab)(1-ABb^3)(1-AB^2b^3)],\end{aligned} \quad (22)$$

TABLE II. Fixed class generating functions for SU(2).

| $\rho$: | $\psi_\rho(A,B)$: |
|---|---|
| (1) | $1/(1-AB)$ |
| $(1^2)$ | $1/(1-AB^2)(1-A)$ |
| (2) | $1/(1-AB^2)(1+A)$ |
| $(1^3)$ | $(1+AB+A^2B^2)/(1-AB^3)(1-AB)(1-A^2)$ |
| (12) | $(1-AB+A^2B^2)/(1-AB^3)(1-AB)(1+A^2)$ |
| (3) | $(1-AB)/(1-AB^3)(1-A^2)$ |
| $(1^4)$ | $(1+2AB^2+A^2B^4)/(1-AB^4)(1-AB^2)(1-A)^2$ |
| $(1^22)$ | $(1+A^2B^4)/(1-AB^4)(1-AB^2)(1-A^2)$ |
| (13) | $(1-A)(1-AB^2+A^2B^4)/(1-AB^4)(1-AB^2)(1-A^3)$ |
| $(2^2)$ | $(1-AB^2)/(1-AB^4)(1-A)^2$ |
| (4) | $(1-AB^2)/(1-AB^4)(1-A^2)$ |
| $(1^5)$ | $[(1+3A^2+A^4)+A(4+2A^2-A^4)B+A^2(10-5A^2)B^2+A(3+4A^2-7A^4)B^3$ $+A^2(7-4A^2-3A^4)B^4+A^3(5-10A^2)B^5+A^2(1-2A^2-4A^4)B^6-A^3(1+3A^2+A^4)B^7]$ $\times[(1-AB^5)(1-AB^3)(1-AB)(1-A^2)^3]^{-1}$ |
| $(1^32)$ | $[(1-A^2+A^4)+A^3(2-A^2)B+A^2(2-A^2)B^2+A(1-A^4)B^3+A^2(1-A^4)B^4$ $+A^3(1-2A^2)B^5+A^2(1-2A^2)B^6-A^3(1-A^2+A^4)B^7]$ $\times[(1-AB^5)(1-AB^3)(1-AB)(1-A^2)(1-A^4)]^{-1}$ |
| $(1^23)$ | $[(1+A^4)-A(2+A^2+A^4)B+A^2(1+A^2)B^2+A^3(1-A^2)B^3+A^2(1-A^2)B^4$ $-A^3(1+A^2)B^5+A^2(1+A^2+2A^4)B^6-A^3(1+A^4)B^7]$ $\times[(1-AB^5)(1-AB^3)(1-AB)(1-A^6)]^{-1}$ |
| $(12^2)$ | $[(1-A^2+A^4)+A^3(2-A^2)B+A^2(2-A^2)B^2-A(1-A^4)B^3-A^2(1-A^4)B^4$ $+A^3(1-2A^2)B^5+A^2(1-2A^2)B^6-A^3(1-A^2+A^4)B^7]$ $\times[(1-AB^5)(1-AB^3)(1-AB)(1+A^2)(1-A^4)]^{-1}$ |
| (14) | $[(1+A^2+A^4)-A(1+A^2)B+A^2B^2-A(1+A^2+A^4)B^3+A^4B^4-A^3(1+A^2)B^5$ $+A^2(1+A^2+A^4)B^6][(1-AB^5)(1-AB^3)(1+A^2)(1+A^4)]^{-1}$ |
| (23) | $(1-A^2)[(1+A^2)-A^3B-A^2B^2-AB^3+A^2(1+A^2)B^4]/(1-AB^5)(1-AB)(1-A^6)$ |
| (5) | $(1-AB^3)/(1-AB^5)(1-A^2)$ |
| $(1^6)$ | $[(1+A+A^2)+A(8-A-A^2)B^2+A(4+10A-11A^2)B^4+A^2(11-10A-4A^2)B^6$ $+A^2(1+A+8A^2)B^8-A^3(1+A+A^2)B^{10}]/(1-AB^6)(1-AB^4)(1-AB^2)(1-A)^4$ |
| $(1^42)$ | $[(1-A+A^2)+A(2+A-A^2)B^2+A(2+2A-3A^2)B^4+A^2(3-2A-2A^2)B^6$ $+A^2(1-A-2A^2)B^8-A^3(1-A+A^2)B^{10}]/(1-AB^6)(1-AB^4)(1-AB^2)(1-A)^3(1+A)$ |
| $(1^23)$ | $[(1-A)-A(1-A)B^2+A(1+2A)B^4+A^2(2+A)B^6+A^2(1-A)B^8-A^3(1-A)B^{10}]$ $\times[(1-AB^6)(1-AB^4)(1-AB^2)(1-A^3)]^{-1}$ |
| $(1^22^2)$ | $[(1+A+A^2)-A^2(1+A)B^2+A^2(2+A)B^4-A^2(1+2A)B^6+A^2(1+A)B^8$ $-A^3(1+A+A^2)B^{10}]/(1-AB^6)(1-AB^4)(1-AB^2)(1-A^2)^2$ |
| $(1^24)$ | $[(1-A+A^2)-A(2-A+A^2)B^2+A^2(2-A)B^4+A^4(1-2A)B^6$ $+A^2(1-A+2A^2)B^8-A^3(1-A+A^2)B^{10}]/(1-AB^6)(1-AB^4)(1-AB^2)(1-A^4)$ |
| (123) | $[(1+A)-A(1+A)B^2-AB^4+A^3B^6+A^2(1+A)B^8-A^3(1+A)B^{10}]$ $\times[(1-AB^6)(1-AB^4)(1-AB^2)(1-A^3)]^{-1}$ |
| (15) | $[(1-A^3)-A(1-A)B^2-A(1-A^3)B^4-A^3(1-A)B^6+A^2(1-A^3)B^8]$ $\times[(1-AB^6)(1-AB^4)(1-A^5)]^{-1}$ |
| $(2^3)$ | $[(1-A+A^2)+A(2+A-A^2)B^2-A(1-A-2A^2)B^4+A^2(1-A+A^2)B^6]$ $\times[(1-AB^6)(1-AB^2)(1-A^2)(1+A)^2]^{-1}$ |
| (24) | $[(1+A+A^2)-A^2(1+A)B^2-A(1+A)B^4+A^2(1+A+A^2)B^6]$ $\times[(1-AB^6)(1-AB^2)(1+A^2)(1+A)^2]^{-1}$ |
| $(3^2)$ | $(1-AB^4)/(1-AB^6)(1-A)^2$ |
| (6) | $(1-AB^4)/(1-AB^6)(1-A^2)$ |

$$\begin{aligned}
&\psi_{(12)}(A,B;a,b)\\
&\quad=[(1-Aa^3)(1-Bb^3)(1+AB)(1+A)(1+B)]^{-1}\\
&\qquad\times[(1+A^2B^2a^2b^2+A^4B^4a^4b^4)/(1-ABb^3)(1-A^2Ba^3)(1-AB^2b^3)\\
&\qquad+ABa^3(1+A^2B^2a^2b^2+A^4B^4a^4b^4)/(1-ABa^3)(1-A^2Ba^3)(1-AB^2b^3)\\
&\qquad+AB^2ab^4(1+AB)/(1-Bab)(1-ABb^3)(1-AB^2b^3)\\
&\qquad+A^2Ba^4b(1+AB)/(1-Aab)(1-ABa^3)(1-A^2Ba^3)\\
&\qquad+AB^3ab^4(1+A^2Ba^3)/(1-Bab)(1-ABa^3)(1-AB^2b^3)\\
&\qquad+A^3Ba^4b(1+AB^2b^3)/(1-Aab)(1-ABb^3)(1-A^2Ba^3)\\
&\qquad+(1+A^2Ba^3)(-B^2a^3+B^3a^4b+B^2a^2b^2)/(1-Bab)(1-ABa^3)(1+B^2a^3)\\
&\qquad+(1+AB^2b^3)(-A^2b^3+A^3ab^4+A^2a^2b^2)/(1-Aab)(1-ABb^3)(1+A^2b^3)],
\end{aligned}\tag{23}$$

$$\begin{aligned}
&\psi_{(3)}(A,B;a,b)\\
&\quad=[(1-Aa^3)(1-Bb^3)(1-AB)]^{-1}[(1-Bab)/(1-B)(1-B^2a^3)+A^2b^3(1-Aab)/(1-A)(1-A^2b^3)\\
&\qquad+A(1-ab)/(1-A)(1-B)].
\end{aligned}\tag{24}$$

For $G = \mathrm{SO}(5)$ the fixed class generating functions for $p = 2$ are

$$\psi_{(1^2)}(A,B;a,b) = (1 + ABa^2b)[(1 - Aa^2)(1 - Bb^2)(1 - A)(1 - B)(1 - Ab)(1 - Ba^2)]^{-1}, \tag{25}$$

and

$$\Psi_{(2)}(A,B;a,b) = (1 - ABa^2b)[(1 - Aa^2)(1 - Bb^2)(1 + A)(1 - B)(1 + Ab)(1 + Ba^2)]^{-1}. \tag{26}$$

For $p = 3$ we find

$$\psi_{(1^3)}(A,B;0,0) = (1 + 2A^2B + 2A^2B^2 + A^4B^3)[(1 - A^2)(1 - B^2)(1 - A^2B)(1 - A^2B^2)]^{-1}, \tag{27}$$

$$\psi_{(12)}(A,B;0,0) = (1 - A^4B^3)[(1 + A^2)(1 - B^2)(1 + A^2B)(1 + A^2B^2)]^{-1}, \tag{28}$$

$$\psi_{(3)}(A,B;a,b) = [(1 + Ba^2)(1 + ABa^3)(1 - Ba^2b)/(1 - B^2a^6) + (1 + B)(A^2b^3 - Aab)/(1 - A^2b^3)][(1 - A^2)(1 - B^2)(1 - Aa^3)(1 - Bb^3)]^{-1}. \tag{29}$$

For the classes $(1^3)$ and $(12)$ we have given the generating functions enumerating only the scalars contained in the direct product. The full result for the class (3) is given because it is the generating function for branching rules for the subjoining $\mathrm{SO}(5) > \mathrm{SO}(5)$ corresponding to dilation of weight space by a factor of 3; it reduces to $[(1 - A^2)(1 - B^2)]^{-1}$ in the scalar limit.

## IV. CONCLUDING REMARKS

The construction of fixed symmetry generating functions follows the usual procedure. First we construct a generating function for the compound character and then we project out the corresponding irreducible representations. The use of Eq. (12) for the compound character generator simplifies the construction in that the $\mathrm{EX}(\eta) \geqslant 0$ operation is performed at an early stage when there is a large number of relatively simple terms. Each term in the result of the projection will contain spurious poles which must cancel out when the terms are combined. This provides a guide to the manipulations needed to combine the terms to produce the final result.

The introduction of fixed class generating functions not only simplifies the construction but also allows us to present the results in a more compact form. The fixed class generators for a given $p$ have, in general, different denominator factors so that their combinations, written over common denominators, have far larger numerators. For practical purposes of determining individual plethysms, it is simpler to isolate the desired terms in the expansion of the fixed class generators and then to combine these with the coefficients $C_\rho^{(\lambda)}$.

An interesting observation arises from examining the three box mixed symmetry $\lambda = (21)$. The SU(2) and SU(3) generating functions with this symmetry contain no terms in their expansions which are independent of $B_i$. This means that the part of the direct product of three copies of any irrep of SU(2) or SU(3) with this symmetry has no scalar component. In fact, it can be shown that SU(2) and SU(3) are the only simple compact groups for which this is true. For example in the case of SO(5) the scalars contained in this plethysm are enumerated by the generating function

$$\phi_{(21)}(A,B;0,0) = A^2B/(1 - A^2)(1 - B)(1 - A^2B)(1 - A^2B^2). \tag{30}$$

This was constructed with the help of the generating functions of Sec. III.


## ACKNOWLEDGMENTS

We are grateful to J. Patera for helpful discussions. One of us (RG) would like to thank the Aspen Center for Physics, where a portion of this work was done, for their hospitality.

Work supported in part by the National Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.



[1] J. Patera and R. T. Sharp, J. Math. Phys. **22**, 261 (1981).
[2] D. E. Littlewood, *The Theory of Group Characters* (Clarendon, Oxford, 1950), 2nd ed.
[3] H. Weyl, Math. Z. **23**, 271 (1925), reprinted in H. Weyl, *Selecta* (Birkhauser-Verlag, Basel, 1956).
[4] C. T. Benson and L. C. Grove, *Finite Reflection Groups* (Bogden and Quigley, New York, 1971).
[5] R. W. Gaskell, J. Math. Phys. **24**, 2379 (1983).
[6] C. Bodine and R. W. Gaskell, J. Math. Phys. **23**, 2217 (1982).
[7] J. Patera, R. T. Sharp, and R. Slansky, J. Math. Phys. **21**, 2335 (1980).

# Comments on superposition rules for nonlinear coupled first-order differential equations

P. Winternitz
*Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec H3C 3J7, Canada*

(Received 1 November 1983; accepted for publication 16 December 1983)

Some comments are made on the classification of finite-dimensional subalgebras of the Lie algebra of vector fields in $n$ variables and of the related nonlinear ordinary differential equations with superposition principles. In particular for $n = 2$ a very natural requirement of indecomposability implies that only two types of equations need be considered.

PACS numbers: 02.20.Sv, 02.30.Hq, 11.10.Lm, 11.30.Na

A recent publication[1] has been devoted to the determination of all pairs of ordinary real differential equations of the type

$$\dot{x}(t) = \sum_{i=1}^{l} Z_i(t)\xi_i(x,y), \quad \dot{y}(t) = \sum_{i=1}^{l} Z_i(t)\eta_i(x,y) \tag{1}$$

such that (i) the system (1) allows a superposition principle, i.e., the general solution of (1) can be written as a function of a finite number of particular solutions and of two significant constants; and (ii) the functions $\xi_i(x,y)$ and $\eta_i(x,y)$ are polynomials of at most second order in $x$ and $y$.

In view of a classical theorem due to Lie[2] the construction of all equations of type (1) with superposition principles is equivalent to the construction of all finite-dimensional Lie algebras that can be realized in terms of vector fields in two variables:

$$\hat{Y}_i = \xi_i(x,y)\frac{\partial}{\partial x} + \eta_i(x,y)\frac{\partial}{\partial y}. \tag{2}$$

The results of Ref. 1 thus amount to a classification of such algebras with the restriction that the coefficients in (2) should be polynomials of at most second order.

The purpose of this short note is twofold.

a) We correct the result reported in Ref. 1 by recalling the work of Lie (1880).[3,4]

b) In view of the increased interest in this area we wish to make some comments which summarize the proper formulation of the mathematical questions involved in classifying nonlinear ordinary differential equations with superposition principles.

*Comment 1:* When solving a classification problem several basic rules should be followed.

a) False generality should be avoided, i.e., the objects should be classified into equivalence classes under some well-defined equivalence relation. Each class should be represented precisely once in a representative list.

b) Triviality should be avoided, i.e., it should be decided beforehand which objects are of interest and then only these should be classified.

Thus, if we are interested in Eqs. (1), it is natural to classify the Lie algebras (2) under local changes of variables. Two sets of Eqs. (1) are then equivalent if they can be transformed into each other by a change of dependent variables $u = \phi(x,y)$, $v = \psi(x,y)$, where $x$, $y$, $u$, and $v$ are real, and $\phi$ and $\psi$ are sufficiently smooth functions, such that the inverse transformation is locally well defined. Such a classification of finite-dimensional Lie algebras that act on two-dimensional manifolds was performed by Lie himself in a different context, without any restriction on the form of the coefficients (see Refs. 3 and 4 and Hermann's comments in Ref. 4 for an exposition of Lie's results in modern terms).

Restricting ourselves to real variables $x$ and $y$ and to quadratic polynomials as in Ref. 1, we can extract the following very simple results directly from Lie's list (without going into the extensive algebraic calculations of Ref. 1).

*Proposition:* Any finite-dimensional Lie algebra that can be realized in terms of vector fields in two variables with polynomial coefficients of at most second order is equivalent (under local changes of variables) to one of the following Lie algebras, or one of their subalgebras:

(i) sl(3,$\mathbb{R}$):

$$\{\partial_x, \partial_y, x\,\partial_x, y\,\partial_x, x\partial_y, y\,\partial_y, x(x\,\partial_x + y\,\partial_y), y(x\,\partial_x + y\,\partial_y)\}; \tag{3}$$

(ii) o(3,1):

$$\{\partial_x, \partial_y, x\,\partial_x + y\,\partial_y, x\,\partial_y - y\,\partial_x, (x^2 - y^2)\partial_x + 2xy\,\partial_y, 2xy\,\partial_x - (x^2 - y^2)\partial y\}; \tag{4}$$

(iii) o(2,2) ~ o(2,1) ⊕ o(2,1):

$$\{\partial_x, \partial_y, x\,\partial_x + y\,\partial_y, x\,\partial_y + y\,\partial_x, (x^2 + y^2)\partial_x + 2xy\,\partial_y, 2xy\partial_x + (x^2 + y^2)\partial_y\}, \tag{5a}$$

or equivalently

$$\{\partial_u, u\,\partial_u, u^2\,\partial_u\} \oplus \{\partial_v, v\,\partial_v, v^2\,\partial_v\}, \tag{5b}$$

$$u = x + y, \quad v = x - y;$$

(iv) gl(2,$\mathbb{R}$)$\oplus\!\!\!\!+\, t_3$

$$[\{\partial_x, x\,\partial_x + y\,\partial_y, x^2\,\partial_x + 2xy\,\partial_y\} \oplus \{y\,\partial_y\}] \oplus\!\!\!\!+ \{\partial_y, x\,\partial_y, x^2\,\partial_y\}. \tag{6}$$

The equations (1) for the Lie algebras sl(3,$\mathbb{R}$) and o(3,1) are special cases of projective and conformal Riccati equations.[5,6] Superposition formulas for these equations, as well as for the more general matrix Riccati equations[7] have been obtained for the general case of sl($n$,$\mathbb{R}$) and o($p$,$q$) algebras.[5–7] The special cases of $n = 3$ and $p + q = 4$ do not need a separate treatment. The equations corresponding to algebras (5) and (6) are "trivial" in the following sense. For algebra (5b) we obtain two uncoupled scalar Riccati equations with independent superposition formulas for $u$ and $v$. For algebra (6) we obtain a scalar Riccati equation in $x$ and an equation in $y$ that turns into a linear scalar equation, once $x(t)$ is substituted into it. We thus have a Riccati superposition formula for $x(t)$ and a subsequent linear one for $y(t)$.

 0022-2488/84/072149-02$02.50  

Subalgebras of the algebras (3), ..., (6) lead to the same types of equations with some of the coefficients $Z_i(t)$ set equal to zero. Thus, out of infinitely many different equivalence classes of Lie algebras that can be realized in terms of the vector fields (2), only 2 need be considered, namely $sl(3,\mathbb{R})$ and $o(3,1)$.

Returning to the list of algebras and equations given in Ref. 1 (and leaving aside the fact that the problem was already solved by Lie), we see that the above rule (a) has not been followed. The list is too long, since it contains many algebras that are mutually equivalent. On the other hand, one of the only two "nontrivial" cases, namely the $o(3,1)$ algebra (4) is missing. This algebra is also missing from Lie's list but that is because he considers its complexification $o(4,\mathbb{C})$ which is decomposable: $o(4,\mathbb{C}) \sim o(3,\mathbb{C}) \oplus o(3,\mathbb{C})$ [compare to (5b)].

We are now also in the position to comment on rule (b). When classifying systems of ordinary differential equations we should restrict ourselves, on one hand, to equations that are not equivalent to linear ones, and on the other hand, to "indecomposable" systems of equations. By this we mean that it should not be possible to split off a subsystem of equations in fewer variables that has a superposition formula of its own. If indecomposability is ignored, seemingly very general systems of equations can be written. For example, one of Lie's algebras is[3]

$$\{\partial_y, x\,\partial_y, F_1(x)\partial_y, \ldots, F_r(x)\partial_y, y\,\partial_y\}, \quad r \geqslant 1, \tag{7}$$

where $1, x, F_1(x), \ldots, F_r(x)$ are linearly independent and the $F_i(x)$ are otherwise arbitrary differentiable functions. The corresponding "decomposable" system of equations is

$$\begin{aligned} \dot{x} &= 0, \\ \dot{y} &= Z_1(t) + Z_2(t)x + Z_3(t)F_1(x) \\ &\quad + \ldots + Z_{r+2}(t)F_r(x) + Z_{r+3}(t)y. \end{aligned} \tag{8}$$

For all practical purposes this is a system of linear equations and is of no interest. Clearly, such "false generality" should be avoided.

*Comment 2:* Ultimately the aim should be to classify all systems of ODE's

$$\dot{x}^\mu(t) = \sum_{i=1}^{l} Z_i(t)\xi_i^\mu(x^1, \ldots, x^n), \quad 1 \leqslant \mu \leqslant n \tag{9}$$

with superposition principles. A "brute force" classification of all finite-dimensional Lie algebras that can be realized in terms of vector fields in $n$ variables, even with a restriction to second-order polynomial coefficients, is an extremely difficult task. A more geometric approach, taking the above classification rules into account, goes a long way towards providing the required results.[8,9] Instead of constructing a Lie algebra $L$ of vector fields in $n$ variables directly, consider the action of a Lie group $G$ on a manifold $M$. If we restrict ourselves to transitive and effective group actions, then $M$ can be identified with a homogeneous space $G/H$, where $H$ is a subgroup of $G$ not containing a normal subgroup of $G$. Let $L$ be the Lie algebra of $G$, $L_0$ that of $H$. Then $L_0$ is realized by vector fields that vanish at the origin. The "nontriviality" requirement that the system of equations (9) should be indecomposable then implies that no coordinates exist in a neighborhood $U$ of the origin in which all vector fields constituting $L$ can be written as

$$\begin{aligned} \hat{X}(x^1, \ldots, x^n) &= \sum_{i=1}^{k} x_i(y^1 \ldots y^k)\frac{\partial}{\partial y^k} \\ &\quad + \sum_{j=k+1}^{n} b_j(y^1 \ldots y^k, z^{k+1} \ldots z^n)\frac{\partial}{\partial z^i}, \\ \{x\} &= \{y,z\}, \quad 1 \leqslant k \leqslant n-1 \end{aligned} \tag{10}$$

(the coefficients of the first $k$ derivatives depend on the first $k$ coordinates only). If such coordinates do exist, then an invariant foliation of $U$ exists. To exclude this we must require that the action of $G$ on $M$ be not only locally transitive and effective, but also locally primitive. These requirements take us directly to a classification of transitive primitive filtered Lie algebras, a task that has essentially been solved by differential geometers.[10–12] For a discussion of this classification and its implications for the construction of systems of ordinary differential equations with superposition principles see Refs. 8 and 9.

## ACKNOWLEDGMENTS

Work supported in part by the Natural Science and Engineering Research Council of Canada and the Fonds FCAC pour l'aide et le soutien à la recherche du Gouvernement du Québec.

[1]G. L. Strobel, J. Math. Phys. **24**, 1715 (1983).
[2]S. Lie and G. Sheffers, *Vorlesungen über continuirliche Gruppen* (B. G. Teubner, Leipzig, 1893), p. 799.
[3]S. Lie and F. Engel, *Theorie der Transformationsgruppen III* (B. G. Teubner, Leipzig, 1893), pp. 71–73.
[4]R. Hermann and M. Ackermann, *Sophus Lie's 1980 Transformation Group Paper* (Math. Sci. Press, Brookline, MA, 1975).
[5]R. L. Anderson, Lett. Math. Phys. **4**, 1 (1980).
[6]R. L. Anderson, J. Harnad, and P. Winternitz, Lett. Math. Phys. **5**, 143 (1981); and Physica **4D**, 164 (1982).
[7]J. Harnad, P. Winternitz, and R. L. Anderson, J. Math. Phys. **24**, 1062 (1983).
[8]S. Shnider and P. Winternitz, Centre de recherche de mathématiques appliqueés, Preprint CRMA-1164, Montreal, 1983.
[9]S. Shnider and P. Winternitz, Lett. Math. Phys. (to be published).
[10]V. W. Guillemin and S. Sternberg, Bull. Am. Math. Soc. **70**, 16 (1964).
[11]M. Golubitsky, J. Diff. Geometry **7**, 175 (1972).
[12]S. Kobayashi and T. Nagano, J. Math. Mech. **14**, 513 (1965).

# The summation of Bessel products

E. A. Evangelidis
*Plasma Physics Division, Pelindaba, Pretoria, South Africa*

(Received 30 June 1983; accepted for publication 16 March 1984)

We produce in what follows the closed forms for the summation of certain products of Bessel functions pertinent to a number of distinct fields of research, such as the theory of plasma waves, charged particle beam interaction with plasma, and density wave theory in galactic dynamics.

PACS numbers: 02.30 – f, 02.30.Lt, 52.40.Mj, 98.50.Lh

## I. INTRODUCTION

Gross[1] in his pioneering article "On Plasma Oscillations" was the first to state explicitly expressions of the following type:

$$K_p = \sum_{m=-\infty}^{\infty} \frac{J_{m+p}^{(a)} J_m^{(a)}}{m + 1 + w/w_c}, \tag{1}$$

$$L_p = \sum_{m=-\infty}^{\infty} \frac{J_{m+p}^{(a)} J_m^{(a)}}{m - 1 + w/w_c}. \tag{2}$$

He also was the first to attempt their summation into closed forms and he partially succeeded, in the sense that he stopped short only of the final integration.

More than twenty years later, similar expressions arose in the study of density wave theory and closed forms were derived and used in a series of publications[2–4] for expressions which were conveniently summarized as

$$\sum_{m=-\infty}^{\infty} \frac{J_{m+p}^{(a)} J_m^{(a)}}{m - \mu + q}. \tag{3}$$

A recent attempt[5] at the summation of similar expressions appeared, in the context of laser-beam–plasma interactions, as a particular application of a more general expression.

We would like to stress from the outset that we are not concerned at all with the general expressions of Ref. 5. The purpose of the present communication is to provide the correct expressions and domains for the particular case at hand.

In order to make our point clear, we employ Eqs. (2.3) and (2.7) of Ref. 5. It is then

$$S_1 = \sum_{n=-\infty}^{\infty} (-1)^n \frac{J_{m-n} J_n}{n + \mu} = \frac{2}{\sin(\mu\pi)} \int_0^{\pi/2} J_m(2z\cos\theta)\cos[(m + 2\mu)\theta]\, d\theta,$$

for $\alpha = m$, $\beta = 0$, $\gamma = 1$. For the definition of $\alpha, \beta, \gamma$, cf. Ref. 5.

This integral is of the standard type found in Ref. 6 (p. 738, expression 6.681.1); it exists only under the condition $m > -1$. This expression (2.8) of Ref. 5 already shows the limited applicability of the expressions derived therein. An erratum (Ref. 7) published recently provides the correct expression for $m \leqslant 0$ but in no way lifts the condition $m > -1$.

Thus we produce in what follows a simple derivation of the closed forms (1) and (2); then we show how these results can be continued to $p = -1$; and then how meaningful expressions can be obtained for $p < -1$. Finally we show how more results can be derived from expressions (4) and (5). Hence, the results presented here extend the results of Refs. 2–5 and, to the best of our knowledge, are presented for the first time.

## II. THE DERIVATION

To this effect we employ Graf's addition theorem.[8] For any complex quantities, $a, b, c, \beta, \gamma$, for which the relation

$$ce^{i\beta} = a - be^{-i\gamma} \tag{4}$$

holds true, it is also true that

$$\sum_{m=-\infty}^{\infty} J_{m+p}(a)\, J_m(b)\, e^{im\gamma} = J_p(c) e^{i\beta p}. \tag{5}$$

Under the restrictions

$$a = b, \quad \beta = \pi/2 - \gamma/2, \tag{6}$$

consistency with expression (4) requires that

$$c = 2a\sin(\gamma/2). \tag{7}$$

For reasons dictated by the divergence of integrals to be indicated further below, we first accomplish the summation of expressions (3) under the restriction

$$\text{(i)} \quad \operatorname{Re}\{p\} > -1. \tag{8}$$

Then we derive the expressions for the continuation to

$$\text{(ii)} \quad \operatorname{Re}\{p\} = -1. \tag{9}$$

Then we show how these results can be extended to

$$\text{(iii)} \quad \operatorname{Re}\{p\} < -1. \tag{10}$$

(i) For $\operatorname{Re} p > -1$ a multiplication of (5) by a factor $e^{-i(\mu+q)}$, $q$ any number, yields after an integration over $\gamma$ from 0 to $2\pi$, the sought-for form

$$\sum_{m=-\infty}^{\infty} \frac{J_{m+p}(a)\, J_m(a)}{m - \mu - q} = -\{\pi/\sin[(\mu + q)\pi]\}\, J_{p+\mu+q}(a)\, J_{-(\mu+q)}(a). \tag{11}$$

The integrals involved can be found in Ref. 6 (p. 739, expressions 6.681.8 and 6.681.9). A perusal of those expressions shows clearly the necessity of the condition $\operatorname{Re}\{p\} > -1$. It is noted that under the restriction $q$ = integer the expression (11) is written in the form

$$\sum_{m=-\infty}^{\infty} \frac{J_{m+p}(a)\, J_m(a)}{m - \mu - q} = \frac{\pi}{\sin(\mu\pi)} e^{iq\pi} J_{p+\mu+q}(a)\, J_{-(\mu+q)}(a) \tag{12}$$

—a closed form first given in Ref. 3. Expression (12) represents the closed forms of Gross's functions $K_p$, $L_p$ in a succinct form.

 0022-2488/84/072151-03$02.50  

(ii) With the observation that integrals of the form

$$\int_0^{\pi} \sin(2Mx)\, J_{2\nu}(2a \sin x)\, dx$$

(cf. Ref. 6, p. 739, expression 6.681.8) are convergent for $\mathrm{Re}\{\nu\} > -1$, i.e., $\mathrm{Re}\{p\} > -2$, we are able to find the closed form for $p = -1$ in the form

$$\sum_{m=-\infty}^{\infty} \frac{J_{m-1}(a)\, J_m(a)}{m-\mu-q} = -\frac{\pi}{\sin[(\mu+q)\pi]} J_{-(\mu+q)}(a)\, J_{(\mu+q)-1}(a). \quad (13)$$

This is the crucial expression upon which hinges the continuation on the axis of the integers to values of $p$ less than $-1$. Thus we have the following.

(iii) For $p < -1$ it suffices to make the trivial observation that

$$\nu J_\nu(a) = (a/2)\{J_{\nu-1}(a) + J_{\nu+1}(a)\}. \quad (14)$$

Thus Bessel functions of lower order, $\nu - 1$, say, are expressed in terms of Bessel functions of higher orders $\nu$ and $\nu + 1$, thus enabling the computation in terms of (13) and (11).

As a first application we retrieve from (11) the long-known expression of plasma physics

$$\sum_{m=-\infty}^{\infty} \frac{J_m^2(a)}{m-\mu} = -\frac{\pi}{\sin(\mu\pi)} J_\mu(a)\, J_{-\mu}(a) \quad (15)$$

by putting $p = 0$, $q = 0$. A combination of (12)–(14) yields

$$\sum_{m=-\infty}^{\infty} m \frac{J_m^2(a)}{m-\mu} = -\frac{\mu\pi}{\sin(\mu\pi)} J_{-\mu}(a)\, J_\mu(a). \quad (16)$$

Further, a combination of (14) with (13) shows that

$$\sum_{m=-\infty}^{\infty} m \frac{J_{m+1}(a)\, J_m(a)}{m-\mu} = -\frac{\pi(\mu+2)}{\sin(\mu\pi)} J_{-\mu}(a)\, J_{\mu+1}(a). \quad (17)$$

Differentiation of (16) with respect to the argument and a combination with (14) and

$$J'_\nu(a) = J_{\nu-1}(a) - (\nu/a)\, J_\nu(a) \quad (18)$$

shows that

$$\sum_{m=-\infty}^{\infty} m \frac{J_m(a)\, J_{m-1}(a)}{m-\mu} = -\frac{\pi}{\sin(\mu\pi)} \{2 J_{-\mu}(a)\, J_{\mu+1}(a) - \mu J_\mu(a)\, J_{1-\mu}(a)\}. \quad (19)$$

Finally, the combination of (17) and (19) yields

$$\sum_{m=-\infty}^{\infty} m^2 \frac{J_m^2(a)}{m-\mu} = -\frac{a}{2} \frac{\pi}{\sin(\mu\pi)} \{(\mu+4)\, J_{-\mu}(a)\, J_{\mu+1}(a) - \mu J_\mu(a)\, J_{1-\mu}(a)\}. \quad (20)$$

One can derive in this fashion the closed forms of products of Bessel functions up to any desired order. However, we turn our attention to the derivation of closed forms along a different direction.

Let us go back to expressions (4) and (5) and modify the restrictions (6). Namely, we require now that

$$a = -b. \quad (21)$$

Then it is

$$c e^{i\beta} = -2a \cos(\gamma/2) e^{-i(\gamma/2)}. \quad (22)$$

Thus $c = 2a\cos(\gamma/2)$,

$$e^{i\beta} = -e^{-i\gamma/2}, \quad (23)$$

implying $\beta = \pi - \gamma/2$.

Hence employing the relation

$$J_\nu(e^{im\pi} z) = e^{im\nu\pi} J_\nu(z), \quad (24)$$

we find that

$$\sum_{m=-\infty}^{\infty} J_{m+p}(b)\, J_{-m}(b)\, e^{im\gamma} = J_p\left(2b \cos\frac{\gamma}{2}\right) e^{-ip\gamma/2}. \quad (25)$$

Because of (24) we stress that this expression is true strictly for $p$ noninteger. Following the same steps as previously (and expression 6.681.1 of Ref. 6) we find that

$$\sum_{m=-\infty}^{\infty} \frac{J_{m+p}(b)\, J_{-m}(b)}{m-\mu-q} = -\frac{\pi}{\sin[2\pi(\mu+q)]} J_{p+\mu+q}(b)\, J_{-(\mu+q)}(b), \quad (26)$$

$p > -2$ noninteger.

Let us now transform Graf's addition theorem (5) according to the prescription of Ref. 8, p. 361 so that we obtain

$$\sum_{m=-\infty}^{\infty} N_{m+p}(a)\, J_m(b)\, e^{im\gamma} = N_p(c)\, e^{i\beta p}. \quad (27)$$

(We write $N_p$ instead of $Y_p$ of Ref. 8, in agreement with Ref. 6.)

Then for $a = b$, repeating the steps after Eq. (6) and using expression 6.681.2, p. 738 of Ref. 6, we find that

$$\sum_{m=-\infty}^{\infty} \frac{N_{m+p}(a)\, J_m(a)}{m+\mu} = [\pi/\sin(\mu\pi)]\{\cot(p\pi)\, J_{p-\mu}(a)\, J_\mu(a) - [1/\sin(\mu\pi)]\, J_{-\mu}(a)\, J_{\mu-p}(a)\}, \quad (28)$$

for $-1 < p < 1$ only.

It is a simple exercise now to prove that

$$\sum_{m=-\infty}^{\infty} (-1)^m \frac{I_{m+p}(a)\, I_m(a)}{m+\mu} = \frac{\pi}{\sin(\mu\pi)} I_{p-\mu}(a)\, I_\mu(a), \quad (29)$$

$p > -1$,

upon using Eq. 7, p. 361 of Ref. 8, and 6.681.3 of Ref. 6. Similarly one finds that[9]

$$\sum_{m=-\infty}^{\infty} \frac{K_{m+p}(a)\, I_m(a)}{m+\mu} = \frac{\pi}{\sin(\mu\pi)\sin(p\pi)}\{I_{\mu-p}(a)\, I_{-\mu}(a) - I_{p-\mu}(a)\, I_\mu(a)\}, \quad -1 < \mathrm{Re}\{p\} < 1. \quad (30)$$

One could certainly use Eqs. (4) and (5) and "variations" on them to produce more closed forms of Bessel function products along the lines detailed above. However, our initial purpose was to produce the correct expressions of products which would enable the exposition of the wave plasma theory in terms of closed forms. Apart from aesthetic reasons, which to our mind are more than sufficient motivation, the necessity to obtain concrete numbers dictates the retention of a few terms only of the infinite series involved in the dispersion relations; thus singularities do disappear as happened in the theory of galactic density waves—or, it turns out, they do not contribute to the growth/decay rate of the wave involved.

[1]E. P. Gross, Phys. Rev. **82**, 232 (1951).
[2]E. A. Evangelidis, Astrophys. Space Sci. **46**, 309 (1977).
[3]E. A. Evangelidis, Astrophys. Space Sci. **49**, 481 (1977).
[4]E. A. Evangelidis, Astrophys. Space Sci. **54**, 467 (1978).
[5]B. S. Newberger, J. Math Phys. **23**, 1278 (1982).
[6]I. S. Gradshteyn and I. A. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).
[7]B. S. Newberger, J. Math. Phys. **24**, 2250 (1983).
[8]G. N. Watson, *A Treatise of the Theory of Bessel Functions* (Cambridge U. P., London, 1958).
[9]It is mentioned that expression 6.681.12 of Ref. 6 used in the derivation of expression (30) of the present text is in disagreement, by a factor $\pi/2$, with expression 19.6.40, p. 378 of *Tables of Integral Transforms*, edited by A. Erdelyi (McGraw–Hill, New York, 1954), Vol. II. We have adopted the latter expression as it is obviously the correct one.

# Poisson reduction and quantization for the $n + 1$ photon

Mark J. Gotay
*Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada*

(Received 18 August 1983; accepted for publication 16 December 1983)

For a dynamical system in which the constraints are given by the vanishing of a singular momentum map $J$, reduction in the usual group-theoretic sense may not be possible. Nonetheless, one may still "reduce" $J^{-1}(0)$, at least on the level of Poisson algebras. An example of such a singular constrained system is the "$n + 1$ photon," that is, a massless, spinless particle in $(n + 1)$-dimensional Minkowski space-time. We apply the generalized reduction procedure to the $n + 1$ photon, explicitly constructing the Poisson algebra of gauge invariant observables. This technique also enables us to completely analyze the effects of the singularities in $J^{-1}(0)$ on the system. We then quantize, obtaining results which are in agreement with a quantization of the extended phase space and the subsequent imposition of the constraint.

PACS numbers: 02.40. + m, 03.20. + i

## I. INTRODUCTION

Let $(X,\omega)$ be a symplectic manifold and let $G$ be a connected Lie group with Lie algebra $g$. Assume that there is a Hamiltonian action of $G$ on $(X,\omega)$ with a $G$-equivariant momentum map $J: X \to g^*$. If $0 \in g^*$ is a regular value of $J$ and if the action of $G$ on $J^{-1}(0)$ is sufficiently nice, then the Marsden–Weinstein reduced space $J^{-1}(0)/G$ will be a symplectic manifold.[1]

These constructs are particularly relevant to physics. In this context, $(X,\omega)$ represents the extended phase space of a dynamical system, $G$ is the gauge group, and, typically, the constraints are given by $J = 0$.[2] The reduced phase space $J^{-1}(0)/G$ is then interpreted as the space of gauge invariant states of the system.

In many interesting situations, however, this group-theoretical reduction procedure does not work. For instance, it may happen that 0 is not a regular value of $J$ as in gravity and Yang–Mills theory. Moreover, even if $J^{-1}(0)$ is smooth, $J^{-1}(0)/G$ need not exist as a symplectic manifold. In either case $J$ is said to be "singular."

For systems with singular momentum maps, then, reduction in the usual sense often cannot be carried out. Nonetheless, Śniatycki and Weinstein[3] have recently pointed out that it is still possible to "reduce" $J^{-1}(0)$, at least on the level of Poisson algebras. This generalized reduction procedure allows one to determine the effects of the singularities of $J$ on the structure of the system as well as uncover certain dynamical features which would otherwise remain inaccessible. In particular, it identifies the gauge-invariant observables and equips them with the structure of a Poisson algebra. This is very useful when quantizing such a system.

Under sufficiently regular conditions, one may quantize a constrained system in two equivalent ways. The first is to quantize the extended phase space $(X,\omega)$ and then impose the constraints $J = 0$ on the quantum wave functions; this ensures that the physically admissible states are gauge invariant.[4,5] Alternatively, one may quantize the reduced phase space $J^{-1}(0)/G$,[5,6] in which case gauge invariance is directly incorporated. When $J$ is singular the latter technique is, of course, no longer applicable. But then the reduction procedure of Śniatycki and Weinstein enables one to do the next best thing, viz., to quantize the Poisson algebra of gauge-invariant observables.

Probably the simplest physically interesting example of a singular constrained system is that of a massless, spinless relativistic particle in $(n + 1)$-dimensional Minkowski space-time, which we refer to as the "$n + 1$ photon." The extended phase space is $\mathbb{R}^{2n+2}$ with coordinates $(\mathbf{p},p_t,\mathbf{x},t)$ and symplectic form

$$\omega = dp_t \wedge dt + \sum_{i=1}^{n} dp_i \wedge dx_i .$$

The gauge group is $\mathbb{R}$ with momentum map

$$J(\mathbf{p},p_t,\mathbf{x},t) = p_t^2 - \|\mathbf{p}\|^2.$$

Since the particle is massless, $J$ must vanish. The constraint set is thus

$$J^{-1}(0) = C^n \times \mathbb{R}^{n+1},$$

where $C^n$ is the null cone in $\mathbb{R}^{n+1}$. In this paper we reduce $J^{-1}(0)$ on the Poisson algebra level and then quantize, obtaining results which are in exact agreement with the quantization of the extended phase space $(\mathbb{R}^{2n+2},\omega)$ and the subsequent imposition of the constraint $J = 0$.

This example serves three purposes: First, it illustrates the usefulness and essential correctness, at least in this instance, of the generalized reduction procedure. Second, it is simple enough that we can both identify and completely analyze the effects of the singularities in $J^{-1}(0)$ on this system. In this regard, our presentation seems to be the first which treats the singularities seriously (compare with standard discussions of the $3 + 1$ photon, e.g., that given in Ref. 7). Finally, Arms, Marsden, and Moncrief[8] have shown that singular momentum mappings typically have quadratic singularities so that $J^{-1}(0)$ is always a "cone." Since the $n + 1$ photon is an elementary, and in some sense canonical, example of this phenomenon, its elucidation is essential for further progress in understanding the structure of singular constrained systems.

In the next section we briefly recall the basic features of the Śniatycki–Weinstein reduction procedure. The details for the $1 + 1$ photon are then worked out in Sec. III. The $n = 1$ case is done separately, since it is rather "special" and technically much easier than the $n > 1$ case, which is elabor-

 0022-2488/84/072154-06$02.50  

ated upon in Sec. IV. The physical interpretation of these results is discussed in the last section.

## II. POISSON ALGEBRAS, REDUCTION AND QUANTIZATION

Let $\mathscr{F}$ be a commutative algebra over $\mathbb{R}$. If $[\cdot,\cdot]$ is a bracket operation on $\mathscr{F}$ such that (i) the pair $(\mathscr{F},[\cdot,\cdot])$ is a Lie algebra and (ii) the Leibniz rule

$$[f,f_1f_2] = [f,f_1]f_2 + [f,f_2]f_1$$

holds, then $(\mathscr{F},[\cdot,\cdot])$ is called a *Poisson algebra*. The basic example of a Poisson algebra is $C^\infty(X)$, where $(X,\omega)$ is symplectic and the Poisson bracket is given by

$$\{f,g\} = -\omega(\xi_f,\xi_g).$$

Here $\xi_f$, the Hamiltonian vector field of $f$, is defined via

$$i_{\xi_f}\omega = -df.$$

Now let $(X,\omega)$, $G$, and $J$ be as in the Introduction. For each $a\in\mathfrak{g}$ define the function $J_a$ on $X$ by $J_a(x) = \langle J(x),a\rangle$, and denote by $\mathscr{J}$ the ideal (relative to the associative algebra structure) in $C^\infty(X)$ generated by the $J_a$. Since $J$ is $G$-equivariant, the action of $G$ on $C^\infty(X)$ induces an action of $G$ on $C^\infty(X)/\mathscr{J}$ in such a way that the projection homomorphism $j$: $C^\infty(X)\rightarrow C^\infty(X)/\mathscr{J}$ is $G$-equivariant. Let $\mathscr{F}$ be the space of $G$-invariant elements of $C^\infty(X)/\mathscr{J}$, that is, the collection of all equivalence classes $jf$ for which $j(\{f,\mathscr{J}\}) = 0$. Again by equivariance, the Poisson bracket $\{\cdot,\cdot\}$ on $C^\infty(X)$ descends to a bracket $[\cdot,\cdot]$ on $\mathscr{F}$ given by

$$[jf,jg] = j(\{f,g\}). \tag{2.1}$$

The pair $(\mathscr{F},[\cdot,\cdot])$ is the *reduced Poisson algebra* of the constrained system under consideration.

If 0 is a regular value of $J$, then $C^\infty(X)/\mathscr{J} = C^\infty(J^{-1}(0))$. Furthermore, if $J^{-1}(0)/G$ is a quotient manifold of $J^{-1}(0)$, then the reduced Poisson algebra $\mathscr{F}$ is canonically isomorphic to the Poisson algebra of the reduced symplectic space $J^{-1}(0)/G$. Under sufficiently regular conditions, then, this generalized reduction procedure is consistent with the Marsden–Weinstein technique, and we may therefore interpret $(\mathscr{F},[\cdot,\cdot])$ as the Poisson algebra of gauge-invariant observables. It is important to note, however, that in the singular case $\mathscr{F}$ need not be the Poisson algebra of any symplectic manifold nor must it be nondegenerate (in the sense that the only elements of $\mathscr{F}$ which Poisson commute with everything are "constant"[9]).

We close this section with some remarks concerning the quantization of a Poisson algebra $(\mathscr{F},[\cdot,\cdot])$. The problem is to construct the quantum state space from a knowledge of this Poisson algebra. This is fairly straightforward, using the techniques of geometric quantization theory,[7] when $\mathscr{F}$ is associated with a symplectic manifold. In the singular case it is necessary to proceed by analogy; briefly, this works as follows.[3]

Let $\Gamma = \mathscr{F}\otimes\mathbb{C}$ be the complexification of $\mathscr{F}$; elements $\sigma\in\Gamma$ are the algebraic counterparts of sections of the prequantization line bundle (which we take to be trivial). Given a derivation $\xi$ of $\mathscr{F}$, we may compute the "covariant derivative" $\nabla_\xi\sigma$ of a section $\sigma$ once a connection $\nabla$ on $\Gamma$ has been specified. A *polarization* $\mathscr{P}$ is a maximal commuting subalgebra of $(\mathscr{F},[\cdot,\cdot])$. A section $\sigma\in\Gamma$ is said to be "polarized" provided $\nabla_{\xi_f}\sigma = 0$ for all $f\in\mathscr{P}$, where $\xi_f$ is the derivation $g\rightarrow[g,f]$ corresponding to the Hamiltonian vector field of $f$. The quantum state space relative to this data is then defined to be the set of all linear functionals on the space of polarized sections in $\Gamma$.

For our purposes we may choose a connection $\nabla$ such that

$$\nabla_{\xi_f}\sigma = [\sigma,f]$$

for all $f\in\mathscr{P}$. Then the space of polarized sections in $\Gamma$ is precisely $\mathscr{P}\otimes\mathbb{C}$, and the quantum wave functions are elements of the dual $(\mathscr{P}\otimes\mathbb{C})'$.

Turning now to the example, we compute the reduced Poisson algebra for the $n+1$ photon and quantize it.

## III. THE 1 + 1 PHOTON

The analysis of the $n+1$ photon is considerably easier when $n=1$, for then the constraint $J=0$ factors. This circumstance simplifies the algebraic computations required for the construction of the reduced Poisson algebra as well as its presentation. This simplicity is also reflected in the structure of the constraint set $J^{-1}(0) = C^n\times\mathbb{R}^{n+1}$, which is essentially trivial when $n=1$.

We begin by changing to null coordinates

$$u = t-x,\quad v = t+x,$$

and their corresponding momenta

$$\mu = p_t - p_x,\quad \nu = p_t + p_x.$$

The symplectic form on $\mathbb{R}^4$ is then

$$\omega = \tfrac{1}{2}(d\mu\wedge du + d\nu\wedge dv)$$

and the momentum map becomes

$$J(\mu,\nu,u,v) = \mu\nu.$$

The ideal $\mathscr{J}$ of $C^\infty(\mathbb{R}^4)$ is thus generated by the product $\mu\nu$. Define $j$: $C^\infty(\mathbb{R}^4)\rightarrow C^\infty(\mathbb{R}^3)\times C^\infty(\mathbb{R}^3)$ by

$$jf = (f(\mu,0,u,v), f(0,\nu,u,v)). \tag{3.1}$$

*Proposition 3.1*: The quotient $C^\infty(\mathbb{R}^4)/\mathscr{J}$ may be identified with the image of $C^\infty(\mathbb{R}^4)$ in $C^\infty(\mathbb{R}^3)\times C^\infty(\mathbb{R}^3)$ under $j$.

*Proof*: If $f\in\mathscr{J}$, then clearly $jf=0$. On the other hand, suppose that $jf=0$. Then $f(\mu,0,u,v)=0$ which, by Hadamard's lemma, implies that $f$ is divisible by $\nu$. Thus $f=\nu h$ for some smooth $h$. Then $f(0,\nu,u,v)=0$ yields $h(0,\nu,u,v)=0$, which similarly implies that $h$ is divisible by $\mu$ and so $f\in\mathscr{J}$. Thus $\ker j = \mathscr{J}$ and the claim follows. Q.E.D.

Now $jf\in\mathscr{F}$ iff $j(\{f,J\})=0$. From (3.1) this will be the case iff

$$\frac{\partial f}{\partial v}(\mu,0,u,v) = 0 = \frac{\partial f}{\partial u}(0,\nu,u,v),$$

so that the invariant elements of $C^\infty(\mathbb{R}^4)/\mathscr{J}$ are of the form

$$(f(\mu,0,u,0), f(0,\nu,0,v))$$

with $f(0,0,u,v)$ constant. We may thus regard $\mathscr{F}$ as consisting of pairs of functions

$$(\psi(\mu,u),\ \phi(\nu,v))\in C^\infty(\mathbb{R}^2)\times C^\infty(\mathbb{R}^2)$$

subject to the *compatibility conditions*

$$\psi(0,u) = \phi(0,v)\quad (=\text{const}). \tag{3.2}$$

In these terms, a direct calculation shows that the induced Poisson bracket (2.1) on $\mathscr{F}$ is given by

$$[(\psi_1,\phi_1), (\psi_2,\phi_2)] = (2[\psi_1,\psi_2]_{u,\mu}, 2[\phi_1,\phi_2]_{v,\nu}), \quad (3.3)$$

where

$$[\psi_1,\psi_2]_{u,\mu} = \frac{\partial\psi_1}{\partial u}\frac{\partial\psi_2}{\partial\mu} - \frac{\partial\psi_1}{\partial\mu}\frac{\partial\psi_2}{\partial u}$$

denotes the ordinary Poisson bracket with respect to the pair $u, \mu$ etc. It is straightforward to check that $[\cdot,\cdot]$ is nondegenerate.

In view of (3.3), the reduced Poisson algebra $\mathscr{F}$ is closely related to the Poisson algebra $C^\infty(\mathbb{R}^2)\times C^\infty(\mathbb{R}^2)$ of the symplectic manifold consisting of two disjoint copies of $\mathbb{R}^2$. Due to the compatibility conditions (3.2), however, $\mathscr{F}$ is strictly a subalgebra of this Poisson algebra, and so is not the Poisson algebra of any symplectic manifold. These conditions therefore express the influence of the singularities in $J^{-1}(0)$ upon the system. In fact, a correlation between these two Poisson algebras might have been expected from a consideration of the case when the photon has a mass $m$. Then the constraint set $J^{-1}(m^2)$ is nonsingular, but disconnected, and the reduced phase space is symplectomorphic to $\mathbb{R}^2\cup\mathbb{R}^2$. It follows that the reduced Poisson algebra for a massive particle is exactly $C^\infty(\mathbb{R}^2)\times C^\infty(\mathbb{R}^2)$. The effect of letting $m\to 0$ is thus to reduce the number of gauge-invariant observables. We shall have more to say about the physical interpretation of this phenomenon, and its relationship to the singular space $J^{-1}(0)/\mathbb{R}$, in Sec. V.

To construct the quantum state space, we must choose a polarization $\mathscr{P}$ of $\mathscr{F}$. Noting that the horizontal polarization $P$ on $\mathbb{R}^4$ spanned by the vector fields $\xi_\mu$ and $\xi_\nu$ projects onto $J^{-1}(0)$, a natural choice for $\mathscr{P}$ is

$$\mathscr{P} = \{(\psi(\mu), \phi(\nu)) \mid \psi(0) = \phi(0)\}. \quad (3.4)$$

According to general considerations, then, the quantum wave functions are elements of $(\mathscr{P}\otimes\mathbb{C})'$.

To represent these states, we need the following result: Consider $\mathbb{R}^2$ with coordinates $\mu$ and $\nu$, and let $\hat{\mathscr{J}}$ be the ideal in $C^\infty(\mathbb{R}^2,\mathbb{C})$ generated by the product $\mu\nu$.

*Lemma*: $C^\infty(\mathbb{R}^2,\mathbb{C})/\hat{\mathscr{J}} = \mathscr{P}\otimes\mathbb{C}$.

*Proof*: Mimicking the proof of Proposition 3.1, we have that $C^\infty(\mathbb{R}^2)/\hat{\mathscr{J}}$ may be identified with the image of $C^\infty(\mathbb{R}^2)$ in $C^\infty(\mathbb{R})\times C^\infty(\mathbb{R})$ under the map $f\to(f(\mu,0), f(0,\nu))$. Comparison with (3.4) and complexification then yields the desired result. Q.E.D.

With this in hand, we now establish:

*Proposition 3.2*: $(\mathscr{P}\otimes\mathbb{C})'$ is isomorphic to the space of all complex-valued distributions $\Phi$ on $\mathbb{R}^2$ satisfying

$$\mu\nu\Phi = 0. \quad (3.5)$$

*Proof*: Let $\Phi$ be such a distribution, in which case $\Phi$ annihilates all functions which are divisible by $\mu\nu$. Then $\Phi$ induces a linear functional $\hat{\Phi}$ on $C^\infty(\mathbb{R}^2,\mathbb{C})/\hat{\mathscr{J}}$ so that, by the Lemma, $\hat{\Phi}\in(\mathscr{P}\otimes\mathbb{C})'$. Conversely, every linear functional on $\mathscr{P}\otimes\mathbb{C} = C^\infty(\mathbb{R}^2,\mathbb{C})/\hat{\mathscr{J}}$ can be lifted to a distribution on $\mathbb{R}^2$ satisfying (3.5). Q.E.D.

These distributions $\Phi$ take the form

$$\Phi(\mu,\nu) = \lambda(\mu)\otimes\delta(\nu) + \delta(\mu)\otimes\chi(\nu),$$

where $\lambda$ and $\chi$ are distributions on $\mathbb{R}$. Then for $f\in C^\infty(\mathbb{R}^2,\mathbb{C})$,

$$\langle\Phi,f\rangle = \langle\lambda(\mu), f(\mu,0)\rangle + \langle\chi(\nu), f(0,\nu)\rangle,$$

from which we obtain the explicit representation

$$\hat{\Phi}(\mu,\nu) = (\lambda(\mu), \chi(\nu))$$

of $\hat{\Phi}$ as a linear functional on $\mathscr{P}\otimes\mathbb{C}$.

Proposition 3.2 is the main result of this section. Not surprisingly, it shows that the gauge invariant wave functions must satisfy the 1 + 1 wave equation, which is just the Fourier transform of (3.5). It also guarantees that this quantization is equivalent to that of the extended phase space $(\mathbb{R}^4,\omega)$. In fact, quantizing in the momentum representation defined by the polarization $P$, we find that the quantum Hilbert space is $L^2(\mathbb{R}^2)$ and that the quantum operator $\mathscr{Q}J$ corresponding to $J$ is given by

$$\mathscr{Q}J[\Phi] = \mu\nu\Phi.$$

Thus, from this point of view as well, the physically admissible photon states must coincide with the distributional solutions of (3.5).

Finally, note the crucial role of the compatibility conditions (3.2), in the guise of (3.4), in Proposition 3.2. Without them (3.5) would not follow and the correlation with the wave equation would be lost.

## IV. THE $n + 1$ PHOTON

For the 1 + 1 photon the constraint set consists simply of two intersecting hyperplanes in $\mathbb{R}^4$. This enabled us to compute directly on $J^{-1}(0)$; in effect, we worked on each of the two hyperplanes and then "glued" along their intersection by means of the compatibility conditions. For $n > 1$, $J^{-1}(0)$ is more complicated and we can no longer proceed in this straightforward manner. In particular, it is now necessary to "resolve" the singularity.

Our first task is to construct the quotient $C^\infty(\mathbb{R}^{2n+2})/\mathscr{J}$. The following result is the higher-dimensional analog of Proposition 3.1. Let $f\in C^\infty(\mathbb{R}^{2n+2})$.

*Proposition 4.1*: $f\in\mathscr{J}$ iff $f|J^{-1}(0) = 0$.

*Proof*: The obverse is apparent. For the converse, it is clear from the structure of the constraint set $J^{-1}(0) = C^n\times\mathbb{R}^{n+1}$ that the configuration variables $(\mathbf{x},t)$ are largely irrelevant and may accordingly be factored out. We are thus effectively reduced to proving that if $g: \mathbb{R}^{n+1}\to\mathbb{R}$ is such that $g|C^n = 0$, then $g$ is globally divisible by $p_t^2 - ||\mathbf{p}||^2$.

There is no problem off $C^n$. On either of the regular components of $C^n$, this follows from the inverse function theorem and Hadamard's lemma. It remains only to demonstrate that $g$ is divisible by $p_t^2 - ||\mathbf{p}||^2$ at the vertex of $C^n$, and for this it suffices (Ref. 10, p. 72) to show that the formal Taylor series of $g$ at the origin is divisible by $p_t^2 - ||\mathbf{p}||^2$. We now establish this for $n = 2$; this case is prototypical, and the generalization to arbitrary $n$ is immediate.

Thus let

$$T_0^r g = \sum_{i+j+k=r} \frac{1}{i!j!k!} g_{i,j}^k p_x^i p_y^j p_t^k \quad (4.1)$$

be the homogeneous part of the $r$th Taylor polynomial of $g$ at the origin of $\mathbb{R}^3$, where

$$g_{i,j}^k = \frac{\partial^{i+j+k} g}{\partial p_x^i \partial p_y^j \partial p_t^k}(0,0,0).$$

In (4.1) view all variables other than $p_t$ as parameters. Then to say that $T_0^r g$ is divisible by $p_t^2 - (p_x^2 + p_y^2)$ is equivalent to requiring that both $p_t = \pm(p_x^2 + p_y^2)^{1/2}$ be roots of $T_0^r g$. Substituting these values for $p_t$ into (4.1), decomposing the sum into even and odd powers of $(p_x^2+p_y^2)^{1/2}$, expanding these powers in a binomial series and reorganizing gives

$$\Big(\sum_{m+n=r} a_{mn}p_x^m p_y^n\Big) \pm (p_x^2+p_y^2)^{1/2}\Big(\sum_{m+n=r-1} b_{mn}p_x^m p_y^n\Big), \tag{4.2}$$

where

$$a_{mn} = \sum_{l=0}^{[m/2]}\sum_{k=l}^{[n/2+l]} \binom{k}{l} \times \frac{1}{(m-2l)!(n-2k+2l)!(2k)!} g^{2k}_{m-2l,n-2k+2l}, \tag{4.3}$$

$$b_{mn} = \sum_{l=0}^{[m/2]}\sum_{k=l}^{[n/2+l]} \binom{k}{l} \times \frac{1}{(m-2l)!(n-2k+2l)!(2k+1)!} g^{2k+1}_{m-2l,n-2k+2l}, \tag{4.4}$$

and $[k]$ denotes the greatest integer less than or equal to $k$. From (4.2) it follows that $p_t = \pm(p_x^2+p_y^2)^{1/2}$ will be roots of $T_0^r g$ iff the coefficients $a_{mn}$ and $b_{mn}$ vanish.

Now let $\mathbf{v}$ be a vector at the origin which points along a generator of the cone, and consider the $r$th derivative of $g$ in the direction $\mathbf{v}$:

$$D_{\mathbf{v}}^r g(0,0,0) = \Big[\Big(\mathrm{v}_x\frac{\partial}{\partial p_x}+\mathrm{v}_y\frac{\partial}{\partial p_y}+\mathrm{v}_t\frac{\partial}{\partial p_t}\Big)^r g\Big](0,0,0).$$

Another lengthy calculation, consisting of expanding this expression out, separating into even and odd powers of $\mathrm{v}_t$, and then using the fact that $\mathrm{v}_t^2=\mathrm{v}_x^2+\mathrm{v}_y^2$, yields

$$D_{\mathbf{v}}^r g(0,0,0) = r!\Big(\sum_{m+n=r} a_{mn}\mathrm{v}_x^m\mathrm{v}_y^n\Big) \pm (r-1)!(\mathrm{v}_x^2+\mathrm{v}_y^2)^{1/2}\Big(\sum_{m+n=r-1} b_{mn}\mathrm{v}_x^m\mathrm{v}_y^n\Big),$$

where $a_{mn}$ and $b_{mn}$ are given by (4.3) and (4.4), respectively. But by assumption $g|C^n=0$ so that $D_{\mathbf{v}}^r g(0,0,0)=0$ for all such $\mathbf{v}$. This implies that $a_{mn}=0$ and $b_{mn}=0$, and we are finished. Q.E.D.

This proposition shows that

$$C^\infty(\mathbb{R}^{2n+2})/\mathscr{J} = C^\infty(J^{-1}(0)),$$

the smooth functions on $J^{-1}(0)$ in the sense of Whitney.[11] Unfortunately, $C^\infty(J^{-1}(0))$ is rather difficult to handle. To obtain a more tractable representation of $C^\infty(\mathbb{R}^{2n+2})/\mathscr{J}$, we "resolve" the singularity by means of the map $\tilde\phi: \mathbb{R}^{2n+2}\to\mathbb{R}^{2n+2}$ given by

$$\tilde\phi(\boldsymbol{\pi},p_t,\mathbf{x},t) = (p_t\boldsymbol{\pi},p_t,\mathbf{x},t).$$

Note that now the physical momenta are given by $p_t$ and $\mathbf{p}=p_t\boldsymbol{\pi}$. If we define $K$: $\mathbb{R}^{2n+2}\to\mathbb{R}$ via

$$K(\boldsymbol{\pi},p_t,\mathbf{x},t) = 1-||\boldsymbol{\pi}||^2,$$

then $K^{-1}(0)=(S^{n-1}\times\mathbb{R})\times\mathbb{R}^{n+1}$ and $\tilde\phi(K^{-1}(0))=J^{-1}(0)$. Let $\phi$ be the restriction of $\tilde\phi$ to $K^{-1}(0)$. Note that $\phi$ is a local diffeomorphism away from the "equator" $p_t=0$ and collapses the equator $(S^{n-1}\times\{0\})\times\mathbb{R}^{n+1}$ onto the *singular set* $S=\{(\mathbf{0},0)\}\times\mathbb{R}^{n+1}$ in $J^{-1}(0)$.

We think of $K^{-1}(0)$ as being a "covering manifold" of the singular space $J^{-1}(0)$; using $\phi$, we pull the entire formalism on $J^{-1}(0)$ back to $K^{-1}(0)$. The advantages of this procedure are (i) $K^{-1}(0)$ is a manifold and (ii) we can dispense with $C^\infty(J^{-1}(0))$ directly and work instead with its more manageable isomorph $\phi^*C^\infty(J^{-1}(0))\subset C^\infty(K^{-1}(0))$. The key fact which makes this possible is that $\phi^*C^\infty(J^{-1}(0))$ admits a relatively simple characterization in $C^\infty(K^{-1}(0))$ in terms of formal Taylor series.[12]

*Proposition 4.2*: Let $F\in C^\infty(K^{-1}(0))$. Then $F\in\phi^*C^\infty(J^{-1}(0))$ iff for each $s\in S$ there exists a formal power series $\not{f}_s$ at $s$ such that

$$T_qF = \not{f}_s\circ T_q\phi \tag{4.5}$$

for all $q\in\phi^{-1}(s)$.

*Proof*: Suppose that $F=f\circ\phi$ for some $f\in C^\infty(J^{-1}(0))$. Let $\tilde f$ be any extension of $f$ to $\mathbb{R}^{2n+2}$; then $\not{f}_s=T_s\tilde f$ will do in (4.5). The reverse implication follows from the inverse function theorem and Theorem 3.2 of Ref. 12. Q.E.D.

Note that (4.5) is a very strong condition: for a smooth function $F$ on $K^{-1}(0)$ to lie in $\phi^*C^\infty(J^{-1}(0))$, it does *not* suffice for $F$ simply to factor through $\phi$. Rather, (4.5) requires that $F$ *and* all its formal Taylor series $T_qF$ factor through $\phi$.

In summary, we henceforth work on $K^{-1}(0)$ and identify

$$C^\infty(\mathbb{R}^{2n+2})/\mathscr{J} = \phi^*C^\infty(J^{-1}(0)).$$

From this standpoint, the conditions (4.5) reflect the presence of the singularities in $J^{-1}(0)$.[13] With these considerations out of the way, we are now ready to construct the reduced Poisson algebra.

Let $F\in\phi^*C^\infty(J^{-1}(0))$ so that there exists a smooth function $\tilde f$ on $\mathbb{R}^{2n+2}$ with $F=\tilde f\circ\phi$. Then $F$ will be invariant provided $\{\tilde f,J\}\circ\phi=0$ which, on $K^{-1}(0)$, translates into

$$\frac{\partial F}{\partial t} - \sum_{i=1}^n \pi_i\frac{\partial F}{\partial x_i} = 0.$$

Setting $\mathbf{w}=\mathbf{x}+\boldsymbol{\pi}t$, this implies that $F=F(\boldsymbol{\pi},p_t,\mathbf{w})$ only. Since $F$ must also factor through $\phi$, it follows (with a slight abuse of notation) that

$$\mathscr{F} = \{F\in\phi^*C^\infty(J^{-1}(0))\,|\,F=F(p_t\boldsymbol{\pi},p_t,p_t\mathbf{w})\}. \tag{4.6}$$

Now if $F$ and $G$ are two elements of $\mathscr{F}$ with $F=\tilde f\circ\phi$ and $G=\tilde g\circ\phi$, then the induced Poisson bracket (2.1) on $\mathscr{F}$ is $[F,G]=\{\tilde f,\tilde g\}\circ\phi$. After making the coordinate change $(\boldsymbol{\pi},p_t,\mathbf{x},t)\mapsto(\boldsymbol{\pi},p_t,\mathbf{w},t)$ on $K^{-1}(0)$, a straightforward computation yields

$$[F,G] = \sum_{i=1}^n [F,G]_{w_i,p_t}\pi_i + \frac{1}{p_t}\sum_{i,j=1}^n [F,G]_{\omega_i,\pi_j}(\delta_{ij}-\pi_i\pi_j). \tag{4.7}$$

Although this expression would appear to be singular when $p_t=0$, in fact it is not because of (4.6).

We show that (4.7) is nondegenerate. Indeed, suppose that $[F,G]=0$ for all $G$ in $\mathscr{F}$. Take $G=p_tw_k$. Then $[F,p_tw_k]=0$ reduces to

$$\pi_k\Big(p_t\frac{\partial F}{\partial p_t} - \sum_{i=1}^n\pi_i\frac{\partial F}{\partial\pi_i}\Big) + \frac{\partial F}{\partial\pi_k} = 0.$$

Multiply this by $\pi_k$ and sum; since $||\boldsymbol{\pi}||^2=1$, it follows that $\partial F/\partial p_t=0$. But then, by (4.6), $F(p_t\boldsymbol{\pi},p_t,p_t\mathbf{w})=F(\mathbf{0},0,\mathbf{0})$ is constant and nondegeneracy is proven.

The quantization of the $n+1$ photon is patterned after that of the $1+1$ photon given in Sec. III. The analog of the horizontal polarization $P$ on $\mathbb{R}^{2n+2}$ spanned by the vector fields $\xi_{p_t}$ and $\xi_{p_i}$, $i = 1,...,n$, is the maximal commuting subalgebra

$$\mathscr{P} = \{F \in \mathscr{F} \,|\, F = F(p_t, \pi, p_t)\} \qquad (4.8)$$

of $\mathscr{F}$. We now construct the quantum state space $(\mathscr{P} \otimes \mathbb{C})'$.

Let $\hat{J}$ and $\hat{K}$ be the restrictions of $J$ and $K$ to the first factor of $\mathbb{R}^{n+1}$ in $\mathbb{R}^{2n+2}$, and denote by $\hat{\mathscr{J}}$ the ideal in $C^\infty(\mathbb{R}^{n+1})$ generated by $\hat{J}$. From the proof of Proposition 4.1 we see that

$$C^\infty(\mathbb{R}^{n+1})/\hat{\mathscr{J}} = C^\infty(C^n).$$

Letting $\hat{\phi}$ be the restriction of $\tilde{\phi}$ to $\hat{K}^{-1}(0)$, we may then identify $C^\infty(\mathbb{R}^{n+1})/\hat{\mathscr{J}}$ with the subalgebra $\hat{\phi}^* C^\infty(C^n)$ of $C^\infty(S^{n-1} \times \mathbb{R})$. From (4.8), (4.5), and the analog of Proposition 4.2 applied to $\hat{\phi}^* C^\infty(C^n) \subset C^\infty(S^{n-1} \times \mathbb{R})$, it follows that $\hat{\phi}^* C^\infty(C^n)$ is isomorphic to $\mathscr{P}$. Upon complexifying, we finally obtain

$$C^\infty(\mathbb{R}^{n+1}, \mathbb{C})/\hat{\mathscr{J}} = \mathscr{P} \otimes \mathbb{C}.$$

Imitating the proof of Proposition 3.2, this last result yields:

*Proposition 4.3*: $(\mathscr{P} \otimes \mathbb{C})'$ is isomorphic to the space of all complex-valued distributions $\Phi$ on $\mathbb{R}^{n+1}$ satisfying

$$(p_t^2 - \|\mathbf{p}\|^2)\Phi = 0.$$

Thus, as before, the physically admissible photon states must satisfy the Fourier transformed $n+1$ wave equation. As expected, this is consistent with the quantization of the extended phase space $(\mathbb{R}^{2n+2}, \omega)$ in the polarization $P$. Indeed, we compute

$$\mathscr{Q}J[\Phi] = (p_t^2 - \|\mathbf{p}\|^2)\Phi$$

on $L^2(\mathbb{R}^{n+1})$ and gauge invariance demands $\mathscr{Q}J[\Phi] = 0$.

## V. DISCUSSION

We spend a moment correlating our results with the structure of the singular reduced space $J^{-1}(0)/\mathbb{R}$. This will incidentally help clarify the physical significance of the compatibility conditions (3.2) and their higher-dimensional analogs (4.6) which arise both from the presence of singularities and the requirements of gauge invariance.

The action of the gauge group $\mathbb{R}$ on $\mathbb{R}^{2n+2}$ is given by

$$(\lambda; \mathbf{p}, p_t, \mathbf{x}, t) \mapsto (\mathbf{p}, p_t, \mathbf{x} - 2\lambda\mathbf{p}, t + 2\lambda p_t).$$

On $J^{-1}(0) = C^n \times \mathbb{R}^{n+1}$ this action fixes every point of the singular set $S$ and is otherwise free. We may therefore schematically represent $J^{-1}(0)/\mathbb{R}$ as shown in Fig. 1. The trouble with $J^{-1}(0)/\mathbb{R}$, aside from the expected conical singularity, stems from the anomalous factor of $\mathbb{R}^{n+1}$ associated with the vertex. This is actually a remnant of a slight defect in the extended phase space description of the $n+1$ photon concerning the physical interpretation of states in the singular set $S \subset J^{-1}(0)$. Such a state $(\mathbf{0}, 0, \mathbf{x}, t)$ represents a photon with vanishing momentum located at $(\mathbf{x}, t)$, that is, a vacuum state. But presumably there is only a *single* vacuum state, not one located at every space-time point. It is this $(n+1)$-dimensional array of unphysical vacua which contributes to the pathology in $J^{-1}(0)/\mathbb{R}$ and prevents the latter from being construed as the space of all gauge-invariant states.

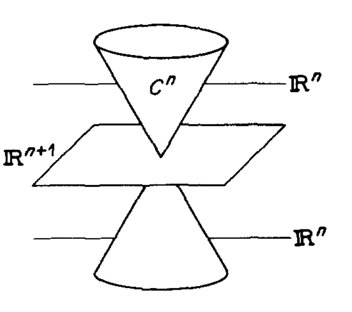


FIG. 1. The singular reduced space $J^{-1}(0)/\mathbb{R}$.

On the other hand, a physical observable should be unable to distinguish between these spurious vacua. The topology of the reduced space indicates that this will be the case: since $J^{-1}(0)/\mathbb{R}$ fails to be Hausdorff along this $\mathbb{R}^{n+1}$, continuous functions cannot separate these states. This observation is substantiated by our analysis above, and here is where both gauge invariance and the compatibility conditions enter. For $n = 1$, (3.2) guarantees that a physical observable is constant on $S$. Similarly, for $n > 1$, the form (4.6) of a gauge invariant function ensures that it is constant along the equator $\phi^{-1}(S)$ and hence also cannot differentiate between these states. Consequently, the generalized reduction process "corrects" the flaws in both the original description of the system and the reduced phase space, at least to the extent that it guarantees that the gauge invariant observables "detect" but a single vacuum state, as required.

Our analysis of the $n+1$ photon thus demonstrates the utility of the Poisson algebra approach: even though a system may be singular, one can still construct the essential components of the reduced canonical formalism. Moreover, subsequent quantization yields results in exact correspondence with those obtained by standard methods. We hope that this example will encourage further study of the structure of singular constrained systems. Techniques for resolving singularities and, in particular, the work of Bierstone and Milman[12] on composite differentiable functions (of which Proposition 4.2 is a special case) should prove to be quite valuable in this regard.

## ACKNOWLEDGMENTS

I would like to thank P. D. Milman, R. Montgomery, D. L. Rod, J. Śniatycki, R. E. Woodrow, and the referee for helpful discussions. I am especially indebted to L. Bos for both his interest and his substantial contributions to this paper. This work was partially supported by a Natural Sciences and Engineering Research Council of Canada operating grant to the University of Calgary.

[1]J. E. Marsden and A. Weinstein, Rep. Math. Phys. 5, 121–30 (1974).

[2]M. J. Gotay, J. A. Isenberg, J. E. Marsden, R. Montgomery, J. Śniatycki, and P. B. Yasskin, "Constraints and Momentum Mappings in Relativistic Field Theory," in preparation.

[3]J. Śniatycki and A. Weinstein, Lett. Math. Phys. 7, 155–61 (1983).

[4]M. J. Gotay and J. Śniatycki, Commun. Math. Phys. 82, 377–89 (1981).

[5]J. Śniatycki, "Constraints and Quantization," Preprint #506, Department of Mathematics and Statistics, University of Calgary, 1982.

[6]V. Guillemin and S. Sternberg, Inv. Math. 67, 515–38 (1982); I. Vaisman, "Geometric Quantization on Presymplectic Manifolds," Report #66, Department of Mathematics, University of Haifa, 1983.

[7]N. M. J. Woodhouse, *Geometric Quantization* (Clarendon, Oxford, 1980).

[8]J. M. Arms, J. E. Marsden, and V. Moncrief, Commun. Math. Phys. 78, 455–78 (1981).

[9]Here is an example of a degenerate reduced Poisson algebra. Lift the irrational flow on the torus $T^2$ to $T^*T^2$ and let $J$ be the usual momentum map for this cotangent action. Then $J^{-1}(0) \approx T^2 \times \mathbb{R}$, $\mathscr{F} \approx C^\infty(\mathbb{R})$, and $[\cdot,\cdot] = 0$. However, it is still possible that nondegeneracy will hold whenever the group action is proper.

[10] B. Malgrange, *Ideals of Differentiable Functions* (Oxford U.P., Oxford, 1966).

[11] That is, a function $f$ on $J^{-1}(0)$ belongs to $C^\infty(J^{-1}(0))$ iff $f$ extends to a smooth function on an open set in $\mathbf{R}^{2n+2}$ containing $J^{-1}(0)$.

[12] E. Bierstone and P. D. Milman, Ann. Math. **116**, 541–58 (1982).

[13] Insofar as $\phi^* C^\infty(J^{-1}(0))$ is strictly a subspace of $C^\infty(K^{-1}(0))$. Note that our method of resolving the singularity (using the map $\tilde{\phi}$) yields the cylinder $K^{-1}(0)$ as the "nonsingular model" for the cone $J^{-1}(0)$ rather than a two-component hyperboloid as might be expected on physical grounds. In fact, it does not seem possible to resolve $J^{-1}(0)$ as $J^{-1}(m^2)$ for any mass $m$; this indicates that the $m \rightarrow 0$ limit is in some sense highly singular.

# Estimation of inverse temperature and other Lagrange multipliers: The dual distribution

Y. Tikochinsky[a] and R. D. Levine
*The Fritz Haber Research Center for Molecular Dynamics, The Hebrew University, Jerusalem 91904, Israel*

(Received 17 November 1983; accepted for publication 10 February 1984)

It is shown that the problem of parameter estimation for distributions of the exponential type, has a unique consistent Bayesian solution: The requirement that Bayes' rule and maximum entropy lead to the same inverse distribution determines the loss function. Similarly, the demand that the best estimate for a random variable, given an observed value of that variable, coincides with the observed value, determines the prior distribution for the corresponding conjugate parameter. Properties of the dual distribution thus determined are investigated. In particular, the symmetrical role of parameter and constraint as a pair of conjugate variables is shown to imply an inherent uncertainty principle. Possible applications to temperature fluctuations and to an imbedding of classical mechanics in a statistical background are indicated.

PACS numbers: 02.50. + s, 03.20. + i, 05.40. + j

## I. INTRODUCTION

Thermodynamics and hence statistical physics[1] introduces a set of extensive variables to characterize the state of the system. Corresponding to these is a set of conjugate intensive variables. In the maximum entropy approach[1,2] the conjugate variables are introduced as Lagrange multipliers in the procedure of seeking the constrained extremum of the entropy. The two sets of variables do not appear therefore to be on equal footing. A glaring example of this "asymmetry" is that given the mean value of an extensive variable, the theory clearly predicts that fluctuations in that variable are possible. [For example, given the mean energy, we generate a distribution of energy (cf. Sec. II below) and hence can compute the variance of the energy which is closely related to the specific heat.] Yet, given a mean value of an extensive variable, the existing theory assigns a unique numerical value to the conjugate Lagrange multiplier and does not appear to recognize the possibility of fluctuation about that value.

One can, of course, take the stand that the symmetry between the two possible sets of variables is guaranteed in classical thermodynamics by the well-understood changes of variables via the Legendre transform.[3] It is clearly desirable however to trace this symmetry to the fundamental theory. Furthermore, the maximum entropy formalism is being extensively applied[4] to the description of collisions of composite projectiles (be they nuclei or molecules) and to other areas of statistical physics (e.g., irreversible processes,[5] statistical optics[6]) where there is no corresponding phenomenological thermodynamics.

A technical resolution of the problem is to proceed not via the maximum entropy formalism but via a classical Bayesian approach.[6–8] There, the problem of determining the Lagrange multiplier becomes one of parameter estimation as is discussed in Sec. II. The problem is then that the two routes do not necessarily coincide. The conditions under which they do are determined in Sec. III.

The result of our considerations is the characterization of a unique dual distribution: the distribution of the value of the Lagrange multiplier given the mean value of the constraint. Some properties of this distribution are explored in Sec. IV. Particular attention is given therein to the "uncertainty relation" between the pair of conjugate extensive and intensive variables. Generalizations to several variables are provided in Sec. V. We conclude with potential applications to physics in Sec. VI.

## II. BACKGROUND

Suppose we are given a vessel containing $N$ ideal gas molecules in thermal equilibrium. We know that the energies of the molecules are distributed according to the Boltzmann law

$$f(E\,|\,\beta) = \frac{\Omega(E)e^{-\beta E}}{z(\beta)}, \quad z(\beta) = \int \Omega(E)e^{-\beta E}\,dE\,, \qquad (1)$$

but we do not know the temperature $T = 1/\beta$. We are allowed to pierce a hole in the vessel and let $n \ll N$ molecules escape, meanwhile recording their energies. Given the evidence $E_1,...,E_n$, what value should we assign to the unknown parameter $\beta$ and how reliable should this assignment be considered?

The problem just described, namely, parameter estimation, is basic to statistical theory. Given the outcome $x_1,...,x_n$ for a random variable $X$ distributed according to

$$f(x|\lambda) = \frac{\Omega(x)e^{-\lambda A(x)}}{z(\lambda)}, \quad z(\lambda) = \int \Omega(x)e^{-\lambda A(x)}\,dx\,, \qquad (2)$$

what is the distribution $\bar{P}(\lambda\,|x_1,...,x_n)$ and what is the best guess $\hat{\lambda} = \hat{\lambda}(x_1,...,x_n)$ for the unknown parameter $\lambda$. In the (now, generally accepted) Bayesian approach,[7] one proceeds as follows.[8]

(a) Choose a "prior" or "marginal" distribution $\bar{f}_0(\lambda)$. The distribution inverse to the "sample distribution"

$$P(x_1,...,x_n\,|\lambda) = \prod_{i=1}^{n} f(x_i\,|\lambda)\,, \qquad (3)$$

---

[a] Permanent address: Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel.

  

is given by Bayes' rule

$$\bar{P}(\lambda|x_1,...,x_n)=\frac{P(x_1,...,x_n|\lambda)\bar{f}_0(\lambda)}{\int P(x_1,...,x_n|\lambda)\bar{f}_0(\lambda)d\lambda}=\frac{\bar{m}(\lambda)e^{-\lambda n\bar{A}}}{\bar{Z}(\bar{A})}. \tag{4a}$$

Here

$$\bar{A}=\frac{1}{n}\sum_{i=1}^{n}A(x_i) \tag{4b}$$

is the "sample average,"

$$\bar{m}(\lambda)\propto\bar{f}_0(\lambda)/z^n(\lambda) \tag{4c}$$

is the "density of states" for the parameter $\lambda$, and

$$\bar{Z}(\bar{A})=\int\bar{m}(\lambda)e^{-\lambda\bar{A}}d\lambda \tag{4d}$$

is the "partition function." Note that the distribution of $\lambda$ given the sample $x_1,...,x_n$, is completely determined [once $\bar{f}_0(\lambda)$ has been chosen] by the value of the "sufficient statistic"[9] $\bar{A}(x_1,...,x_n)$. That is, all the information relevant to the distribution of $\lambda$ obtained by sampling can be summarized by a single number—the sample average $\bar{A}$. We shall henceforth denote the distribution (4) by $\bar{P}(\lambda|\bar{A})$.

(b) In order to determine the "best estimate" $\hat{\lambda}$ for the parameter $\lambda$, choose a non-negative "loss function" $L(\lambda,\hat{\lambda})$ and determine the best estimate $\hat{\lambda}=\hat{\lambda}(\bar{A})$ by minimizing the "average loss"

$$R(\hat{\lambda})=\int L(\lambda,\hat{\lambda})\bar{P}(\lambda|\bar{A})d\lambda \tag{5}$$

over $\hat{\lambda}$. For example, by choosing

$$L(\lambda,\hat{\lambda})=(\lambda-\hat{\lambda})^2, \tag{6a}$$

one obtains

$$\hat{\lambda}(\bar{A})=\int\lambda\bar{P}(\lambda|\bar{A})d\lambda\equiv\langle\lambda|\bar{A}\rangle. \tag{6b}$$

Similarly, the choice $L(\lambda,\hat{\lambda})=|\lambda-\hat{\lambda}|$ leads to

$$\hat{\lambda}(\bar{A})=\text{median of }\bar{P}(\lambda|\bar{A})$$

as the best estimate for $\lambda$. To ensure uniqueness of the minimum (or infimum), $R(\hat{\lambda})$ is made convex by requiring $L(\lambda,\hat{\lambda})$ to be convex in $\hat{\lambda}$ for all $\lambda$.

The shortcoming of the above procedure stems from its indeterminate nature. Two non-negative functions, namely $\bar{f}_0(\lambda)$ and $L(\lambda,\hat{\lambda})$, are to be chosen almost freely. Can one somehow narrow the choice? It is our intention to demonstrate that this is indeed the case. We shall show in Sec. III that requirements of consistency lead to a unique choice for the loss function $L(\lambda,\hat{\lambda})$ and the prior $\bar{f}_0(\lambda)$, at least for distributions of the "exponential type"[8] of which Eqs. (1)–(4) are examples. Thus, to every distribution $f(x|\lambda)$ of the form (2) there corresponds a unique *dual* distribution

$$\bar{f}(\lambda|A(x))=\frac{\bar{\Omega}(\lambda)e^{-\lambda A(x)}}{\bar{z}(A(x))},\quad \bar{z}(A)=\int\bar{\Omega}(\lambda)e^{-\lambda A}d\lambda, \tag{7a}$$

with a density of states

$$\bar{\Omega}(\lambda)\propto\bar{f}_0(\lambda)/z(\lambda). \tag{7b}$$

Owing to the reflexive property of duality (which we shall prove), $f(x|\lambda)$ is the dual distribution to $\bar{f}(\lambda|A(x))$.

In Sec. IV some further properties of the dual distribution are explored. It is shown that the random variable $A(x)$ and its conjugate parameter $\lambda$ satisfy an inherent uncertainty relation. Next we discuss the connections between samples of size $n$ and samples of size 1. We show that sampling of the variable $X$ induces a corresponding sampling in the dual space $\lambda$ in such a way that only the sufficient statistic $\bar{\lambda}=(1/n)\sum\lambda_i$ (but not the individual $\lambda_i$'s) is observable.

## III. THE DUAL DISTRIBUTION

In this section we confine ourselves to samples of size 1. Hence only $f(x|\lambda)$ and $\bar{f}(\lambda|A(x))$ [Eqs. (2) and (7)] enter.

### A. Determination of the loss function

Having made a choice for the density of states $\bar{\Omega}(\lambda)$ and the loss function $L(\lambda,\hat{\lambda})$, the best estimate $\hat{\lambda}=\hat{\lambda}(A(x))$ is determined by solving

$$R'(\hat{\lambda})=\int\frac{\partial L}{\partial\hat{\lambda}}(\lambda,\hat{\lambda})\bar{f}(\lambda|A)d\lambda=\left\langle\frac{\partial L}{\partial\hat{\lambda}}\right\rangle=0. \tag{8}$$

But given the average $\langle(\partial L/\partial\hat{\lambda})(\lambda,\hat{\lambda})\rangle$ the principle of maximum entropy[2,7] predicts

$$F(\lambda|A(x))=\frac{\bar{\Omega}(\lambda)\exp[-\mu(\partial L/\partial\hat{\lambda})(\lambda,\hat{\lambda})]}{Z(\mu,\hat{\lambda})}, \tag{9a}$$

where

$$Z(\mu,\hat{\lambda})=\int\bar{\Omega}(\lambda)\exp\left(-\mu\frac{\partial L}{\partial\hat{\lambda}}(\lambda,\hat{\lambda})\right)d\lambda. \tag{9b}$$

Here $\mu$ is a Lagrange multiplier whose value $\mu=\mu(\hat{\lambda}(A))$ is determined by solving

$$0=\left\langle\frac{\partial L}{\partial\hat{\lambda}}\right\rangle=-\frac{\partial\log Z(\mu,\hat{\lambda})}{\partial\mu}. \tag{10}$$

We now have *two predictions* [for the same data $A(x)$!], namely,

$$\bar{f}(\lambda|A)=\Omega(\lambda)e^{-\lambda A}/z(A)\quad\text{(from Bayes rule)}, \tag{11a}$$

and

$$F(\lambda|A)=\frac{\bar{\Omega}(\lambda)\exp[-\mu(\partial L/\partial\hat{\lambda})(\lambda,\hat{\lambda})]}{Z(\mu,\hat{\lambda})}\quad\text{(from maximum entropy)}. \tag{11b}$$

*Proposition*: The two predictions (11a) and (11b) coincide if and only if the loss function is quadratic. That is, $L(\lambda,\hat{\lambda})=c(\lambda-\hat{\lambda})^2$, where $c>0$ is constant. Indeed, if $L=c(\lambda-\hat{\lambda})^2$ then $\langle\partial L/\partial\hat{\lambda}\rangle=-2c\langle\lambda-\hat{\lambda}\rangle=0$ implies

$$\hat{\lambda}=\langle\lambda|A\rangle, \tag{12}$$

and

$$Z=\int\bar{\Omega}(\lambda)e^{2\mu c(\lambda-\hat{\lambda})}d\lambda=e^{-2\mu c\hat{\lambda}}\bar{z}(-2\mu c). \tag{13}$$

Applying Eq. (10) to the last expression, we obtain

$$0=\left\langle\frac{\partial L}{\partial\hat{\lambda}}\right\rangle=-\frac{\partial\log Z}{\partial\mu}=2c\hat{\lambda}+2c\frac{\partial\log\bar{z}(-2\mu c)}{\partial A}. \tag{14}$$

Hence, with the aid of Eqs. (12) and (7a),

$$\hat{\lambda} = -\frac{\partial \log \bar{z}(-2\mu c)}{\partial A} = \langle \lambda | A \rangle = -\frac{\partial \log \bar{z}(A)}{\partial A}. \tag{15}$$

Now

$$\frac{\partial^2 \log \bar{z}(A)}{\partial A^2} = \langle (\lambda - \langle \lambda \rangle)^2 \rangle \geqslant 0. \tag{16}$$

Hence $-(\partial \log \bar{z}(A)/\partial A)$ is monotonic and by Eqs. (15) and (11) we have $-2\mu c = A$ and $F(\lambda | A) = \bar{f}(\lambda | A)$. Thus $L = c(\lambda - \hat{\lambda})^2$ is *sufficient* to establish harmony between the predictions of maximum entropy and Bayes' rule.

In order to prove that the condition is also *necessary*, we shall somewhat restrict the choice of loss functions. We assume (in addition to non-negativity and convexity) that $L(\lambda,\hat{\lambda}) = L(\lambda - \hat{\lambda})$ is a function of the difference $(\lambda - \hat{\lambda})$ only, satisfying $L(0) = 0$. Now $F(\lambda | A) = \bar{f}(\lambda | A)$ implies

$$-\mu \frac{\partial L}{\partial \hat{\lambda}} = -\lambda A + \log Z - \log \bar{z}(A). \tag{17}$$

Invoking Eq. (10), we have

$$-\mu \left\langle \frac{\partial L}{\partial \hat{\lambda}} \right\rangle = 0 = -\langle \lambda \rangle A + \log Z - \log \bar{z}(A), \tag{18}$$

and Eq. (17) reduces to

$$\frac{\partial L}{\partial \hat{\lambda}} = \frac{A}{\mu}(\lambda - \langle \lambda \rangle). \tag{19}$$

The last equation can be viewed either as an explicit expression for the Lagrange parameter $\mu$, or as a condition satisfied by the loss function $L(\lambda - \hat{\lambda})$. Taking the latter point of view, we have

$$\frac{\partial L}{\partial \lambda} = -\frac{\partial L}{\partial \hat{\lambda}} = -\frac{A}{\mu}(\lambda - \langle \lambda \rangle), \tag{20}$$

hence, by integration,

$$L = -(A/\mu)(\lambda^2/2 - \langle \lambda \rangle \lambda) + h(\hat{\lambda}). \tag{21}$$

Here $\mu = \mu(\hat{\lambda}(A))$ and $\langle \lambda \rangle = -(\partial \log \bar{z}(A)/\partial A)$ are functions of $A$. Inverting $A = A(\hat{\lambda})$ (which is certainly valid for some range of $\hat{\lambda}$), and taking the derivative of (21) with respect to $\hat{\lambda}$, we have

$$\begin{aligned}\frac{\partial L}{\partial \hat{\lambda}} &= -\frac{d}{d\hat{\lambda}}\left(\frac{A}{\mu}\right)\left(\frac{\lambda^2}{2} - \langle \lambda \rangle \lambda\right) \\ &\quad + \left(\frac{A}{\mu}\right)\lambda \frac{d\langle \lambda \rangle}{d\hat{\lambda}} + h'(\hat{\lambda}) \\ &= (A/\mu)(\lambda - \langle \lambda \rangle).\end{aligned} \tag{22}$$

Comparing equal powers of $\lambda$, we secure

$$A/\mu = -D, \tag{23a}$$

where $D$ is a constant,

$$\frac{d\langle \lambda \rangle}{d\hat{\lambda}} = 1, \tag{23b}$$

$$h'(\hat{\lambda}) = D\langle \lambda \rangle. \tag{23c}$$

Hence

$$\langle \lambda \rangle = \hat{\lambda} + G, \tag{24a}$$

and

$$h(\hat{\lambda}) = D(\hat{\lambda}^2/2 + G\hat{\lambda}) + H, \tag{24b}$$

where $G$ and $H$ are constants. Inserting these expressions into (21), we obtain

$$\begin{aligned} L &= (D/2)(\lambda - \hat{\lambda})^2 - DG(\lambda - \hat{\lambda}) + H \\ &= (D/2)(\lambda - \langle \lambda \rangle) - DG^2/2 + H. \end{aligned} \tag{25}$$

The constants $G$ and $H$ are now determined by use of the assumptions $L(0) = 0$, $\partial^2 L/\partial \hat{\lambda}^2 = D > 0$ and $L \geqslant 0$. Putting $\lambda = \hat{\lambda}$ we have $H = 0$. Again, putting $\lambda = \langle \lambda \rangle$ we obtain $L = -(D/2)G^2 \geqslant 0$, which can be satisfied only by $G = 0$. Thus,

$$\hat{\lambda} = \langle \lambda \rangle, \tag{26}$$

and

$$L = (D/2)(\lambda - \hat{\lambda})^2, \quad D > 0. \tag{27}$$

Incidentally, by Eq. (26) we have

$$\frac{d\hat{\lambda}}{dA} = \frac{d\langle \lambda \rangle}{dA} = -\frac{\partial^2 \log \bar{z}(A)}{\partial A^2} = -\mathrm{var}(A) \leqslant 0.$$

Hence, $\hat{\lambda}(A)$ is a monotonic function of $A$ and the inversion $A = A(\hat{\lambda})$ is valid for all $\hat{\lambda}$. Note that the class of loss functions for which our proof applies, could be enlarged to include $L(\lambda,\hat{\lambda}) = g(\lambda)\, l(\lambda - \hat{\lambda})$, where $g(\lambda) > 0$ could be absorbed in the yet undetermined density of states $\bar{\Omega}(\lambda)$. Having established $l = c(\lambda - \lambda)^2$ as the only loss function which brings harmony between the predictions of maximum entropy and Bayes' rule, we shall now turn to determine $\bar{\Omega}(\lambda)$, or, equivalently, the prior distribution for $\lambda$.

### B. Determination of the density of states

Observing an outcome $A(x)$, our best estimate for $\lambda$ is $\hat{\lambda}(A) = \langle \lambda | A \rangle$, but given $\hat{\lambda}$, our best estimate for the parameter $A(x)$ in Eq. (7a) is $\hat{A}(\hat{\lambda}) = \langle A | \hat{\lambda} \rangle$. We now demand self-consistency: *the best estimate of A given A is A*. That is,

$$\hat{A}(\hat{\lambda}) = \langle A | \hat{\lambda} \rangle = -\frac{\partial \log z(\hat{\lambda})}{\partial \hat{\lambda}} = A, \tag{28a}$$

where

$$\hat{\lambda} = \langle \lambda | A \rangle = -\frac{\partial \log \bar{z}(A)}{\partial A}. \tag{28b}$$

The last equality in Eq. (28a) can be interpreted in a slightly different way. The estimate $\hat{\lambda}$ (given $A$), determines an estimate for the average $\langle A \rangle$ via

$$\langle \hat{A} \rangle = -\frac{\partial \log z(\hat{\lambda})}{\partial \hat{\lambda}}. \tag{29}$$

Demanding that the *best estimate for* $\langle A \rangle$ *given A is A*, we have

$$\langle \hat{A} \rangle = -\frac{\partial \log z(\hat{\lambda})}{\partial \hat{\lambda}} = A. \tag{30}$$

We shall now show that the requirement of self-consistency [Eqs. (28)], is enough to determine the partition function $\bar{z}(A)$ and the density of states $\bar{\Omega}(\lambda)$ uniquely (up to an irrelevant factor). Indeed, determining $\hat{\lambda}(A)$ as the (unique) solution of Eq. (28a), condition (28b) serves as a differential equation for the unknown function $\bar{z}(A)$. Let $\tilde{z}(A)$ be such that $-\log \tilde{z}(A)$ is the Legendre transform[3,10] of $-\log z(\lambda)$, that is,

$$-\log \tilde{z}(A) = \tilde{\lambda} A - [-\log z(\tilde{\lambda})], \tag{31}$$

where $\tilde{\lambda} = \tilde{\lambda}(A)$ is the (unique) solution of

$$A = -\frac{\partial \log z(\tilde{\lambda})}{\partial \tilde{\lambda}}. \quad (32)$$

But by Eqs. (31), (32), and (28),

$$-\frac{\partial \log \tilde{z}(A)}{\partial A} = \tilde{\lambda} + A\frac{d\tilde{\lambda}}{dA} + \frac{\partial \log z(\tilde{\lambda})}{\partial \tilde{\lambda}}\frac{d\tilde{\lambda}}{dA} = \tilde{\lambda} = \hat{\lambda}, \quad (33)$$

hence

$$-\frac{\partial \log \tilde{z}(A)}{\partial A} = -\frac{\partial \log \bar{z}(A)}{\partial A}, \quad (34)$$

and

$$\bar{z}(A) = C\tilde{z}(A). \quad (35)$$

Since the constant $C$ is irrelevant, we shall standardize the solution $\bar{z}(A)$ by adopting $C = 1$, that is

$$-\log \bar{z}(A) = \langle \lambda \rangle A + \log z(\langle \lambda \rangle), \quad (36a)$$

or

$$\bar{z}(A) = e^{-\langle \lambda \rangle A}/z(\langle \lambda \rangle), \quad (36b)$$

where $\langle \lambda \rangle$ is the solution of (32). Finally, the integral equation

$$\bar{z}(A) = \int \bar{\Omega}(\lambda) e^{-\lambda A}\, dx \quad (37)$$

determines (under broad conditions) a unique solution for the density of states $\bar{\Omega}(\lambda)$, given the "moment generating function"[8] $\bar{z}(A)$. In summary, given the distribution

$$f(x|\langle \lambda \rangle) = \frac{\Omega(x) e^{-\langle \lambda \rangle A(x)}}{z(\langle \lambda \rangle)}, \quad \langle A \rangle = -\frac{\partial \log z(\langle \lambda \rangle)}{\partial \langle \lambda \rangle}, \quad (38)$$

a *dual* distribution

$$\bar{f}(\lambda|\langle A \rangle) = \frac{\bar{\Omega}(\lambda) e^{-\lambda \langle A \rangle}}{\bar{z}(\langle A \rangle)}, \quad \langle \lambda \rangle = -\frac{\partial \log \bar{z}(\langle A \rangle)}{\partial \langle A \rangle} \quad (39)$$

is uniquely determined via the Legendre transform of $-\log z(\langle \lambda \rangle)$. Since the Legendre transform is a reflexive one [that is, $-\log z(\langle \lambda \rangle)$ is the transform of $-\log \bar{z}(\langle A \rangle)$], $f(x|\langle \lambda \rangle)$ is the dual distribution to $\bar{f}(\lambda|\langle A \rangle)$.

### C. Examples

By way of illustration, consider the following two examples.

(a) The Maxwell–Boltzmann distribution for the energy of ideal gas molecules is

$$f(E|\langle \beta \rangle) = (2/\sqrt{\pi})\langle \beta \rangle^{3/2}\sqrt{E}\, e^{-\langle \beta \rangle E}, \quad 0 \leqslant E < \infty. \quad (40a)$$

Here

$$\Omega(E) = C\sqrt{E}, \quad (40b)$$

and

$$z(\langle \beta \rangle) = C\tfrac{1}{2}\sqrt{\pi}\langle \beta \rangle^{-3/2}, \quad (40c)$$

where $C$ is a constant. Equation (38) yields

$$\langle \beta \rangle = 3/2\langle E \rangle. \quad (41)$$

Hence, substituting in Eq. (36), we obtain

$$\bar{z}(\langle E \rangle) = \frac{2}{\sqrt{\pi}}\frac{1}{C}\left(\frac{3}{2e\langle E \rangle}\right)^{3/2}. \quad (42a)$$

The integral equation (37) is now solved by

$$\bar{\Omega}(\beta) = \frac{4}{\pi}\frac{1}{C}\left(\frac{3}{2e}\right)^{3/2}\sqrt{\beta}, \quad (42b)$$

and the distribution dual to (40a) is

$$\bar{f}(\beta|\langle E \rangle) = (2/\sqrt{\pi})\langle E \rangle^{3/2}\sqrt{\beta}\, e^{-\langle E \rangle \beta}, \quad 0 \leqslant \beta < \infty. \quad (42c)$$

Using Eqs. (7b), (42b), and (40c) we can also determine the prior $\bar{f}_0(\beta)$, namely,

$$\bar{f}_0(\beta) \propto \bar{\Omega}(\beta) z(\beta) \propto 1/\beta, \quad (43)$$

which is the result obtained by Jeffreys[11] for a scale parameter.

(b) As a second example we take

$$f(x|\langle \lambda \rangle) = (1/\sqrt{2\pi}) e^{-(1/2)x^2} e^{-\langle \lambda \rangle x}/z(\langle \lambda \rangle)$$
$$= (1/\sqrt{2\pi}) e^{-(1/2)(x + \langle \lambda \rangle)^2}, \quad -\infty < x < \infty. \quad (44a)$$

Here

$$\Omega(x) = (1/\sqrt{2\pi}) e^{-(1/2)x^2}, \quad z(\langle \lambda \rangle) = e^{(1/2)\langle \lambda \rangle^2}. \quad (44b)$$

Hence

$$\langle \lambda \rangle = -\langle x \rangle, \quad (45)$$

$$\bar{\Omega}(\lambda) = (1/\sqrt{2\pi}) e^{-(1/2)\lambda^2}, \quad \bar{z}(\langle x \rangle) = e^{(1/2)\langle x \rangle^2}, \quad (46a)$$

and

$$\bar{f}(\lambda|\langle x \rangle) = (1/\sqrt{2\pi}) e^{-(1/2)\lambda^2} e^{-\lambda\langle x \rangle}/\bar{z}(\langle x \rangle)$$
$$= (1/\sqrt{2\pi}) e^{-(1/2)(\lambda + \langle x \rangle)^2}, \quad -\infty < \lambda < \infty. \quad (46b)$$

In this example the prior probability is uniform:

$$\bar{f}(\lambda) \propto \bar{\Omega}(\lambda) z(\lambda) \propto 1. \quad (47)$$

## IV. PROPERTIES OF THE DUAL DISTRIBUTION

Having observed $x_1$ and hence $A_1 = A(x_1)$ and $\hat{\lambda}_1 = \hat{\lambda}(A_1) = -\partial \log \bar{z}(A_1)/\partial A_1$, we expect the *next* observation to fulfill

$$A_2 = A_1 \pm \Delta A, \quad (48)$$

where

$$\Delta A = \langle (A - \langle A \rangle)^2 \rangle^{1/2} = [\mathrm{var}(A)]^{1/2}, \quad (49)$$

and

$$\hat{\lambda}_2 = \hat{\lambda}(A_2) = \hat{\lambda}(A_1 \pm \Delta A) \approx \hat{\lambda}(A_1) \pm \frac{\partial \hat{\lambda}}{\partial A}(A_1)\Delta A. \quad (50)$$

But

$$\frac{\partial \hat{\lambda}}{\partial A_1} = -\frac{\partial^2 \log \bar{z}(A_1)}{\partial A_1^2} = -\mathrm{var}(\lambda) = -(\Delta\lambda)^2. \quad (51)$$

Hence

$$\hat{\lambda}_2 \approx \hat{\lambda}_1 \mp (\Delta\lambda)^2 \Delta A. \quad (52)$$

That is $\Delta\lambda \approx (\Delta\lambda)^2 \Delta A$, or

$$\Delta\lambda(A_1)\Delta A(A_1) \approx 1. \quad (53)$$

Thus the expected uncertainty in $\lambda$ having seen $A_1$, times the expected uncertainty in $A$ having seen $A_1$, is of the order of unity.

We turn now to a more careful discussion where it proves possible (nay, essential) to distinguish between experimental and inherent uncertainties.[12]

### A. Inherent uncertainty relation

Let $\langle A \rangle$ be known and let $\Delta A(\langle A \rangle)$ and $\Delta\lambda(\langle A \rangle)$ denote the *inherent* uncertainties in $A$ and $\lambda$ given $\langle A \rangle$ (or $\langle \lambda \rangle$). Then

$$\langle (A - \langle A \rangle)^2 \rangle = (\Delta A)^2 = \frac{\partial^2 \log z(\langle \lambda \rangle)}{\partial \langle \lambda \rangle^2} = -\frac{\partial \langle A \rangle}{\partial \langle \lambda \rangle}, \tag{54}$$

and

$$\langle (\lambda - \langle \lambda \rangle)^2 \rangle = (\Delta \lambda)^2 = \frac{\partial^2 \log \bar{z}(\langle A \rangle)}{\partial \langle A \rangle^2} = -\frac{\partial \langle \lambda \rangle}{\partial \langle A \rangle}. \tag{55}$$

Hence

$$(\Delta A)^2 (\Delta \lambda)^2 = -\frac{\partial \langle A \rangle}{\partial \langle \lambda \rangle}\left(-\frac{\partial \langle \lambda \rangle}{\partial \langle A \rangle}\right) = 1 .$$

That is,

$$\Delta A(\langle A \rangle) \Delta \lambda(\langle A \rangle) = 1 . \tag{56}$$

The inherent uncertainties are related as above *regardless of the accuracy by which* $\langle A \rangle$ *is known*. Of course, the individual uncertainties $\Delta A$ and $\Delta \lambda$ *are* dependent on the accuracy of $\langle A \rangle$. The better we know $\langle A \rangle$, the better are our estimates for $\Delta A$ and $\Delta \lambda$. This leads us to discuss the accuracy of the estimation that is, the relation between samples of size 1 and samples of size $n > 1$.

### B. Accuracy of the estimation, induced sampling in the dual space

Given a distribution

$$f(x|\lambda) = \Omega(x) e^{-\lambda A(x)}/z(\lambda), \tag{57}$$

the sample distribution is

$$P(x_1,...,x_n|\lambda) = \Pi f(x_i|\lambda) = \frac{\Pi\, \Omega(x_i) \exp[-\lambda \Sigma A(x_i)]}{z^n(\lambda)}. \tag{58}$$

With the aid of $P$, the distribution of the sample average

$$\bar{A} = \frac{1}{n}\sum A(x_i), \tag{59}$$

can be expressed as

$$P(\bar{A}|\lambda) = \int P(x_1,...,x_n|\lambda) \times \delta\left(\frac{1}{n}\sum A(x_i) - \bar{A}\right) dx_1 \cdots dx_n = \frac{m(\bar{A}) e^{-\lambda n \bar{A}}}{Z(\lambda)}, \tag{60a}$$

where

$$m(\bar{A}) = \int \Omega(x_1) \cdots \Omega(x_n) \times \delta\left(\frac{1}{n}\sum A(x_i) - \bar{A}\right) dx_1 \cdots dx_n, \tag{60b}$$

and

$$Z(\lambda) = z^n(\lambda). \tag{60c}$$

It is easy to show that the following relations hold[8]:

$$\langle \bar{A} \rangle_P = \langle A \rangle_f, \tag{61a}$$

$$\mathrm{var}_P(\bar{A}) = (1/n)\mathrm{var}_f(A), \tag{61b}$$

hence

$$\left(\frac{\Delta \bar{A}}{\langle \bar{A} \rangle}\right)_P = \frac{1}{\sqrt{n}}\left(\frac{\Delta A}{\langle A \rangle}\right)_f. \tag{61c}$$

The last result can be regarded as a form of the law of large numbers[8]: the larger the sample the better is our estimate $\bar{A}$ for the average $\langle A \rangle_f$.

Consider now the distribution $\bar{P}(\lambda|\bar{A})$ dual to $P(\bar{A}|\lambda)$:

$$\bar{P}(\lambda|\bar{A}) = \bar{m}(\lambda) e^{-\lambda n \bar{A}}/\bar{Z}(\bar{A}), \tag{62a}$$

where

$$\bar{Z}(\bar{A}) = \frac{e^{-\tilde{\lambda}\bar{A}}}{Z(\tilde{\lambda})} = \int \bar{m}(\lambda) e^{-\lambda n \bar{A}} d\lambda, \tag{62b}$$

and $\tilde{\lambda}$ is the solution of

$$\bar{A} = -\frac{\partial \log Z(\tilde{\lambda})}{\partial (n\tilde{\lambda})}. \tag{62c}$$

But $Z(\tilde{\lambda}) = z^n(\tilde{\lambda})$ implies

$$\bar{A} = -\frac{\partial \log z^n(\tilde{\lambda})}{\partial (n\tilde{\lambda})} = -\frac{\partial \log z(\tilde{\lambda})}{\partial \tilde{\lambda}}. \tag{63}$$

Hence

$$\tilde{\lambda} = \langle \lambda | \bar{A} \rangle_{\bar{f}} = -\frac{\partial \log \bar{z}(\bar{A})}{\partial \bar{A}}$$

and

$$\bar{Z}(\bar{A}) = [e^{-\tilde{\lambda}\bar{A}}/z(\tilde{\lambda})]^n = \bar{z}^n(\bar{A}). \tag{64}$$

Thus, given the sample average $\bar{A}$, we have

$$\langle \lambda \rangle_{\bar{P}} = -\frac{\partial \log Z(\bar{A})}{\partial (n\bar{A})} = -\frac{\partial \log \bar{z}(\bar{A})}{\partial \bar{A}} = \langle \lambda \rangle_{\bar{f}}, \tag{65a}$$

$$\mathrm{var}_{\bar{P}}(\lambda) = \frac{\partial^2 \log Z(\bar{A})}{\partial (n\bar{A})^2} = \frac{1}{n}\frac{\partial^2 \log \bar{z}(\bar{A})}{\partial \bar{A}^2} = \frac{1}{n}\mathrm{var}_{\bar{f}}(\lambda), \tag{65b}$$

and therefore

$$\left(\frac{\Delta \lambda}{\langle \lambda \rangle}\right)_{\bar{P}} = \frac{1}{\sqrt{n}}\left(\frac{\Delta \lambda}{\langle \lambda \rangle}\right)_{\bar{f}}. \tag{65c}$$

The property (65a) *characterizes* the dual distribution $\bar{P}(\lambda|\bar{A})$. Any Bayes' distribution $\bar{P}(\lambda|\bar{A})$ satisfying (65a) is necessarily the dual distribution to $P(\bar{A}|\lambda)$. Indeed,

$$\langle \lambda|\bar{A} \rangle_{\bar{P}} = -\frac{\partial \log \bar{Z}(\bar{A})}{\partial (n\bar{A})} = \langle \lambda|\bar{A} \rangle_{\bar{f}} = -\frac{\partial \log \bar{z}(\bar{A})}{\partial \bar{A}}$$

implies $\bar{Z}(\bar{A}) = \mathrm{const} \cdot \bar{z}^n(\bar{A})$. Property (65b) and its counterpart (61b) allow us to connect the uncertainty product for a sample of size $n$ to the corresponding product for a sample of size one, namely,

$$[\Delta \lambda(\bar{A})]_{\bar{P}} [\Delta \bar{A}(\bar{A})]_P = \frac{1}{\sqrt{n}}[\Delta \lambda(\bar{A})]_{\bar{f}} \frac{1}{\sqrt{n}}[\Delta A(\bar{A})]_f .$$

In view of Eq. (56), we secure

$$[\Delta\lambda(\bar{\bar{A}})]_{\bar{P}}[\Delta\bar{\bar{A}}(\bar{A})]_P = 1/n. \quad (66)$$

The two "uncertainty products" [Eqs. (56) and (66)] have quite different meanings. Equation (56) is the inherent uncertainty: However well we know the mean value $\langle A\rangle$ of $A$, there will be a finite variance $\Delta A$ and a finite variance $\Delta\lambda$ and no further measurement can reduce their product below unity. Equation (66) deals with a more mundane aspect: the variance of our estimate of $\langle A\rangle$. It is a purely experimental uncertainty and does therefore reduce [cf. (61)] as more measurements are being made. Now, $\Delta\bar{\bar{A}}$ is what the experimentalist reports as his estimate for the uncertainty in the measured mean value of $A$. Often, of course, one does not estimate $\lambda$ for each measured value of $A$ but rather reports only $\bar{A}$ and $\Delta\bar{\bar{A}}$ from which $\lambda$ and $\Delta\lambda$ are to be computed. In that case the experimental uncertainties satisfy (66) with $n = 1$. Note however that even when many measurements are made so that the experimental uncertainties are quite small [i.e., $n$ in (66) is large], the inherent uncertainties continue to satisfy (56). As we said in the beginning of this paragraph, $\Delta A$, defined by (54) is an inherent variance [of the distribution $f(x|\lambda)$] and is quite distinct from $\Delta\bar{\bar{A}}$, the uncertainty of our estimate for $\langle A\rangle$.

Let us now calculate the density of states $\bar{m}(\lambda)$. From Eqs. (64) and (7a), we have

$$\begin{aligned}\bar{Z}(\bar{\bar{A}}) &= \bar{z}^n(\bar{\bar{A}})\\ &= \int \bar{\Omega}(\lambda_1)e^{-\lambda_1\bar{\bar{A}}}d\lambda_1\cdots\int\bar{\Omega}(\lambda_n)e^{-\lambda_n\bar{\bar{A}}}d\lambda_n\\ &= \int d\lambda\, e^{-n\lambda\bar{\bar{A}}}\int\bar{\Omega}(\lambda_1)\cdots\bar{\Omega}(\lambda_n)\\ &\quad\times\delta\left(\frac{1}{n}\sum\lambda_i-\lambda\right)d\lambda_1\cdots d\lambda_n. \end{aligned} \quad (67)$$

Since $\bar{Z}(\bar{\bar{A}})$ determines $\bar{m}(\lambda)$ uniquely, we obtain, comparing Eq. (67) with Eq. (62b),

$$\bar{m}(\bar{\lambda}) = \int\bar{\Omega}(\lambda_1)\cdots\bar{\Omega}(\lambda_n)\times\delta\left(\frac{1}{n}\sum\lambda_i-\bar{\lambda}\right)d\lambda_1\cdots d\lambda_n, \quad (68)$$

which is the exact counterpart to (60b). The last result suggests that Eqs. (65) should be rewritten, in analogy to Eqs. (61), as

$$\langle\bar{\lambda}\rangle_{\bar{P}} = \langle\lambda\rangle_{\bar{f}}, \quad (65a')$$

$$\text{var}_{\bar{P}}(\bar{\lambda}) = (1/n)\text{var}_{\bar{f}}(\lambda), \quad (65b')$$

and

$$\left(\frac{\Delta\bar{\lambda}}{\langle\bar{\lambda}\rangle}\right)_{\bar{P}} = \frac{1}{\sqrt{n}}\left(\frac{\Delta\lambda}{\langle\lambda\rangle}\right)_{\bar{f}}. \quad (65c')$$

We can summarize the structure revealed by Eqs. (57)–(68) as follows. The sampling of the random variable $X$ has induced a corresponding sampling in the dual space $\lambda$, with all the properties one usually associates with a sample. The sample distribution in the dual space is given by

$$\bar{P}(\lambda_1,\dots,\lambda_n|\bar{\bar{A}}) = \prod\bar{f}(\lambda_i|\bar{\bar{A}}) = \frac{\Pi\bar{\Omega}(\lambda_i)e^{-\bar{\bar{A}}\Sigma\lambda_i}}{\bar{z}^n(\bar{\bar{A}})}. \quad (69)$$

Hence the sample average

$$\bar{\lambda} = (1/n)\sum\lambda_i \quad (70)$$

is distributed according to

$$\bar{P}(\bar{\lambda}|\bar{\bar{A}}) = \bar{m}(\bar{\lambda})e^{-n\bar{\lambda}\bar{\bar{A}}}/\bar{z}^n(\bar{\bar{A}}), \quad (62a')$$

with $\bar{m}(\bar{\lambda})$ given by Eq. (68). The quantity $\bar{\lambda}$ serves as a sufficient statistic for the parameter $\bar{\bar{A}}$ in Eq. (62a), that is, any Bayes' inverse distribution

$$\begin{aligned}P(\bar{\bar{A}}|\lambda_1,\dots,\lambda_n) &= \frac{\bar{P}(\lambda_1,\dots,\lambda_n|\bar{\bar{A}})f_0(\bar{\bar{A}})}{\int\bar{P}(\lambda_1,\dots,\lambda_n|\bar{\bar{A}})f_0(\bar{\bar{A}})d\bar{\bar{A}}}\\ &= \frac{m(\bar{\bar{A}})e^{-n\bar{\lambda}\bar{\bar{A}}}}{Z(\bar{\lambda})} \equiv P(\bar{\bar{A}}|\bar{\lambda})\end{aligned}$$

is completely determined by the single number $\bar{\lambda}$. In particular, the distribution dual to $\bar{P}(\bar{\lambda}|\bar{\bar{A}})$ is $P(\bar{\bar{A}}|\bar{\lambda})$, where $Z(\bar{\lambda}) = z^n(\bar{\lambda})$. Note that [given $\bar{\bar{A}}(x_1,\dots,x_n)$] the individual $\lambda_1,\dots,\lambda_n$ are *not* observable (and not needed). The only observable quantity is the sufficient statistic $\bar{\lambda} = (1/n)\Sigma\lambda_i$, which *is* needed. Given an observation $\bar{\bar{A}}(x_1,\dots,x_n)$, a corresponding observation (or best guess) $\bar{\lambda}$ is formed via $\bar{\lambda} = -\partial\log\bar{z}(\bar{\bar{A}})/\partial\bar{\bar{A}}$.

We end this section with the following *conjecture*. Instead of solving directly for the distribution $\bar{P}(\bar{\bar{A}}|\lambda)$ [Eqs. (64) and (68)], we could have used the prior

$$\bar{f}_0(\lambda) \propto \bar{\Omega}(\lambda)z(\lambda)$$

obtained from the solution of Eq. (37) and determine $\bar{P}$ as the Bayes' distribution (4). Thus, we should expect the following relations to hold:

$$\frac{\bar{\Omega}(\lambda)}{z^{n-1}(\lambda)} \propto \int\bar{\Omega}(\lambda_1)\cdots\bar{\Omega}(\lambda_n)\times\delta\left(\frac{1}{n}\sum\lambda_i-\lambda\right)d\lambda_1\cdots d\lambda_n \quad (71a)$$

or, equivalently

$$\bar{Z}(\bar{\bar{A}}) = \int\frac{\bar{\Omega}(\lambda)}{z^{n-1}(\lambda)}e^{-n\lambda\bar{\bar{A}}}d\lambda \propto \bar{z}^n(\bar{\bar{A}}). \quad (71b)$$

Although all the examples checked by us do fulfill these relations, we failed to prove them.

## V. GENERALIZATION

Most of the results derived in the preceding sections for a single parameter can be generalized to the multiparameter case. Thus, given the distribution

$$f(\mathbf{x}|\lambda_1,\dots,\lambda_m) = \frac{\Omega(\mathbf{x})\exp[-\Sigma_{r=1}^m\lambda_r A_r(\mathbf{x})]}{z(\lambda_1,\dots,\lambda_m)} \quad (72)$$

(e.g., maximum entropy distribution with $m$ constraints), we can Legendre-transform any group $A_1,\dots,A_s$, $1\leqslant s\leqslant m$, to their conjugate variables $\lambda_1,\dots,\lambda_s$ and obtain the corresponding dual distribution. For example, transforming all the variables $A_1,\dots,A_m$, we have

$$\bar{f}(\lambda_1,...,\lambda_m|A_1,...,A_m) = \frac{\bar{\Omega}(\lambda_1,...,\lambda_m)\exp(-\Sigma_{r=1}^m \lambda_r A_r)}{\bar{z}(A_1,...,A_m)}, \tag{73a}$$

where

$$\bar{z}(A_1,...,A_m) = \exp\frac{(-\Sigma\tilde{\lambda}_r A_r)}{z(\tilde{\lambda}_1,...,\tilde{\lambda}_m)}, \tag{73b}$$

with $\tilde{\lambda}_1,...,\tilde{\lambda}_m$ the solution of

$$A_r = \frac{-\partial \log z(\tilde{\lambda}_1,...,\tilde{\lambda}_m)}{\partial\tilde{\lambda}_r}, \quad r = 1,...,m, \tag{73c}$$

and $\bar{\Omega}(\lambda_1,...,\lambda_m)$ the solution of

$$\bar{z}(A_1,...,A_m) = \int \bar{\Omega}(\lambda_1,...,\lambda_m)\exp\left(-\sum\lambda_r A_r\right)d\lambda_1\cdots d\lambda_m. \tag{73d}$$

To assure uniqueness of the solution $\tilde{\lambda}_1,...,\tilde{\lambda}_m$ [Eq. (73c)], we assume that the $m+1$ constraints $A_0(\mathbf{x}) = 1, A_1(\mathbf{x}),...,A_m(\mathbf{x})$ are linearly independent.[13] The uncertainty relation (56) is now replaced by

$$\Delta A_r \Delta\lambda_r \geqslant 1, \quad r = 1,...,m, \tag{74}$$

where equality holds if and only if the covariance matrix

$$C_{rs} = -\frac{\partial\langle A_r\rangle}{\partial\langle\lambda_s\rangle} = \frac{\partial^2 \log z}{\partial\langle\lambda_r\rangle\,\partial\langle\lambda_s\rangle} = \langle(A_r - \langle A_r\rangle)(A_s - \langle A_s\rangle)\rangle \tag{75}$$

is diagonal. In order to derive (74), we make use of the fact that $C$ is a positive definite symmetric matrix[13] with inverse $C^{-1} = \bar{C}$, where

$$\bar{C}_{rs} = -\frac{\partial\langle\lambda_r\rangle}{\partial\langle A_s\rangle} = \frac{\partial^2 \log \bar{z}}{\partial\langle A_r\rangle\,\partial\langle A_s\rangle} = \langle(\lambda_r - \langle\lambda_r\rangle)(\lambda_s - \langle\lambda_s\rangle)\rangle. \tag{76}$$

It is shown in the Appendix that any positive definite symmetric matrix $C$ satisfies

$$C_{rr}(C^{-1})_{rr} \geqslant 1, \tag{77}$$

with equality if and only if $C$ is diagonal. In particular, the covariance matrix fulfills

$$C_{rr}\bar{C}_{rr} = (\Delta A_r)^2(\Delta\lambda_r)^2 \geqslant 1.$$

## VI. DISCUSSION

We have seen in the preceding sections how arguments of consistency single out a unique inverse distribution dual to a given direct distribution. We also saw that the only consistent best guess for a random variable is its average. The apparently unsymmetrical role of the constraint–Lagrange multiplier has been removed and equal status has been endowed to both as conjugate variables. Is there any reflection of this symmetry in nature? It is tempting to answer in the affirmative, though lacking concrete evidence in support of such hypothesis, all that follows must be considered as directions for future research.

### A. Temperature fluctuations

At the heart of statistical mechanics lies the Boltzmann distribution

$$f(E|\langle\beta\rangle) = \Omega(E)e^{-\langle\beta\rangle E}/z(\langle\beta\rangle), \tag{78}$$

where $\langle\beta\rangle$ is identified with $1/T$, $T$ being the thermodynamic temperature determined by the second law via the efficiency of a reversible Carnot engine. Equation (78) predicts energy fluctuations

$$(\Delta E)^2 = -\frac{\partial\langle E\rangle}{\partial\langle\beta\rangle} = -\frac{\partial\langle E\rangle}{\partial T}\frac{\partial T}{\partial\langle\beta\rangle} = CT^2, \tag{79}$$

where

$$C = \frac{\partial\langle E\rangle}{\partial T} \tag{80}$$

is the heat capacity. Similarly, the dual distribution

$$\bar{f}(\beta|\langle E\rangle) = \bar{\Omega}(\beta)e^{-\beta\langle E\rangle}/\bar{z}(\langle E\rangle) \tag{81}$$

predicts "beta fluctuations"

$$(\Delta\beta)^2 = -\frac{\partial\langle\beta\rangle}{\partial\langle E\rangle} = -\frac{\partial\langle\beta\rangle}{\partial T}\frac{\partial T}{\partial\langle E\rangle} = \frac{1}{T^2C}. \tag{82}$$

If temperature fluctuations are real, then we should expect

$$\Delta\beta(T) \approx \Delta(1/T) \approx |\Delta T/T^2|, \tag{83}$$

where $\Delta T(T)$ is the inherent uncertainty in $T$ (Sec. III A). Combining with Eqs. (79) and (82) we have

$$\Delta T \approx (1/C)\Delta E. \tag{84}$$

### B. Imbedding classical mechanics in a statistical background

Consider a particle leaving $x_0 = 0$ at time $t_0 = 0$ and arriving at $x$ at the final time $t$. Let

$$A(x,t;x_0 = 0, t_0 = 0) = \int_0^t L(x,\dot{x})dt \equiv A(x,t) \tag{85}$$

denote the action for such a particle, and let

$$\bar{A}(p,t) = px - A(x,t) \tag{86}$$

denote the Legendre-transformed action. Here $x = x(p,t)$ is determined by solving

$$p = \frac{\partial A(x,t)}{\partial x}. \tag{87}$$

Similarly, the transformed action satisfies

$$x = \frac{\partial\bar{A}}{\partial p}(p,t). \tag{88}$$

Suppose that the final $x$ is not known but we are given the average $\langle x\rangle$ at the final time $t$. Furthermore, we are told that the final average $\langle p\rangle$ is related to $\langle x\rangle$ via the classical relations (87) and (88), that is,

$$\langle p\rangle = \frac{\partial A(\langle x\rangle,t)}{\partial\langle x\rangle} \quad\text{and}\quad \langle x\rangle = \frac{\partial\bar{A}(\langle p\rangle,t)}{\partial\langle p\rangle}. \tag{89}$$

What can we say about the probability density for finding the particle at time $t$ in $dx$ around $x$? Similarly, what is the probability density for an arrival with momentum $p$ in $dp$? Now, maximum entropy tells us that both distributions are of the exponential type. Moreover, in view of the symmetry between $x$ and $p$ as conjugate variables, we expect the two distributions to be dual to each other. These expectations together with relation (89) lead to

$$f(x,t\,|\,\langle p\rangle)=\frac{\Omega(x,t)\exp(\langle p\rangle x/\hbar)}{z(\langle p\rangle,t)}\,, \tag{90}$$

and

$$\bar{f}(p,t\,|\,\langle x\rangle)=\frac{\bar{\Omega}(p,t)\exp(\langle x\rangle p/\hbar)}{\bar{z}(\langle x\rangle,t)}\,, \tag{91}$$

where

$$z(\langle p\rangle,t)=\exp(\bar{A}(\langle p\rangle,t)/\hbar)\,, \tag{92}$$

and

$$\bar{z}(\langle x\rangle,t)=\exp(A(\langle x\rangle,t)/\hbar)\,. \tag{93}$$

Here $\hbar$ is an arbitrary constant having the dimensions of action. Having chosen the partition functions, the densities $\Omega(x,t)$ and $\bar{\Omega}(p,t)$ are determined by solving

$$\exp\left(\frac{\bar{A}(\langle p\rangle,t)}{\hbar}\right)=\int\Omega(x,t)\exp\left(\frac{\langle p\rangle x}{\hbar}\right)dx \tag{94}$$

and

$$\exp\left(\frac{A(\langle x\rangle,t)}{\hbar}\right)=\int\Omega(p,t)\exp\left(\frac{\langle x\rangle p}{\hbar}\right)dp\,. \tag{95}$$

For example, if the Hamiltonian is quadratic, it can be shown that

$$\Omega(x,t)\propto\exp(-A(x,t)/\hbar)\,, \tag{96}$$

and

$$\bar{\Omega}(p,t)\propto\exp(-\bar{A}(p,t)/\hbar)\,. \tag{97}$$

There is a general relation between the entropy of a distribution and the corresponding dual partition function, which we have not yet written down, namely,

$$\begin{aligned}S[f]&=-\int f(x|\lambda)\log\frac{f(x|\lambda)}{\Omega(x)}dx\\&=\int f[\lambda A(x)+\log z(\lambda)]dx\\&=\lambda\langle A\rangle+\log z(\lambda)=-\log\bar{z}(\langle A\rangle)\,.\end{aligned} \tag{98}$$

Similarly,

$$S[\bar{f}]=-\log z(\langle\lambda\rangle)\,. \tag{99}$$

In the present context, we have

$$S[f]=-(1/\hbar)A(\langle x\rangle,t)\,, \tag{100}$$

and

$$S[\bar{f}]=-(1/\hbar)\bar{A}(\langle p\rangle,t)\,. \tag{101}$$

Thus the entropy at time $t$ is proportional to the action evaluated at the average position $\langle x\rangle$. All this is, of course, reminiscent of the Feynman path integral approach to quantum mechanics. Here, however, we have outlined a possible extension of classical mechanics where the latter describes the motion of the averages $\langle x\rangle$ and $\langle p\rangle$. One can work out the details of such an extension. For example, for a free particle (starting at the origin $x_0=0$ at time $t_0=0$) one finds

$$\Delta x=(\hbar t/m)^{1/2}\,,\quad \Delta p=(m\hbar/t)^{1/2}\,, \tag{102}$$

and hence

$$\Delta v=\Delta x/t=(1/m)\Delta p\,. \tag{103}$$

Thus, the mass of a free particle plays the role of momentum fluctuation, in analogy to the role of heat capacity as the energy fluctuation. The analogy between Eqs. (84) and (103) is also striking.

## ACKNOWLEDGMENTS

We would like to thank N. Tishby and D. Shalitin for many helpful discussions. One of us (Y.T.) is grateful for the hospitality extended to him by the Fritz Haber Center.

This work was partly supported by the U.S.–Israel Binational Science Foundation (BSF), Jerusalem, Israel. The Fritz Haber Research Center is supported by the Minerva Gesellschaft für die Forschung, mbH, München, West Germany.

## APPENDIX: AN INEQUALITY FOR POSITIVE DEFINITE SYMMETRIC MATRICES

In this appendix the following inequality is proved. Any positive definite symmetric matrix $A$ satisfies

$$A_{ii}(A^{-1})_{ii}\geqslant 1\,, \tag{A1}$$

with equality if and only if $A$ is diagonal. Obviously, we expect the result to be well known, but we were not able to trace it in the literature. The present proof is due to Shalitin.[14]

Let $p$ be an orthogonal matrix diagonalizing $A$, that is,

$$pA\tilde{p}=a,\quad A=\tilde{p}ap\,,\quad A^{-1}=\tilde{p}a^{-1}p\,, \tag{A2}$$

where $a$ is diagonal. Then

$$\begin{aligned}A_{ii}A^{-1}{}_{ii}&=\sum_j\tilde{p}_{ij}a_j\,p_{ji}\sum_k\tilde{p}_{ik}a_k^{-1}p_{ki}\\&=\sum_{j,k}p_{ji}^2\,p_{ki}^2a_ja_k^{-1}\\&=\sum_{j,k}p_{ji}^2p_{ki}^2\,\frac{1}{2}(a_ja_k^{-1}+a_ka_j^{-1})\,.\end{aligned} \tag{A3}$$

But, for any positive $x$,

$$x+1/x\geqslant 2\,, \tag{A4}$$

with equality if and only if $x=1$. Hence

$$\begin{aligned}A_{ii}A^{-1}{}_{ii}&\geqslant\sum_{j,k}p_{ji}^2\,p_{ki}^2\\&=\sum_j\tilde{p}_{ij}p_{ji}\sum_k\tilde{p}_{ik}\,p_{ki}=1\,.\end{aligned} \tag{A5}$$

If $A$ is diagonal $(AA^{-1})_{ii}=A_{ii}A^{-1}{}_{ii}=1$. Conversely, if equality holds in (A5) then $A$ is diagonal. In order to see this, we may assume that the matrix $p$ groups together equal eigenvalues of $A$. That is, the diagonal matrix $a$ consists of diagonal scalar submatrices $\alpha,\beta,\ldots$, with $\alpha_i=\alpha_j$, $\beta_i=\beta_j$, etc. Let $p$ be decomposed into two parts

$$p=\Pi+\Pi'\,, \tag{A6}$$

where $\Pi$ consists of square submatrices $\Pi_\alpha$, $\Pi_\beta$,... along the main diagonal corresponding to $\alpha,\beta,\ldots$, and $\Pi'$ is the rest. If $\Pi'\neq 0$ then (A3) may be rewritten as

$$A_{ii}A^{-1}{}_{ii}=1=\sum_j p_{ji}^2\left(\sum_{k,a_k=a_j}p_{ki}^2+\sum_{k,a_k\neq a_j}p_{ki}^2\,\frac{1}{2}(a_ja_k^{-1}+a_ka_j^{-1})\right)$$

$$> \sum_j p_{ji}^2 \left( \sum_{k, a_k = a_j} p_{ki}^2 + \sum_{k, a_k \neq a_j} p_{ki}^2 \right)$$

$$= \sum_j p_{ji}^2 \sum_k p_{ki}^2 = 1 , \qquad \text{(A7)}$$

where the strict inequality

$$\tfrac{1}{2}(a_j a_k^{-1} + a_k a_j^{-1}) > 1 , \quad \text{for } a_k \neq a_j \qquad \text{(A8)}$$

has been used. Hence, $\Pi' = 0$ and each submatrix $\Pi\gamma$ satisfies

$$\widetilde{\Pi}_\gamma \gamma \Pi_\gamma = \gamma . \qquad \text{(A9)}$$

By Eq. (A2) we then have

$$A = a. \qquad \text{(A10)}$$

[1]See, for example, J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1977), Chap. 3–5 in particular.

[2]E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957). For an updated account including a wealth of applications, see *The Maximum Entropy Formalism*, Ref. 7.

[3]See, for example, H. B. Callen, *Thermodynamics* (Wiley, New York, 1960), p. 90.

[4]See, for example, Y. M. Engel and R. D. Levine, Phys. Rev. C **28**, 2321 (1983); R. D. Levine, Adv. Chem. Phys. **47**, 239 (1981); and in Ref. 7; E. Zamir, R. D. Levine, and R. B. Bernstein, Chem. Phys. **55**, 253 (1980).

[5]B. Robertson, Phys. Rev. **144**, 151 (1966); and in Ref. 7.

[6]C. E. Helstron, *Quantum Detection and Estimation Theory* (Academic, New York, 1976); B. R. Frieden, *Probability, Statistical Optics and Data Testing* (Springer, Berlin, 1982).

[7]For a sweeping though highly personal account of the history of statistical theory and the role of the Bayesian approach, see E. T. Jaynes, "Where do we stand on maximum entropy?," in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT Press, Cambridge, MA, 1979).

[8]See, for example, R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics* (Macmillan, New York, 1978).

[9]See Ref. 8 and E. B. Dynkin, Usp. Mat. Nauk **6**, 68 (1951) [Selected Transl. Math. Stat. Prob. (I.M.S.) **1**, 17 (1961)].

[10]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1950).

[11]H. Jeffreys, *Theory of Probability* (Oxford U. P., New York, 1939).

[12]Y. Alhassid and R. D. Levine, Chem. Phys. Lett. **73**, 16 (1980).

[13]Y. Alhassid, N. Agmon, and R. D. Levine, Chem. Phys. Lett. **53**, 22 (1978); J. M. Manoyan and Y. Alhasid, *ibid* **101**, 265 (1983).

[14]D. Shalitin (private communication).

# Decoupling of a system of partial difference equations with constant coefficients and application

Alain J. Phares
*Department of Physics, Villanova University, Villanova, Pennsylvania 19085*

(Received 7 December 1983; accepted for publication 15 March 1984)

Consider $D$ multi-variable functions, $A_j(\mathbf{n}), j = 1$ to $D$, where $\mathbf{n}$ stands for the evaluation point in the associated multi-dimensional space of coordinates $(n_1,n_2,...)$. Let us assume that the $A_j$'s satisfy a system of $D$ linearly coupled finite difference equations: the value of each function $A_i$ at the evaluation point $\mathbf{n}$ is given as a linear combination of the values of this function and others at shifted evaluation points. By introducing $D$ suitable generating functions, $G_j, j = 1$ to $D$, one is able to replace the $D$ coupled difference equations by a system of $D$ linear equations where the $G_j$'s play the role of the $D$ unknowns. After solving this new system of equations, it is then possible to construct a difference equation for each of the $A_j$'s relating the value of $A_i$ at the evaluation point $\mathbf{n}$ to the values of $A_i$ itself at shifted arguments. The solution of such a decoupled equation can then be handled using the multi-dimensional combinatorics function technique.

PACS numbers: 02.50. + s, 05.50. + q

## I. INTRODUCTION

A one-dimensional multi-term linear recurrence relation is a difference equation relating the value of a function $A(n)$ at point $n$ to the values of the same function at shifted arguments $(n - n_1)$, $(n - n_2)$, etc., i.e.,

$$A(n) = f_1(n)A(n - n_1) + f_2(n)A(n - n_2) + \cdots + I(n), \quad n \in \mathscr{R}. \tag{1}$$

$f_1(n), f_2(n)$ etc. and $I(n)$ are given coefficients that may depend on the evaluation point $n$. If $I(n) = 0$, the equation is said to be homogeneous, and if $I(n) \neq 0$, then the equation is said to be inhomogeneous. Equation (1) does not allow one to completely calculate $A(n)$, certain initial conditions have to be specified such as

$$A(n_{0i}) = \lambda_i, \quad n_{0i} \in \mathscr{I}. \tag{2}$$

$\mathscr{R}$ stands for the region of the one-dimensional space, where Eq. (1) holds and $\mathscr{I}$ represents the set of "boundary" points $\{n_{0i}\}$. The solution of Eq. (1) satisfying the boundary conditions (2) has been obtained in a series of articles introducing for the first time the so-called "combinatorics functions."[1] Further developments then showed the generalization of this work to multi-dimensional multi-term linear difference equations,[2]

$$A(\mathbf{n}) = f_1(\mathbf{n})A(\mathbf{n} - \mathbf{n}_1) + f_2(\mathbf{n})A(\mathbf{n} - \mathbf{n}_2) + \cdots + I(\mathbf{n}), \quad \mathbf{n} \in \mathscr{R}, \tag{3}$$

$$A(\mathbf{n}_{0i}) = \lambda_i, \quad \mathbf{n}_{0i} \in \mathscr{I}, \tag{4}$$

where $\mathbf{n}$ now represents a point in a multi-dimensional space. Applications of the one-dimensional and multi-dimensional combinatorics function technique (CFT) have shown the flexibility and advantages of the new methodology.[3] More recently, the author showed that further extension of the CFT method is possible and leads to the solutions of a system of linearly coupled difference equations.[4] However, the matrix method proposed for the coupled system,[4] although technically feasible, presents some difficulties due to the fact that matrices generally do not commute. It is for this reason that a new approach has been developed to handle the special case of linearly coupled difference equations with *constant* coefficients.

## II. LINEARLY COUPLED DIFFERENCE EQUATIONS

A system of linearly coupled difference equations is a set of equations that relate a set of $D$ multi-variable functions $A_j(\mathbf{n}), j = 1$ to $D$. The value of a given function $A_i(\mathbf{n})$ at the evaluation point $\mathbf{n}$, is related to the values of $A_i$ itself as well as other $A_j$'s at various shifted arguments, namely,

$$A_i(n) = \sum_{j=1}^{D} \sum_{k} c_{ijk} A_j(\mathbf{n} - \mathbf{n}_{ijk}) + I_i, \quad \mathbf{n} \in \mathscr{R}. \tag{5}$$

A set of boundary conditions is given by

$$A_i(\mathbf{n}_{0l}) = \lambda_{il}; \quad n_{0l} \in \mathscr{I}. \tag{6}$$

In general, $c_{ijk}$ and $I_i$ are known coefficients that may depend on the evaluation point $\mathbf{n}$. In this article we will assume these coefficients to be constant.

At this point, it is convenient to give an example of such a system of equations. This example is relevant to the problem discussed by Hock and McQuistan[5] on "the occupation statistics for indistinguishable dumbbells on a $2 \times 2 \times N$ lattice space." Figure 1 shows such a lattice having $N$ portions of $2 \times 2$ compartments. One refers to the complete lattice as $A_1$. One calls $A_2$ the lattice whose last $2 \times 2$ array is missing one compartment. There are two topologically distinct lattices missing two compartments in their last $2 \times 2$ array; we refer to these lattices as $A_3$ and $A_4$ as shown in Fig. 2. Finally, $A_5$ is the lattice whose last $2 \times 2$ array is missing three compartments. For $j = 1$ to 5, $A_j(q,N)$ represents the total number of arrangements of $q$ dumbbells on the $A_j$ lattice having $N$ arrays. Hock and McQuistan were able to derive the following coupled recurrence relations[5]:

$$\begin{aligned} A_1(q,N) = {} & A_1(q,N-1) + 4A_1(q-1,N-1) \\ & + 2A_1(q-2,N-1) + A_1(q-4,N-2) \\ & + 4A_2(q-1,N-1) + 8A_2(q-2,N-1) \\ & + 4A_3(q-2,N-1) + 4A_3(q-3,N-1) \\ & + 2A_4(q-2,N-1) + 4A_5(q-3,N-1), \end{aligned} \tag{7a}$$

  

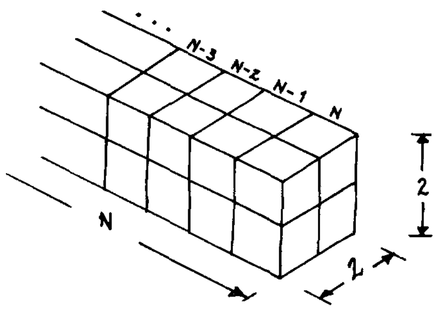


FIG. 1. 2×2×N lattice space.

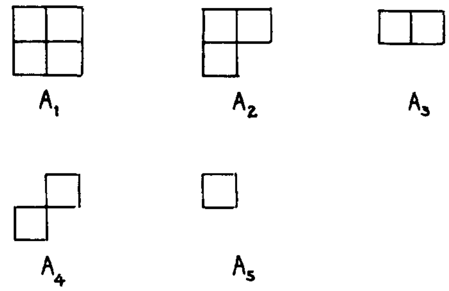


FIG. 2. We show here the last 2×2 array in the $2\times2\times N$ lattice, with no compartment ($A_1$), one compartment ($A_2$), two compartments ($A_3$ and $A_4$), and three compartments ($A_5$) missing.

$$\begin{aligned}A_2(q,N) = {}& A_1(q,N-1) + 2A_1(q-1,N-1)\\ &+ 3A_2(q-1,N-1) + 2A_2(q-2,N-1)\\ &+ 2A_3(q-2,N-1) + A_4(q-2,N-1)\\ &+ A_5(q-3,N-1),\end{aligned} \tag{7b}$$

$$\begin{aligned}A_3(q,N) = {}& A_1(q,N-1) + A_1(q-1,N-1)\\ &+ 2A_2(q-1,N-1) + A_3(q-2,N-1),\end{aligned} \tag{7c}$$

$$\begin{aligned}A_4(q,N) = {}& A_1(q,N-1) + 2A_2(q-1,N-1)\\ &+ A_4(q-2,N-1),\end{aligned} \tag{7d}$$

$$A_5(q,N) = A_1(q,N-1) + A_2(q-1,N-1). \tag{7e}$$

The region $\mathscr{R}$ for which these difference equations are satisfied is defined by

$$\mathscr{R} = \begin{bmatrix} q & \text{non-negative integer} \\ N & \text{positive integer.} \end{bmatrix} \tag{8}$$

The boundary conditions are specified by

$$A_j(q,N) = 0 \quad \text{for } q > 2N > 0, \tag{9a}$$

$$A_1(0,0) = 1; \quad A_j(0,0) = 0 \quad \text{for } j \neq 1, \tag{9b}$$

$$A_j(0,N) = 1 \quad \text{for } N \geqslant 1, \tag{9c}$$

$$A_j(q,N) = 0 \quad \text{for } q \text{ and/or } N \text{ negative integer.} \tag{9d}$$

Region $\mathscr{I}$ is then easily identified from the above. Hock and McQuistan did not make use of these boundary conditions. We will propose a general method of solution valid in the general case, Eqs. (5) and (6), and which will enable us to recover the results of Hock and McQuistan in a much more efficient and straightforward way, while obtaining at the same time, new results with no extra work.

## III. GENERATING FUNCTION METHOD

With every $A_j(n)$ one associates a generating function $G_j(X)$,

$$G_j(X_1,X_2,\ldots) = \sum_{n_1,n_2,\ldots} A_j(n_1,n_2,\ldots)(X_1)^{n_1}(X_2)^{n_2}\cdots, \tag{10}$$

where $n_1$, $n_2$,... run over the possible values of the coordinates of point $\mathbf{n}$ such that $\mathbf{n}\in\mathscr{R}$. For compactness, we will use the notation

$$(\mathbf{X})^{\mathbf{n}} = (X_1)^{n_1}(X_2)^{n_2}\cdots \tag{11}$$

so that Eq. (10) becomes

$$G_j(\mathbf{X}) = \sum_{\mathbf{n}\in\mathscr{R}} A_j(\mathbf{n})(\mathbf{X})^{\mathbf{n}}. \tag{12}$$

Combining Eqs. (5), (6), and (12), it is easy to show that

$$G_i(\mathbf{X}) = \sum_{j=1}^{D} \sum_k c_{ijk} (\mathbf{X})^{\mathbf{n}_{ijk}} G_j(\mathbf{X}) + F_i(\mathbf{X}), \tag{13}$$

where $F_i(\mathbf{X})$ is a function of $\mathbf{X}$ that can be calculated in terms of the boundary values $\lambda_{il}$ and the inhomogeneous term $I_i$. Clearly, Eq. (13) is a system of $D$ equations with $D$ unknowns, $G_j$, $j=1$ to $D$. This system can be written in the form

$$\sum_{j=1}^{D} \left[\delta_{ij} - \sum_k c_{ijk}(\mathbf{X})^{\mathbf{n}_{ijk}}\right] G_j = F_i. \tag{14}$$

$\delta_{ij}$ is the usual Kronecker's delta. Let $M$ be the $D\times D$ matrix defined by

$$M_{ij} = \delta_{ij} - \sum_k c_{ijk}(\mathbf{X})^{\mathbf{n}_{ijk}}. \tag{15}$$

Let $G$ and $F$ be the column matrices representing $G_j$ and $F_j$, then Eq. (14) becomes

$$MG = F, \tag{16}$$

and, solving for $G$, one finds

$$G = M^{-1}F. \tag{17}$$

TABLE I. Total number of arrangements of $q$ dumbbells on the nontruncated $2\times2\times N$ lattice (type $A_1$).

| | $A_1(q,N)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $N$ \ $q$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | 1 | | | | | | | | |
| 1 | 1 | 4 | 2 | | | | | | |
| 2 | 1 | 12 | 42 | 44 | 9 | | | | |
| 3 | 1 | 20 | 142 | 440 | 588 | 288 | 32 | | |
| 4 | 1 | 28 | 306 | 1672 | 4863 | 7416 | 5470 | 1620 | 121 |

TABLE II. Total number of arrangements of $q$ dumbbells on the truncated $2\times2\times N$ lattice of type $A_2$.

| | $A_2(q,N)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $N$ \ $q$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | 0 | | | | | | | | |
| 1 | 1 | 2 | | | | | | | |
| 2 | 1 | 9 | 21 | 11 | | | | | |
| 3 | 1 | 17 | 98 | 230 | 206 | 50 | | | |
| 4 | 1 | 25 | 238 | 1097 | 2574 | 2955 | 1445 | 208 | |

We now apply the generating function method to the problem discussed by Hock and McQuistan. In this case,

$$G_j(x,y) = \sum_{N=0}^{\infty} \sum_{q=0}^{2N} A_j(q,N) x^q y^N. \tag{18}$$

Here $\mathbf{n}$ is a point in a two-dimensional space of coordinates $(q,N)$ and $\mathbf{X}$ stands for $(x,y)$. Equation (14) specialized to this problem becomes

$$\begin{aligned}
&[1 - y(1+4x+2x^2) - x^4y^2]G_1 - 4xy(1+2x)G_2 \\
&\quad - 4x^2y(1+x)G_3 - 2x^2yG_4 - 4x^3yG_5 = 1, \\
&- y(1+2x)G_1 + (1 - 3xy - 2x^2y)G_2 \\
&\quad - 2x^2yG_3 - x^2yG_4 - x^3yG_5 = 0, \\
&- y(1+x)G_1 - 2xyG_2 + (1-x^2y)G_3 = 0, \\
&\quad - yG_1 - 2xyG_2 + (1-x^2y)G_4 = 0, \\
&\quad - yG_1 - xyG_2 + G_5 = 0.
\end{aligned} \tag{19}$$

The solution of this system of five linear equations with five unknowns is

$$G_j = N_j(x,y)/D(x,y), \tag{20}$$

where

$$\begin{aligned}
D(x,y) = 1 &- y(1+7x+6x^2) \\
&- xy^2(1+6x+6x^2-7x^3) \\
&+ 2x^3y^3(1+5x+13x^2+4x^3) \\
&- x^5y^4(1+2x+6x^2+9x^3) \\
&- x^8y^5(1-x+2x^2) + x^{12}y^6,
\end{aligned} \tag{21}$$

$$\begin{aligned}
N_1(x,y) = (1-x^2y)[1 &- 3xy(1+x) \\
&+ x^3y^2(x-3) + x^6y^3],
\end{aligned} \tag{22a}$$

$$\begin{aligned}
N_2(x,y) = (1-x^2y)[&y(1+2x) \\
&+ x^2y^2(2+x) - x^5y^3],
\end{aligned} \tag{22b}$$

$$N_3(x,y) = [y(1+x)N_1 + 2xyN_2]/(1-x^2y), \tag{22c}$$

$$N_4(x,y) = [yN_1 + 2xyN_2]/(1-x^2y), \tag{22d}$$

$$N_5(x,y) = yN_1 + xyN_2. \tag{22e}$$

## IV. DECOUPLING OF THE DIFFERENCE EQUATIONS

The explicit expression of the generating function $G_j(\mathbf{X})$ can be presented in the form

$$G_j(\mathbf{X}) = N_j(\mathbf{X})/D(\mathbf{X}), \tag{23}$$

where $D(\mathbf{X})$ is the determinant of matrix $M$. As exhibited in Eq. (15), matrix element $M_{ij}$ is a finite polynomial. Therefore, $D(\mathbf{X})$ is also a finite polynomial. It is straightforward to show that

$$D(\mathbf{X})G_j(\mathbf{X}) = N_j(\mathbf{X}) \tag{24}$$

generates a multi-term difference equation involving $A_j(\mathbf{n})$ only. Indeed, let

$$D(\mathbf{X}) = \sum_p (\mathbf{X})^{\mathbf{n}_p}\alpha_p. \tag{25}$$

The left-hand side of Eq. (24) becomes

$$\begin{aligned}
D(\mathbf{X})G_j(\mathbf{X}) &= \sum_p (\mathbf{X})^{\mathbf{n}_p}\alpha_p \sum_{\mathbf{n}\in\mathscr{R}} A_j(\mathbf{n})(\mathbf{X})^{\mathbf{n}} \\
&= \sum_{\mathbf{n}\in\mathscr{R}} \sum_p \alpha_p A_j(\mathbf{n})(\mathbf{X})^{\mathbf{n}+\mathbf{n}_p}.
\end{aligned} \tag{26}$$

The equivalence between the right-hand side of Eq. (23) and the right-hand side of Eq. (24) provides the difference equation for $A_j(\mathbf{n})$. By relabeling $\mathbf{n}$ the combination $\mathbf{n}+\mathbf{n}_p$, one finds

$$\sum_{\mathbf{n}} \sum_p \alpha_p A_j(\mathbf{n}-\mathbf{n}_p)(\mathbf{X})^{\mathbf{n}} \equiv N_j(\mathbf{X}) \tag{27}$$

or

$$\sum_p \alpha_p A_j(\mathbf{n}-\mathbf{n}_p) = K_j(\mathbf{n}), \tag{28}$$

where $K_j(\mathbf{n})$ is an inhomogeneous term which comes from the expression of $N_j(\mathbf{X})$; it is the coefficient of $(\mathbf{X})^{\mathbf{n}}$ in the series expansion of $N_j(\mathbf{X})$. This completes the decoupling of our system of linearly coupled difference equations.

TABLE III. Total number of arrangements of $q$ dumbbells on the truncated $2\times2\times N$ lattice of type $A_3$.

| | $A_3(q,N)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $N$ \ $q$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | 0 | | | | | | | | |
| 1 | 1 | 1 | | | | | | | |
| 2 | 1 | 7 | 11 | 3 | | | | | |
| 3 | 1 | 15 | 73 | 135 | 86 | 12 | | | |
| 4 | 1 | 23 | 197 | 793 | 1561 | 1423 | 506 | 44 | |

TABLE IV. Total number of arrangements of $q$ dumbbells on the truncated $2\times2\times N$ lattice of type $A_4$.

| | $A_4(q,N)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $N$ \ $q$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | 0 | | | | | | | | |
| 1 | 1 | | | | | | | | |
| 2 | 1 | 6 | 7 | | | | | | |
| 3 | 1 | 14 | 61 | 92 | 38 | | | | |
| 4 | 1 | 22 | 177 | 650 | 1109 | 792 | 170 | | |

For the purpose of illustration, let us apply our general method to the problem discussed by Hock and McQuistan. It is clear that a monomial $x^\gamma y^\delta$ in the expression of $D(x,y)$ generates a term $A(q-\gamma,N-\delta)$. Since $D(x,y)$ is the sum of 20 monomials, then each of the $A_j$'s satisfies a 20-term recurrence relation. It happens that, in this case, there is no inhomogeneous term. Identifying the left side with the right side of Eq. (24) setting $j=1$, and taking for $D(\mathbf{X})$ expression (21) and for $N_1(\mathbf{X})$ expression (22a), one finds that $A_1(q,N)$ should satisfy the initial values listed in Table I and the relation

$$\begin{aligned}
&A_1(q,N)-A_1(q,N-1)-7A_1(q-1,N-1)\\
&\quad-6A_1(q-2,N-1)-A_1(q-1,N-2)\\
&\quad-6A_1(q-2,N-2)-6A_1(q-3,N-2)\\
&\quad+7A_1(q-4,N-2)+2A_1(q-3,N-3)\\
&\quad+10A_1(q-4,N-3)+26A_1(q-5,N-3)\\
&\quad+8A_1(q-6,N-3)-A_1(q-5,N-4)\\
&\quad-2A_1(q-6,N-4)-6A_1(q-7,N-4)\\
&\quad-9A_1(q-8,N-4)-A_1(q-8,N-5)\\
&\quad+A_1(q-9,N-5)-2A_1(q-10,N-5)\\
&\quad+A_1(q-12,N-6)=0.
\end{aligned}\tag{29}$$

The initial values listed in Table I are precisely the values computed by Hock and McQuistan.[5]

A result not previously obtained by Hock and McQuistan is that $A_2$, $A_3$, $A_4$, and $A_5$ all satisfy the same difference equation (29). However, these quantities do not have the same set of initial values. Since the method of obtaining the initial values for the $A$'s is the same for all the $A$'s, we will drop the indices 1 to 5 on the $G$'s, the generating functions, and the $A$'s. We write $G(x,y)$ and $D(x,y)$ in the form,

$$\begin{aligned}
G(x,y)&=\sum_{N=0}^{\infty}\sum_{q=0}^{2N}A(q,N)x^qy^N,\\
D(x,y)&=\sum_{i=0}^{i_m}\sum_{j=0}^{j_m}d_{ij}x^iy^j,
\end{aligned}\tag{30}$$

so that their product becomes

$$D(x,y)G(x,y)=\sum_{i=0}^{i_m}\sum_{j=0}^{j_m}\sum_{N=0}^{\infty}\sum_{q=0}^{2N}d_{ij}A(q,N)x^{q+i}y^{N+j}.\tag{31}$$

This product must be identical to the polynomial $N(x,y)$ (here again we are dropping the index on function $N$ pretty much the same way we did it for $G$ and $A$). $N(x,y)$ is a polynomial of the form

$$N(x,y)=\sum_{k=0}^{k_m}\sum_{l=0}^{l_m}e_{kl}x^ky^l.\tag{32}$$

Coefficients $d_{ij}$ and $e_{kl}$ are immediately identified knowing the explicit expressions of $D(x,y)$ and $N(x,y)$, respectively. Since expansions (31) and (32) must be equivalent, one finds the condition

$$\sum_{q+i=k}\sum_{N+j=l}d_{ij}A(q,N)=e_{kl}.\tag{33}$$

For $k>k_m$ and $l>l_m$, $A(q,N)$ satisfies the difference equation (29) and Eq. (33) enables one to obtain the initial values listed in Tables II, III, IV, and V for $A_1$, $A_2$, $A_3$, $A_4$, and $A_5$, respectively.

## V. CONCLUSION

We have shown that any system of linearly coupled difference equations with constant coefficients can be decoupled by use of suitably chosen generating functions $G_j(\mathbf{X})$. All functions $A_j(\mathbf{n})$ are shown to satisfy the *same* decoupled difference equation with appropriate initial value condi-

TABLE V. Total number of arrangements of $q$ dumbbells on the truncated $2\times2\times N$ lattice of type $A_5$.

| | $A_5(q,N)$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $N$ \ $q$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | 0 | | | | | | | | |
| 1 | 1 | | | | | | | | |
| 2 | 1 | 5 | 4 | | | | | | |
| 3 | 1 | 13 | 51 | 65 | 20 | | | | |
| 4 | 1 | 21 | 159 | 538 | 818 | 494 | 82 | | |

tions. Application of the general theory to a specific problem discussed by Hock and McQuistan has been successful and enables one to not only elegantly reproduce their results, but also obtain new results with no extra hardship. This is due to the fact that our theory shows that all $A_j(\mathbf{n})$'s satisfy the same decoupled difference equation.[6]


[1]A. F. Antippa and A. J. Phares, J. Math. Phys. **18**, 173 (1977); A. J. Phares, *ibid.* **18**, 1838 (1977).

[2]A. J. Phares, *Advances in Computer Methods for Partial Differential Equations* (IMACS, Rutgers University, New Brunswick, 1979), Vol. III, p. 47; A. J. Phares and R. J. Meier, J. Math. Phys. **22**, 1021 (1981); A. J. Phares, *ibid.* **22**, 1030 (1981).

[3]A. F. Antippa and A. J. Phares, J. Math. Phys. **19**, 308 (1978); A. J. Phares, *ibid.* **19**, 2239 (1978); P. A. Maurone and A. J. Phares, *ibid.* **20**, 2191 (1979); P. A. Maurone and A. J. Phares, *ibid*, **21**, 830 (1980); A. J. Phares and K. H. Simmons, Am. J. Phys. **48**, 159 (1980); A.J. Phares, The Fibonacci Quarterly, **22**, 29 (1984).

[4]A. J. Phares, *Advances in Computer Methods for Partial Differential Equations* (IMACS, Rutgers University, New Brunswick, 1981), Vol. IV, p. 53.

[5]J. L. Hock and R. B. McQuistan, J. Math. Phys. **24**, 1859 (1983).

[6]A systematic application of the technique presented in this paper leads to several interesting results. A. J. Phares, J. Math. Phys. **25**, 1756 (1984).

# Poisson branching point processes

Kuniaki Matsuo and Malvin Carl Teich
*Columbia Radiation Laboratory, Department of Electrical Engineering, Columbia University, New York, New York 10027*

Bahaa E. A. Saleh
*Columbia Radiation Laboratory, Department of Electrical Engineering, Columbia University, New York, New York 10027 and Department of Electrical and Computer Engineering, University of Wisconsin, Madison, Wisconsin 53706*

(Received 10 August 1983; accepted for publication 23 December 1983)

We investigate the statistical properties of a special branching point process. The initial process is assumed to be a homogeneous Poisson point process (HPP). The initiating events at each branching stage are carried forward to the following stage. In addition, each initiating event independently contributes a nonstationary Poisson point process (whose rate is a specified function) located at that point. The additional contributions from all points of a given stage constitute a doubly stochastic Poisson point process (DSPP) whose rate is a filtered version of the initiating point process at that stage. The process studied is a generalization of a Poisson branching process in which random time delays are permitted in the generation of events. Particular attention is given to the limit in which the number of branching stages is infinite while the average number of added events per event of the previous stage is infinitesimal. In the special case when the branching is instantaneous this limit of continuous branching corresponds to the well-known Yule–Furry process with an initial Poisson population. The Poisson branching point process provides a useful description for many problems in various scientific disciplines, such as the behavior of electron multipliers, neutron chain reactions, and cosmic ray showers.

PACS numbers: 02.50. + s, 29.70. – e, 94.40.Pa

## I. INTRODUCTION

The theory of branching processes provides an important set of mathematical tools which may be applied to many problems in modern physics.[1,2] These range from multiple atomic transitions to extensive air showers produced by cosmic rays. In many of the existing mathematical treatments of these problems, the branching is treated as an instantaneous effect. However, in most physical systems, a random time delay (or spatial dispersion) is inherent in the multiplication process.

In a recent set of papers, we examined a special generalized branching process in which the multiplication of each event is Poisson and a random time delay is introduced at every stage. The first model that we analyzed[3–5] is the two-stage cascaded Poisson, in which each event of a primary Poisson point process produces a virtual inhomogeneous rate function which, in turn, generates a secondary Poisson point process. These secondary point processes are superimposed to form the final point process. In that model, primary events themselves are excluded from the final point process.[3–5] The description turns out to be that of a doubly stochastic Poisson point process (DSPP), which we refer to as the shot-noise-driven process (SNDP).[3] The SNDP is also a special case of the Neyman–Scott cluster process.[3,5] Because of the great body of theoretical results available for the DSPP, our calculations for the statistical properties of the process turned out to be relatively straightforward. In another version of this two-stage model, primary events are carried forward to the final process.[6]

The second system which we analyzed[7] is an $m$-stage cascade of Poisson processes buffered by linear filters. Each filtered point process forms the input to the following stage, acting as a rate for a DSPP. This is equivalent to a cascaded SNDP. We obtained the counting and time statistics, as well as the autocovariance function. The results of that study are likely to find use in problems where a series of multiplicative effects occur. Examples are the behavior of photon and charged-particle detectors, the production of cosmic rays, and the transfer of neural information.

In this paper, we consider a cascade model in which primary events are carried forward together with secondary events, to form the point process at the input to each successive stage. Since the primary and secondary events comprising the union process at each stage are not independent,[6] the solution is somewhat more difficult than for the cascaded Poisson case considered previously.[7] The initial point process is assumed to be a homogeneous Poisson process (HPP). The final process is itself homogeneous (stationary). This treatment should allow us to model a wide variety of physical phenomena in which particles produce more particles, and so on, with the original particles remaining. Our process may also be regarded as a special generalized branching process,[1] in which each event of the HPP produces an age-dependent point process. However, our interest is in the union of the branching point processes rather than in the statistics of the number of events at a certain time (or place), as is the customary quantity of interest in age-dependent branching processes.

Branching processes with properties such as age dependence, random walk, and diffusion have been studied extensively from a general theoretical point of view.[1] Few of the statistical properties are obtained in a form amenable to numerical solution, however. The present work examines a relatively simple process that describes branching with time delay. Thanks to the simplicity offered by the Poisson assumption, we can obtain explicit formulas for useful statistical properties that may be experimentally measured. Examples are the counting distribution, moments, and power spectral density, as we demonstrate.

In Sec. II, we review the properties of a Poisson branching process in which the branching is instantaneous. This establishes the properties of the limiting situation, to which our process must converge when time delay is negligible. We also consider the limiting case when the number of branching stages approaches infinity while the average number of secondary events per primary event approaches zero. In the instantaneous multiplication case, this results in the Yule–Furry process,[2] driven by HPP initial events.

In Sec. III, we introduce time delay at each stage of branching and define the process formally. We provide expressions for the moment generating functional of the process, from which we compute the moments, counting probability distribution, and autocorrelation function (or power spectral density). In Sec. IV, we discuss the important limit of the continuous branching point process with time delay, showing how it differs from the instantaneous continuous branching case.

## II. INSTANTANEOUS POISSON BRANCHING PROCESS

This section is divided into three subsections. In Subsec. A, we briefly discuss the well-known general Galton–Watson (GW) branching process.[1] In Subsec. B, a special Galton–Watson branching process, in which the multiplication is Poisson, is examined. The properties of a Poisson Galton–Watson process, in which the initial number of events is itself Poisson, are examined in detail in Subsec. C.

### A. Galton–Watson branching process

Let $N_0, N_1, N_2, \ldots$ be nonnegative integers denoting the successive random variables of a Markov chain, where $N_m$ denotes the size of the population of the $m$th generation of the branching process. The population $N_{m+1}$ at the $(m+1)$st generation is determined by the sum

$$N_{m+1} = \sum_{k=1}^{N_m} Z_k^m \tag{1}$$

of $N_m$ independent, identically distributed (iid) random variables $Z_1^m, Z_2^m, \ldots, Z_N^m$, each with probability distribution

$$\text{Prob}(Z^m = k) = p_k^m, \qquad k = 0,1,2,\ldots . \tag{2}$$

This determines the transition matrix of the Markov chain. It is assumed that $N_0 = 1$. The chain is known as a Galton–Watson (GW) process.

The basic assumption is that each of the members of a generation branches independently and identically to generate the population of the following generation. The statistical properties of the random number $N_m$ may be determined from its probability generating function

$$G_m(z) = \langle z^{N_m} \rangle , \tag{3}$$

which may be calculated by use of recursive equations. These are easily determined by using the iid assumption:

$$G_0(z) = z ,$$
$$G_{m+1}(z) = G_m[O_m(z)] , \quad m = 0,1,\ldots , \tag{4}$$

where

$$O_m(z) = \sum_{k=0}^{\infty} p_k^m z^k \tag{5}$$

is the probability generating function of the random variable $Z^m$.

### B. Poisson Galton–Watson process

We now consider a special case of the GW process by taking

$$p_k^m = \begin{cases} 0, & k = 0, \\ \alpha_m^{k-1} e^{-\alpha_m}/(k-1)!, & k = 1,2,\ldots, \end{cases} \tag{6}$$

i.e., $Z_k^m$ obeys a shifted version of the Poisson distribution[8] of mean $\alpha_m$. This signifies that each member of the $m$th generation survives and remains in the $(m+1)$st generation, adding a cluster of offspring which is Poisson distributed with mean $\alpha_m$. We shall call this special GW process the Poisson GW process (PGW).

By substituting (6) in (5), we obtain

$$O_m(z) = z e^{\alpha_m(z-1)} , \quad m = 0,1,2,\ldots . \tag{7}$$

Therefore, from (4), the probability generating function is

$$G_0(z) = z ,$$
$$G_{m+1}(z) = G_m[z e^{\alpha_m(z-1)}] , \quad m = 0,1,\ldots . \tag{8}$$

### C. Poisson Galton–Watson process with an initial Poisson population

In this subsection, we define a process in which members of an initial population of random size $N_0$ each independently generate identical PGW processes. The final process is the sum of these processes. Furthermore, we assume that $N_0$ is Poisson with mean $a$.

The properties of this process may be obtained by regarding it as a shifted version of a special GW process in which $N_0 = 1$, and the $p_k^m$ are given by

$$p_k^1 = a^k e^{-a}/k! , \quad k = 0,1,\ldots ,$$
$$p_k^m = \left. \begin{cases} 0, & k = 0 \\ \alpha_m^{k-1} e^{-\alpha_m}/(k-1)!, & k = 1,2,\ldots \end{cases} \right\} , \quad m = 2,3,\ldots . \tag{9}$$

Thus $N_1 = Z^1$ is Poisson with mean $a$, and the branching to generations $m = 2,3,\ldots$ occurs in accordance with a shifted Poisson law (in which no deaths occur) with parameters $\alpha_2, \alpha_3, \ldots$ . This allows us to write the probability generating function for this special process as

$$G_0(z) = z ,$$
$$G_1(z) = e^{a(z-1)} , \tag{10}$$
$$G_{m+1}(z) = G_m[z e^{\alpha_m(z-1)}] , \quad m = 1,2,\ldots .$$

Because (10) forms the limiting case for the process we shall define in Sec. III, some of its important statistical properties will be provided in the following. All of these properties may be determined by using (10).

### 1. Moment generating function

The moment generating function (mgf) $Q_m(s) = \langle \exp(-sN_m)\rangle$ may be obtained from the probability generating function $G_m(z)$ by the use of[7]

$$Q_m(s) = G_m(e^{-s}) . \quad (11)$$

With the help of (10) we can show that for a Poisson branching process, with homogeneous branching (i.e., $\alpha_j = \alpha$), and with a Poisson initial population,

$$Q_m(s) = \exp\{a[D_m(s) - 1]\} , \quad m \geqslant 1 , \quad (12)$$

where

$$D_m(s) = D_1(s) \exp \left\{\alpha \sum_{j=1}^{m-1} [D_j(s) - 1]\right\} ,$$

$$D_1(s) = e^{-s} .$$

For $m = 1$ and $m = 2$, we recover the mgf's for the Poisson and Thomas counting distributions, respectively.[6,8,9]

### 2. Moments

The moments of the count number $N_m$ may be obtained from (11). The mean and variance are[7]

$$\langle N_m \rangle = a \prod_{k=1}^{m-1} (1 + \alpha_k) , \quad m \geqslant 2 \quad (13)$$

and

$$\mathrm{Var}[N_m] = a \sum_{k=0}^{m-1} C_k \prod_{r=k+1}^{m-1} (1 + \alpha_r)^2 , \quad (14)$$

where

$$C_0 = 1 ,$$

$$C_1 = \alpha_1 ,$$

$$C_k = \alpha_k \sum_{r=1}^{k-1} (1 + \alpha_r) , \quad k \geqslant 2 .$$

The count variance-to-mean ratio (Fano factor $F$) provides a suitable index for the degree of deviation from a Poisson counting process for which $F = 1$.[9] We form this ratio with the help of (13) and (14):

$$F_m = \frac{\mathrm{Var}[N_m]}{\langle N_m \rangle} = \sum_{k=0}^{m-1} \left\{\alpha_k \left[\prod_{r=1}^{k-1} (1+\alpha_r)\right] \left[\prod_{r=k+1}^{m-1} (1+\alpha_r)^2\right]\right\} \times \left(\prod_{k=1}^{m-1} (1+\alpha_k)\right)^{-1} , \quad (15)$$

where

$$\alpha_0 = 1 ,$$

$$\prod_{r=t}^{s} (\cdot) = 1 \quad \text{for} \quad s < t .$$

For homogeneous branching

$$\langle N_m \rangle = a(1+\alpha)^{m-1} , \quad m \geqslant 1 , \quad (16)$$

$$\mathrm{Var}[N_m] = a(1+\alpha)^{m-2}[(2+\alpha)(1+\alpha)^{m-1} - 1] , \quad m \geqslant 1 , \quad (17)$$

$$F_m = [1/(1+\alpha)][(2+\alpha)(1+\alpha)^{m-1} - 1] , \quad m \geqslant 1 . \quad (18)$$

The results for the one- and two-stage cases are clearly identical to those for the Poisson and Thomas distributions, respectively.[6,9]

When the branching is homogeneous, the $n$th moment of $N_m$ may be determined from the mgf provided in (12). The result is the recurrence relation

$$\langle N_m^{n+1} \rangle = \langle N_m \rangle \sum_{k=0}^{n} \binom{n}{k} \langle N_m^{n-k} \rangle I_m^{(k+1)} , \quad m \geqslant 2 , \quad (19)$$

where

$$I_m^{(1)} = 1 ,$$

$$I_m^{(k+1)} = \frac{(-1)^k}{(1+\alpha)^{m-1}} D_m^{k+1} ,$$

$$D_m^{(k+1)} = D_m^k + \alpha \sum_{l=0}^{k} \binom{k}{l} D_m^{(k-l)} \sum_{j=1}^{m-1} D_j^{(l+1)} ,$$

$$D_m^{(0)} = 1 ,$$

$$D_1^{(k)} = 1 , \quad k \geqslant 1 ,$$

$$\langle N_m \rangle = a(1+\alpha)^{m-1} .$$

### 3. Counting probability distribution

The probability distribution $p_m(n)$ of $N_m$ may be obtained by differentiating the probability generating function $G_m(z)$,[6]

$$p_m(n) = \frac{1}{n!} \frac{\partial^n}{\partial z^n} G_m(z)\bigg|_{z=0} . \quad (20)$$

Using (10) and (20), we obtain the recurrence relation for the homogeneous case,

$$p_m(0) = e^{-a} , \quad (21)$$

$$(n+1) p_m(n+1) = \langle N_m \rangle \sum_{k=0}^{n} p_m(n-k) J_m^{(k+1)} ,$$

where

$$J_m^{(k+1)} = \frac{(-1)^{k+1}}{(1+\alpha)^{m-1} k!} E_m^{(k+1)} ,$$

$$E_m^{(k+1)} = Y_m^{(k)} + \alpha \sum_{l=0}^{k} \binom{k}{l} E_m^{(k-l)} \sum_{j=1}^{m-1} E_j^{(l+1)} ,$$

$$Y_m^{(k+1)} = \alpha \sum_{l=0}^{k} \binom{k}{l} Y_m^{(k-l)} \sum_{j=1}^{m-1} E_j^{(l+1)} ,$$

$$Y_m^{(0)} = \exp\left\{\alpha \sum_{j=1}^{m-1} [E_j^{(0)} - 1]\right\} ,$$

$$E_m^{(k)} = 0 \quad \text{for all } m \geqslant 1, \text{ all } k \geqslant 0, \text{ except } (m,k) = (1,1) ,$$

$$E_1^{(1)} = -1 .$$

In Fig. 1(a), we present a graphical representation of the counting distribution $p_m(n)$ versus the count number $n$ for $m = 2,3,4$, and 10, with $\alpha = 0.5$ and $\langle N_m \rangle = 10$. It is seen that the distribution for $m = 10$ approaches a $\delta$-function at the origin plus a relatively flat component, indicating very strong pulse clustering. In Fig. 1(b), the case for $\alpha = 2.0$ is shown. For both cases, it is clear that the variance of the counting distributions increases as the number of stages increases. It is also apparent that the variance increases with increasing $\alpha$, when $m$ and $\langle N_m \rangle$ are fixed. The results for $m = 2$ are identical to those for the instantaneous Thomas process.[6,9]

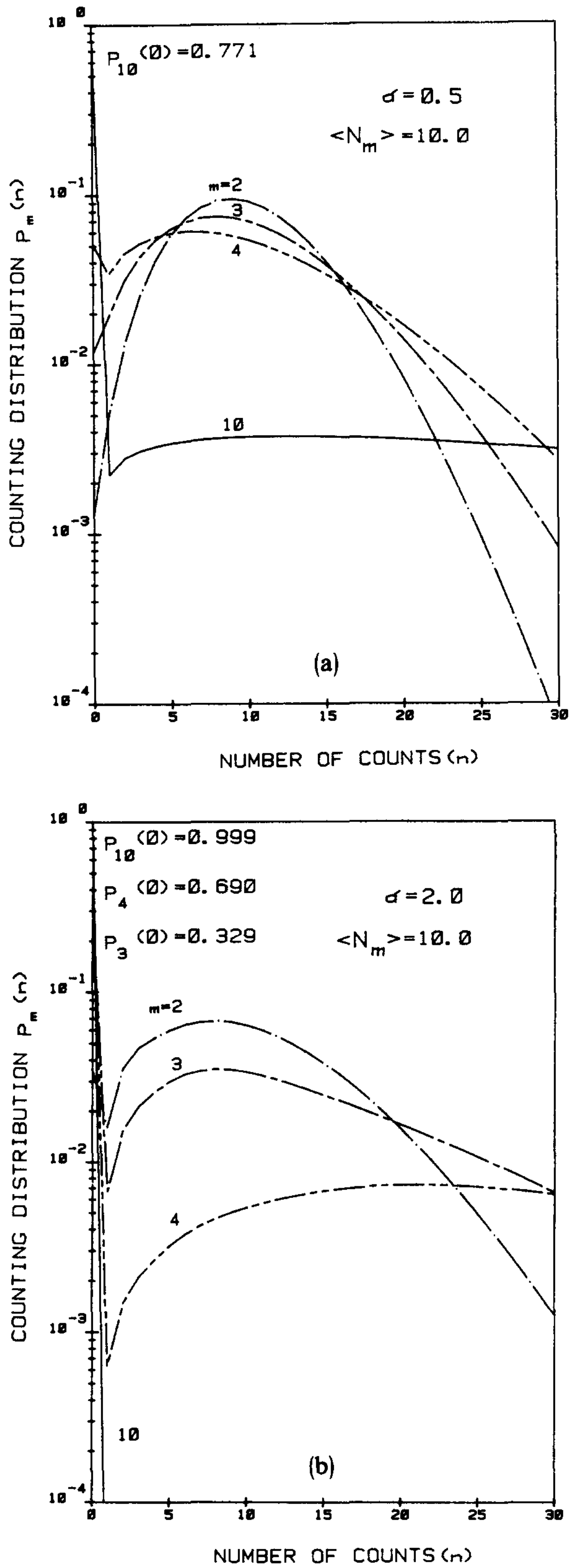


FIG. 1. Counting distribution $p_m(n)$ vs count number $n$ with $m$ as a parameter. The mean count $\langle N_m \rangle = 10.0$ for all cases. (a) $\alpha = 0.5$; (b) $\alpha = 2.0$.

### 4. *Limit of continuous branching*

An important special case is one in which the number of branching stages approaches infinity, while the branching at each stage becomes infinitesimal. Let

$$m \to \infty , \tag{22a}$$

$$\alpha \to 0 , \tag{22b}$$

with the product

$$m\alpha = x \tag{22c}$$

remaining finite. In this limit, we denote $N_m$ and $Q_m(s)$ as $N_x$ and $Q_x(s)$, respectively. The limit of (12) yields

$$Q_x(s) = \exp\{a[D_x(s) - 1]\} , \tag{23a}$$

where $D_x(s)$ satisfies the differential equation

$$\frac{\partial}{\partial x} D_x(s) = D_x(s)\,[D_x(s) - 1] , \tag{23b}$$

and the initial condition is

$$D_0(s) = e^{-s} . \tag{23c}$$

Equation (23) has the solution

$$Q_x(s) = \exp\left\{ -a \frac{1 - e^{-s}}{1 - (1 - e^{-x})e^{-s}} \right\} , \tag{24}$$

which is recognized as the moment generating function for the linear birth (Yule–Furry) process with a Poisson initial population.[10]

The $n$th ordinary moment of $N_x$ is found to satisfy

$$\langle N_x^{n+1} \rangle = \langle N_x \rangle \sum_{k=0}^{n} \binom{n}{k} \langle N_x^{n-k} \rangle I_x^{(k+1)} , \tag{25}$$

where

$$I_x^{(1)} = 1 ,$$

$$I_x^{(k+1)} = \frac{(-1)^{k+1} e^{-2x}}{1 - e^{-x}} \sum_{l=1}^{\infty} \frac{l^{k+1}}{(1 - e^{-x})^l} .$$

The mean count is

$$\langle N_x \rangle = ae^x , \tag{26}$$

and the variance, which is readily obtained from (25), is given by

$$\mathrm{Var}[N_x] = ae^x(2e^x - 1) . \tag{27a}$$

The Fano factor therefore takes the particularly simple form

$$F_x = 2e^x - 1 , \tag{27b}$$

which is, of course, also obtainable from (18).

The probability (counting) distribution $p_x(n)$ of $N_x$ may be determined from (24) or from the limit of (21). The result is

$$p_x(0) = e^{-a} ,$$

$$(n+1)p_x(n+1) = \langle N_x \rangle \sum_{k=0}^{n} p_x(n-k) J_x^{(k+1)} , \tag{28}$$

where

$$J_x^{(k+1)} = e^{-2x}(k+1)(1 - e^{-x})^k .$$

It is of interest to show the manner in which the distribution $p_m(n)$ approaches $p_x(n)$ as $m \to \infty$ and $\alpha = x/m \to 0$. In Fig. 2, we plot the counting distributions $p_m(n)$ for $m = 5$, 10, and 50, with fixed $m\alpha = x = 1.0$. We also plot $p_x(n)$ for $x = 1.0$, which is labeled Y–F (Yule–Furry). The final count mean of all distributions was kept constant at a value $\langle N_m \rangle = 10$ [this means that the initial mean $a$ differs from curve to curve; see (16) and (26)]. The results demonstrate

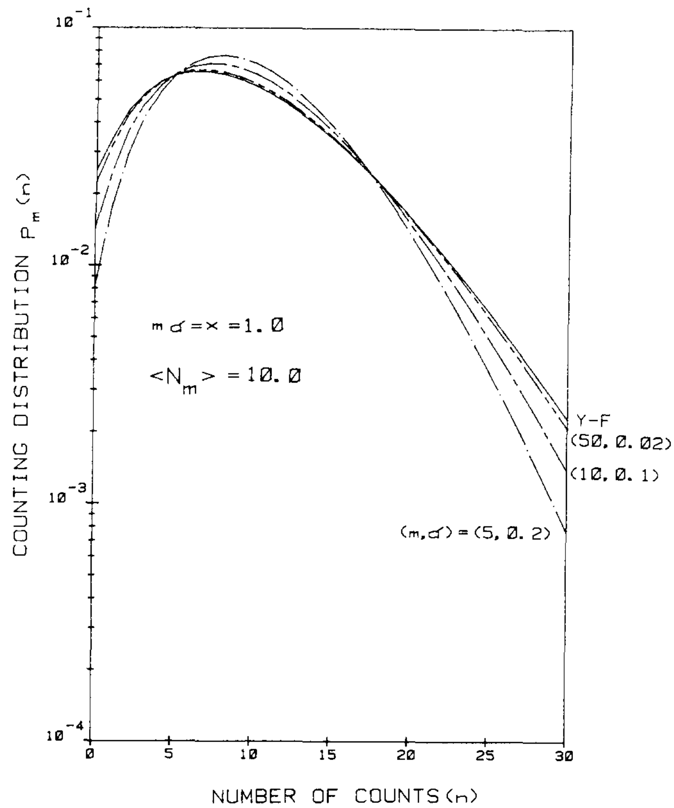


FIG. 2. Counting distributions $p_x(n)$ for the Poisson-driven Yule–Furry process (labeled Y–F) and $p_m(n)$ for the $m$-stage Poisson Galton–Watson branching processes with a Poisson initial population. $\langle N_m \rangle = 10.0$ and $m\alpha = x = 1.0$ for all cases. Note that $p_m(n)$ approaches $p_x(n)$ quite closely for $m = 50$.

that the limiting Yule–Furry distribution $p_x(n)$ is essentially attained (for this particular set of parameters) when $m \geqslant 50$.

## III. POISSON BRANCHING POINT PROCESS

### A. General branching point process

A generalization of the sequence of integers $N_0, N_1, N_2, \ldots$ discussed in Sec. IIA is the sequence of point processes $N_0(t)$, $N_1(t), N_2(t), \ldots$. Events now have times associated with them. The variable $N_m(t)$ represents the numbers of events of the $m$th generation which occur in the time interval $(-\infty, t]$. It is again assumed that the sequence $N_m(t)$ is Markov, i.e., given the point process $N_m(t)$, the statistics of the point process $N_{m+1}(t)$ are completely defined. The transition from the process $N_m(t)$ is obtained as follows. Each event of a given generation independently generates a point process. These point processes are statistically identical when each is measured from the occurrence time of the event that generated it. The following generation is comprised of the union of those point processes. For example, if the process $N_m(t)$ has occurrence (jump) times $t_1^m, t_2^m, t_k^m, \ldots$, the $k$th event of the $m$th generation, which occurs at time $t_k^m$, generates a point process $Z_k^m(t - t_k^m)$. The point processes $Z_1^m(t), Z_2^m(t), \ldots$ are iid. The process $N_{m+1}(t)$ is the union of the processes $Z_k^m(t - t_k^m)$, $k = 1, 2, \ldots$; i.e.,

$$N_{m+1}(t) = \sum_{k=1}^{N_m(t)} Z_k^m(t - t_k^m).$$

The general branching point process $N_0(t), N_1(t), \ldots$ is completely defined once the point processes $Z^m(t)$ are defined for $m = 0, 1, \ldots$.

### B. Poisson branching point process

We shall call a general branching point process Poisson if $Z^m(t)$ is the union of a Poisson point process of rate $h_m(t)$, with a process $u(t)$ $[u(t) = 0, t < 0; u(t) = 1, t \geqslant 0]$ containing only one count at $t = 0$. The initial process $N_0(t)$ also contains a single event at $t = 0$, i.e., $N_0(t) = u(t)$.

### C. Poisson branching point process driven by an initial Poisson point process

Here we assume that the 1st generation $N_1(t)$ is described by an HPP counting process of rate $\mu$. Subsequent branching follows a Poisson branching point process as described in Sec. IIIB. Because of the stationarity of the initial generation $N_1(t)$, the point processes of subsequent generations will remain stationary. This process shall be referred to as the Poisson-driven Poisson branching point process.

To understand the nature of the formation of this process, and its possible applicability to physical systems, we can think of it schematically as a cascade of systems $T_m$ operating on random point signals. Consider an operator $P$ representing a Poisson point generator that operates on a function $X(t)$ to produce a sequence of impulses $dN(t) = \Sigma_k \delta(t - t_k)$; $dN(t)$ represents a Poisson point process of rate $X(t)$. Consider also a unit system designated $h_m(t)$, representing a time-invariant linear system of impulse response $h_m(t)$, that operates on the signal $\Sigma_k \delta(t - t_k)$ to produce the signal $\Sigma_k h_m(t - t_k)$. The functions $h_m(t)$ are assumed nonnegative.

The Poisson branching point process with an initial Poisson population is formed as follows. The first generation $dN_1(t)$ is a homogeneous set of Poisson impulses of rate $\mu$ as shown in Fig. 3(a). This signal is modified by the system $T_1$ to produce a set of random impulses $dN_2(t)$ representing the second generation, and so on, as indicated in the figure. The system $T_m$, which is shown in Fig. 3(b), filters the stream of impulses provided to its input with a linear time-invariant filter of impulse response $h_m(t)$. The filtered signal $X_m(t)$ is a random continuous process, which in turn acts as the stochastic rate of a DSPP, represented by the set of impulses $dM_m(t)$. The union of this set of impulses with the input set $dN_m(t)$ constitutes the final output set of impulses $dN_{m+1}(t)$. [Figure 3(c) will be discussed subsequently.]

We now proceed to determine the statistical properties of the above-described Poisson-driven Poisson branching point process. The quantities we derive in this section include: (i) the moment generating functional for the process $N_m(t)$; (ii) the multifold and singlefold moment generating functions for the numbers of counts in $L$ intervals $[t_j, t_j + T_j]$, $j = 1, 2, \ldots, L$; (iii) the moments of the number of counts $N_m(t)$ in the interval $[0, T]$; (iv) the counting probability distribution for $N_m(T)$ in $[0, T]$; and (v) the correlation function and power spectral density.

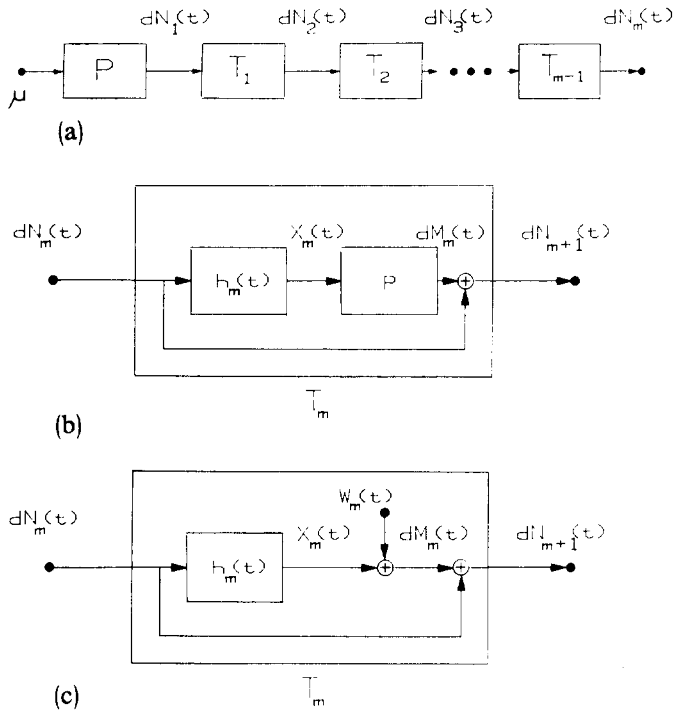


FIG. 3. (a) Schematic representation for the $m$-stage Poisson branching process excited by a homogeneous Poisson process with rate $\mu$. P represents a Poisson point process generator whereas $T_m$ represents a random point process transformation operator. (b) Point process transformation unit cell for each stage. The box $h_m(t)$ represents the impulse-response function for a time-invariant linear filter, and P is a Poisson point process generator. (c) Equivalent unit cell useful for calculating the count mean and variance. $W_m(t)$ is a stationary, zero-mean, white process.

### *1. Moment generating functional*

The moment generating functional associated with a Poisson-driven Poisson branching point process $N_m(t)$, at the $m$th stage, is defined by the expectation

$$L_m(s) = \left\langle \exp\left( - \int_{-\infty}^{\infty} s(t)\, dN_m(t) \right) \right\rangle . \quad (29)$$

It can be shown[6,7] that $L_m(s)$ satisfies the following recurrence relation:

$$L_m(s) = \left\langle \exp\left[ - \int_{-\infty}^{\infty} \{ s(t) - h_{m-1}(-t) * [e^{-s(t)} - 1] \}\, dN_{m-1}(t) \right] \right\rangle$$
$$= L_{m-1}\{ s(t) - h_{m-1}(-t) * [e^{-s(t)} - 1] \}, \quad (30)$$

where the symbol $*$ indicates convolution. The moment generating functional for the first stage is

$$L_1(s) = \exp\left\{ \mu \int_{-\infty}^{\infty} [e^{-s(t)} - 1]\, dt \right\} . \quad (31)$$

For convenience, we define the following operator:

$$q_m(\cdot) = -(\cdot) + \int_{-\infty}^{\infty} h_m(\sigma - t)[\exp\{-(\cdot)\} - 1]\, d\sigma . \quad (32)$$

Combining (29)–(32) then yields

$$L_m(s) = \exp\left\{ \mu \int_{-\infty}^{\infty} [\exp\{ q_1(q_2 \cdots q_{m-1}(s(t))) \} - 1]\, dt \right\},$$
$$m \geqslant 2 . \quad (33)$$

For the case of identical impulse response functions at each stage [$h_m(t) = h(t)$ for all $m$], (33) can be expressed as

$$L_m(s) = \exp\left\{ \mu \int_{-\infty}^{\infty} [D_m(s,t) - 1]\, dt \right\}, \quad m \geqslant 1 , \quad (34)$$

where

$$D_m(s,t) = D_1(s,t) \exp\left\{ h(-t) * \sum_{j=1}^{m-1} [D_j(s,t) - 1] \right\},$$
$$D_1(s,t) = e^{-s(t)} .$$

### *2. Multifold and singlefold moment generating function*

The $L$-fold moment generating function for the numbers of counts in the intervals $[t_j, t_j + T_j]$, $j = 1,2,\ldots,L$, can be obtained from the moment generating functional $L_m(s)$ by the substitution

$$s(t) = \mathbf{s}\mathbf{v}^{\dagger}(t) , \quad (35)$$

where $\mathbf{s}$ and $\mathbf{v}(t)$ are vectors defined by

$$\mathbf{s} = (s_1, s_2, \ldots, s_L) ,$$
$$\mathbf{v}(t) = (v_1(t), v_2(t), \ldots, v_L(t)) ,$$
$$v_j(t) = \begin{cases} 1, & t_j \leqslant t \leqslant t_j + T_j , \\ 0, & \text{otherwise} , \quad j = 1,2,3,\ldots,L . \end{cases}$$

The symbol $\dagger$ indicates vector transposition. This results in

$$Q_1(\mathbf{s}) = \exp\left\{ \mu \int_{-\infty}^{\infty} [\exp\{ -\mathbf{s}\mathbf{v}^{\dagger}(t) \} - 1]\, dt \right\},$$
$$Q_m(\mathbf{s}) = \exp\left\{ \mu \int_{-\infty}^{\infty} [\exp\{ q_1(q_2(q_3 \cdots q_{m-1}(\mathbf{s}\mathbf{v}^{\dagger}(t)))) \} - 1]\, dt \right\}, \quad m \geqslant 2 . \quad (36)$$

For identical branching, it follows that

$$Q_m(\mathbf{s}) = \exp\left\{ \mu \int_{-\infty}^{\infty} [D_m(\mathbf{s},t) - 1]\, dt \right\}, \quad m \geqslant 1 , \quad (37)$$

where

$$D_m(\mathbf{s},t) = D_1(\mathbf{s},t) \exp\left\{ h(-t) * \sum_{j=1}^{m-1} [D_j(\mathbf{s},t) - 1] \right\},$$
$$D_1(\mathbf{s},t) = \exp\left\{ - \sum_{j=1}^{L} s_j\, v_j(t) \right\} .$$

Equation (37) will be used to determine the correlation function and power spectral density for the process.

The statistical properties of $N_m(T)$, the number of counts in an interval $[0,T]$ at the $m$th stage, may be determined from the singlefold moment generating function, which is readily obtained from (36) by substituting $L = 1$:

$$Q_1(s) = \exp\left\{ \mu \int_{-\infty}^{\infty} [\exp\{ -s v(t) \} - 1]\, dt \right\}, \quad (38)$$
$$Q_m(s) = \exp\left\{ \mu \int_{-\infty}^{\infty} [\exp\{ q_1(q_2(q_3 \cdots q_{m-1}(s v(t)))) \} - 1]\, dt \right\}, \quad m \geqslant 2.$$

This recurrence relation is difficult to use unless the branching stages are identical (homogeneous branching), in which case it reduces to

$$Q_m(s) = \exp\left\{\mu \int_{-\infty}^{\infty} [D_m(s,t) - 1]\, dt\right\}, \quad m \geqslant 1, \quad (39)$$

with

$$D_m(s,t) = D_1(s,t) \exp\left\{h(-t) * \sum_{j=1}^{m-1} [D_j(s,t) - 1]\right\},$$

$$D_1(s,t) = e^{-sv(t)},$$

$$v(t) = \begin{cases} 1, & 0 \leqslant t \leqslant T, \\ 0, & \text{otherwise}. \end{cases}$$

### 3. Moments

The $n$th ordinary moment of $N_m(T)$ follows directly from the singlefold mgf by means of the relation[11]

$$\langle N_m^n(T)\rangle = (-1)^n \frac{\partial^n}{\partial s^n} Q_m(s)\bigg|_{s=0}. \quad (40)$$

Using (39) and (40), the recurrence relation for the moments (in the special case of homogeneous branching) becomes

$$\langle N_m^{n+1}(T)\rangle = \langle N_m(T)\rangle \sum_{k=0}^{n} \binom{n}{k} \langle N_m^{n-k}(T)\rangle I_m^{(k+1)}, \quad m \geqslant 2, \quad (41)$$

where

$$I_m^{(1)} = 1,$$

$$I_m^{(k+1)} = \frac{1}{T(1+\alpha)^{m-1}} \int_{-\infty}^{\infty} D_m^{(k+1)}(t)\, dt,$$

$$D_m^{(k+1)}(t) = v(t) D_m^{(k)}(t) + \sum_{l=0}^{k} \binom{k}{l} D_m^{(k-l)}(t) \times \left[h(-t) * \sum_{j=1}^{m-1} D_j^{(l+1)}(t)\right],$$

$$D_m^{(0)}(t) = 1 \quad \text{for all } t,$$

$$D_1^{(k)} = v(t), \quad k \geqslant 1.$$

This should be compared with the expression for the instantaneous case given in (19).

For homogeneous branching, the mean number of counts is

$$\langle N_m(T)\rangle = \langle N_m^1(T)\rangle = \mu T(1+\alpha)^{m-1}, \quad (42)$$

and the variance of $N_m(T)$ is

$$\mathrm{Var}[N_m(T)] = \langle N_m(T)\rangle I_m^2, \quad m \geqslant 2, \quad (43)$$

with

$$I_m^2 = \frac{1}{T(1+\alpha)^{m-1}} \int_{-\infty}^{\infty} D_m^{(2)}(t)\, dt,$$

$$D_m^{(2)}(t) = \{D_m^{(1)}(t)\}^2 + h(-t) * \sum_{j=1}^{m-1} D_j^{(2)}(t),$$

$$D_m^{(1)}(t) = v(t) + h(-t) * \sum_{j=1}^{m-1} D_j^{(1)}(t),$$

$$D_1^{(1)}(t) = v(t).$$

In the limit of long counting times

$$\mathrm{Var}[N_m] = \mu T(1+\alpha)^{m-2}[(2+\alpha)(1+\alpha)^{m-1} - 1], \quad (44)$$

in accord with (17) for instantaneous branching. Though higher statistical properties are difficult to compute for nonhomogeneous branching, the count mean and variance can be obtained.

For this purpose, we consider the representation provided in Fig. 3(c), where $W_m(t)$ is a stationary, zero-mean, white random process, with a cross-correlation function given by

$$R_{W_i W_j}(\tau) = \langle W_i(t+\tau) W_j(t)\rangle = \langle X_j(t)\rangle^{1/2} \delta(t) \delta_{ij}. \quad (45)$$

$\delta(t)$ and $\delta_{ij}$ are the Dirac and Kronecker delta functions, respectively. The system in Fig. 3(c) turns out to be identically equivalent to the one in Fig. 3(b) as far as computation of the first and second moments are concerned.[7,12,13] A straightforward calculation provides

$$\langle N_m(T)\rangle = \mu T \prod_{k=1}^{m-1} (1+\alpha_k), \quad m \geqslant 2 \quad (46)$$

and

$$\mathrm{Var}[N_m(T)] = \mu \sum_{k=0}^{m-1} C_k \int_{-T}^{T} (T - |\tau|) \times \mathop{*}_{r=k+1}^{m-1} [\delta(\tau) + h_r(\tau) + h_r(-\tau) + g_r(\tau)]\, d\tau, \quad m \geqslant 2, \quad (47)$$

where

$$C_0 = 1,$$

$$C_1 = 1,$$

$$C_k = \alpha_k \prod_{r=1}^{k-1} (1+\alpha_r), \quad k \geqslant 2,$$

$$\alpha_r = \int_{-\infty}^{\infty} h_r(t)\, dt,$$

$$g_r(\tau) = h_r(\tau) * h_r(-\tau),$$

$$\mathop{*}_{r=i}^{j} [\delta(\tau) + h_r(\tau) + h_r(-\tau) + g_r(\tau)] = \delta(\tau) \text{ for } j < i.$$

The symbol $*_{k=1}^{n}$ indicates $n$-fold convolution. The Fano factor is therefore

$$F_m(T) = \left[\prod_{k=1}^{m-1} (1+\alpha_k)\right]^{-1} \sum_{k=0}^{m-1} C_k \int_{-T}^{T} \left(1 - \frac{|\tau|}{T}\right) \times \mathop{*}_{r=k+1}^{m-1} [\delta(\tau) + h_r(\tau) + h_r(-\tau) + g_r(\tau)]\, d\tau, \quad m \geqslant 2. \quad (48)$$

When all $\alpha_j$ are identical and equal to $\alpha$, (46) and (47) reduce to (42) and (43), respectively. In the limit of long counting times, the process is effectively instantaneous and the above expressions for the mean, variance, and Fano factor become (13), (14), and (15), with $a = \mu T$, respectively. In the special case $m = 2$, (46)–(48) reproduce the previously obtained results for the Thomas point process.[6]

Because of the importance of the Fano factor as a simple measure characterizing the departure of a process from the HPP, we carry out a parametric study of its dependence in our branching process. For simplicity, we assume that the impulse response functions $h_m(t)$ are identical at each stage, and have the simple exponential form

$$h(t)=\begin{cases}(2\alpha/\tau_p)\exp(-2t/\tau_p), & t\geqslant 0,\\ 0, & t<0.\end{cases} \tag{49}$$

Here $\tau_p/2$ is the characteristic decay time of the filter and $\alpha$ is the area under the function.

In Fig. 4, we plot the Fano factor $F_m(T)$ versus the number of stages $m$, with $2T/\tau_p$ and $\alpha$ as parameters. All of the curves are monotonically increasing functions of $m$ (as are the underlying mean and variance curves). This is to be contrasted with the results for the cascaded Poisson process that we studied earlier,[7] in which the mean and variance decay with increasing $m$ if $\alpha<1$. The distinction arises because of the presence of the feed-forward path [shown in Fig. 3(b)], whch distinguishes the present model as a branching process, rather than as a simple cascade of stages. For $T/\tau_p \gg 1$, the curves will obey (18), which provides essentially exponential growth (straight-line behavior on a logarithmic ordinate). For $T/\tau_p \ll 1$, the particlelike clusters of the points in the process are chopped apart by the small sampling time, leading to the independence that is characteristic of the HPP.[6] Indeed, as the curves for $2T/\tau_p=0.01$ show, $F_m(T)$ remains essentially constant at unity, up to four stages. The small residual clustering is amplified as $m$ increases above this value. Increasing values of $\alpha$, of course, correspond to increased clustering.

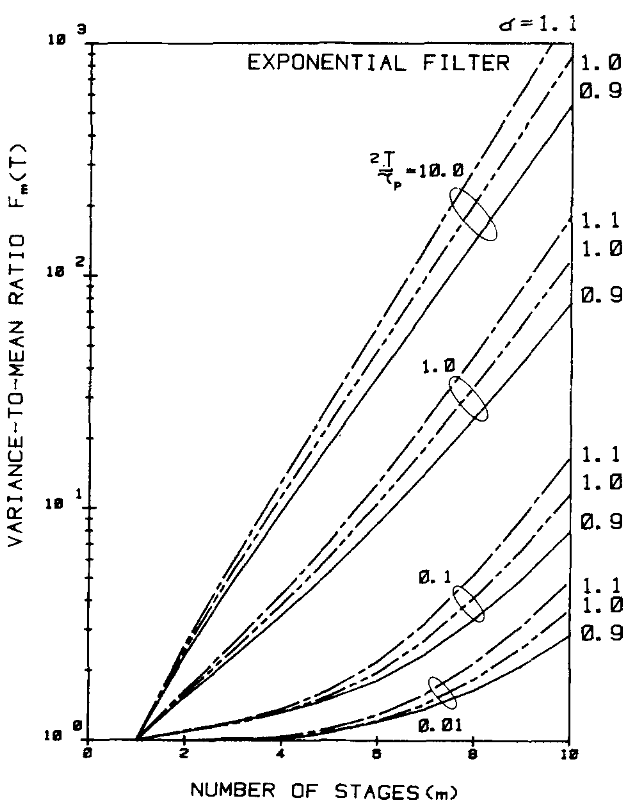


FIG. 4. Count variance-to-mean ratio (Fano factor) $F_m(T) = \mathrm{Var}[N_m(T)]/\langle N_m(T)\rangle$ vs number of stages $m$, with $2T/\tau_p$ and $\alpha$ as parameters. The impulse response functions $h_m(t)$ are all assumed to be identical, exponentially decaying functions with time constant $\tau_p/2$.

### 4. Counting probability distribution

The counting probability distribution function of $N_m(T)$ can be derived by using (11), (20), and (38) from which it follows that

$$p_m(0)=\exp\left\{\mu\int_{-\infty}^{\infty}[E_m^{(0)}(t)-1]\,dt\right\}, \tag{50}$$

$$(n+1)p_m(n+1)=\langle N_m(T)\rangle\sum_{k=0}^{n}p_m(n-k)J_m^{(k+1)},$$

where

$$J_m^{(k+1)}=\frac{(-1)^{k+1}}{T(1+\alpha)^{m-1}k!}\int_{-\infty}^{\infty}E_m^{(k+1)}(t)\,dt,$$

$$E_m^{(k+1)}(t)=E_m^{(1)}Y_m^{(k)}(t)+\sum_{l=0}^{k}\binom{k}{l}E_m^{(k-l)}(t)h(-t)*\sum_{j=1}^{m-1}E_j^{(l+1)}(t),$$

$$Y_m^{(k+1)}(t)=\sum_{l=0}^{k}\binom{k}{l}Y_m^{(k-l)}h(-t)*\sum_{j=1}^{m-1}E_j^{(l+1)}(t),$$

$$Y_m^{(0)}(t)=\exp\left\{h(-t)*\sum_{j=1}^{m-1}[E_j^{(0)}(t)-1]\right\},$$

$$E_m^{(0)}(t)=\begin{cases}0, & 0\leqslant t\leqslant T,\\ \exp\left\{h(-t)*\sum_{j=1}^{m-1}[E_j(t)-1]\right\}, & \text{otherwise},\end{cases}$$

$$E_1^{(0)}(t)=\begin{cases}0, & 0\leqslant t\leqslant T,\\ 1, & \text{otherwise},\end{cases}$$

$$E_1^{(1)}(t)=\begin{cases}-1, & 0\leqslant t\leqslant T\\ 0, & \text{otherwise},\end{cases}$$

$$E_1^{(k)}(t)=0 \quad \text{for all } t,\ k\geqslant 2.$$

Equation (50) reduces to (21) in the limit $T/\tau_p \gg 1$. As $T/\tau_p$ is reduced, $F_m(T)$ will decrease (see Fig. 4), and the counting distributions will narrow. The transition in $p_m(n)$ vs $n$ will not be unlike that demonstrated for the cascaded Poisson process (see Ref. 7, Fig. 8).

### 5. Autocorrelation function and power spectral density

In this subsection, we derive the autocorrelation function and the power spectral density for the Poisson branching point process. The autocorrelation function $r_m(\tau)$ is defined as

$$r_m(\tau)=\lim_{\Delta t\to 0}\frac{1}{(\Delta t)^2}\langle \Delta N_m(t)\Delta N_m(t+\tau)\rangle, \tag{51}$$

where the quantity $\Delta N_m(t)$ represents the number of counts in the time interval $[t,t+\Delta t]$, at the $m$th stage. The equation for $\langle \Delta N_m(t)\Delta N_m(t+\tau)\rangle$ may be obtained from (37) by substituting

$$L=2,$$

$$v_1(t)=\begin{cases}1, & 0\leqslant t\leqslant \Delta t,\\ 0, & \text{otherwise},\end{cases}$$

$$v_2(t)=\begin{cases}1, & \tau\leqslant t\leqslant \tau+\Delta t,\\ 0, & \text{otherwise}.\end{cases}$$

Differentiating (37) with respect to $s_1$ and $s_2$, substituting $s_1=s_2=0$, and letting $\Delta t\to 0$ leads to (see Appendix)

$$r_m(\tau) = \{ \mu(1+\alpha)^{m-1}\}^2 + \mu \int_{-\infty}^{\infty} Y_m(\omega)\, e^{j\omega\tau} \frac{d\omega}{2\pi}, \quad (52a)$$

where

$$Y_m(\omega) = |1 + H(\omega)|^{2(m-1)} + \alpha(1+\alpha)^{m-2} \times \frac{1 - [|1+H(\omega)|^2/(1+\alpha)]^{m-1}}{1 - |1+H(\omega)|^2/(1+\alpha)}. \quad (52b)$$

$H(\omega)$ is the Fourier transform of $h(t)$. Substituting $\tau = 0$ into the second term of (52a) yields the variance

$$\mathrm{Var}\,[dN_m(t)] = \mu \int_{-\infty}^{\infty} Y_m(\omega) \frac{d\omega}{2\pi}, \quad (53)$$

which represents the power fluctuations of the process $dN_m(t)$ in the infinitesimal duration $\Delta t$.

The power spectral density $s_m(\omega)$ is defined as the Fourier transform of the autocorrelation function $r_m(\tau)$, which is clearly

$$s_m(\omega) = 2\pi \{ \mu(1+\alpha)^{m-1}\}^2 \delta(\omega) + \mu Y_m(\omega). \quad (54)$$

The first term of (54) represents the dc power of the process $dN_m(t)$, whereas the second term represents the frequency distribution of the ac power, which depends on the shape of the impulse response function $h(t)$ through $H(\omega)$.

The autocorrelation function between the number of counts in the interval $T$, separated by a time delay $\tau$, is defined as

$$R_m(\tau) = \langle [N_m(t+T) - N_m(t)] \times [N_m(t+T+\tau) - N_m(t+\tau)] \rangle, \quad (55)$$

which can be easily obtained from (52a) by means of

$$R_m(\tau) = \int_0^T \int_0^T r_m(t_1 - t_2 + \tau)\, dt_1\, dt_2. \quad (56)$$

Substituting (52a) into (56) gives rise to

$$R_m(\tau) = \{ \mu T(1+\alpha)^{m-1}\}^2 + \mu T \int_{-\infty}^{\infty} Y_m(\omega) \Phi_T(\omega) e^{j\omega\tau} \frac{d\omega}{2\pi}, \quad (57a)$$

where

$$\Phi_T(\omega) = T\, [\sin(\omega T/2)/(\omega T/2)]^2. \quad (57b)$$

Substituting $\tau = 0$ into the second term of (57a) leads to the variance of the counting process,

$$\mathrm{Var}[N_m(T)] = \mu T \int_{-\infty}^{\infty} Y_m(\omega) \Phi_T(\omega) \frac{d\omega}{2\pi}, \quad (58)$$

which is the frequency-domain representation of (43). The power spectral density for the counts is easily obtained by taking the Fourier transform of (57a), which provides

$$S_m(\omega) = 2\pi \{ \mu T(1+\alpha)^{m-1}\}^2 \delta(\omega) + \mu T Y_m(\omega) \Phi_T(\omega). \quad (59)$$

## IV. POISSON BRANCHING POINT PROCESS IN THE LIMIT OF CONTINUOUS BRANCHING

### A. Introduction

In this section we investigate properties of the Poisson branching point process in the limit of an infinite number of branching stages, when the branching at each stage is infinitesimal. Thus we allow

$$m \to \infty, \quad (22a)$$

$$\alpha \to 0, \quad (22b)$$

with the product

$$m\alpha = x \quad (22c)$$

remaining finite. In this limit we replace the discrete index $m$, which has been used throughout Sec. III to indicate the branching stage number, with the continuous index $x$. Thus $L_m$, $Q_m$, $N_m$,... become $L_x$, $Q_x$, $N_x$, ..., respectively. Furthermore we define a normalized impulse response function $h_0(t)$ such that

$$h(t) = \alpha h_0(t) \quad (60)$$

and

$$\int_{-\infty}^{\infty} h_0(t)\, dt = 1.$$

By applying this limit to the expression derived in Sec. III, we obtain a number of results that form a simple generalization of the Yule–Furry process. Their application to the generation of cosmic ray showers is likely to be useful.

### B. Results

We are able to obtain results for the moment generating functional and moment generating function in the case of instantaneous branching, when the initial process is Poisson. These are, of course, identical to those for the Poisson-driven Yule–Furry process, as provided in Sec. II C. General results, with arbitrary time dynamics, have been derived for the count mean, variance, and Fano factor, and for the autocorrelation function and power spectral density of the point process. It will be evident in the following that the count mean and variance depend critically on $m$. The results below should be compared with those provided in Secs. II C and III C.

#### *1. Moment generating functional*

The moment generating functional (34) becomes

$$L_x(s) = \exp\left\{ \mu \int_{-\infty}^{\infty} [D_x(s,t) - 1]\, dt \right\}, \quad (61a)$$

where $D_x(s,t)$ satisfies the nonlinear integro-differential functional equation

$$\frac{\partial}{\partial x} D_x(s,t) = D_x(s,t)\, \{ h_0(-t) * [D_x(s,t) - 1] \}, \quad (61b)$$

with the initial condition

$$D_0(s,t) = e^{-s(t)}. \quad (61c)$$

We are unable to obtain a general solution to (61b). However, in the simple special case where

$$h_0(t) = \delta(t), \quad (62)$$

(61b) can be shown to have the solution

$$D_x(s,t) = \frac{e^{-x} e^{-s(t)}}{1 - (1 - e^{-x}) e^{-s(t)}}. \quad (63)$$

The moment generating functional is then

$$L_x(s) = \exp\left\{ -\mu \int_{-\infty}^{\infty} \frac{1 - e^{-s(t)}}{1 - (1 - e^{-x}) e^{-s(t)}}\, dt \right\}. \quad (64)$$

### 2. Moment generating function

The moment generating function $Q_x(s)$ of the random variable $N_x(T)$ may be obtained from the moment generating functional $L_x(x)$ by setting $s(t) = sv(t)$. Equation (61b) is then a nonlinear integro-differential equation which is difficult to solve. In the special case of instantaneous branching, we can use (64) to obtain

$$Q_x(s) = \exp\left\{ -\mu T \frac{1 - e^{-s}}{1 - (1 - e^{-x}) e^{-s}} \right\}, \qquad (65)$$

which is identical to (24) with $a = \mu T$, as it should be. Equations (24) and (65) are identified as the moment generating function of a Yule–Furry process driven by a homogeneous Poisson point process, as mentioned above.

### 3. Moments

It is possible to obtain expressions for the mean and variance of $N_x(T)$ for an arbitrary impulse response function $h_0(t)$. Applying the limits of (22) on (42) leads to

$$\langle N_x(T) \rangle = \mu T e^x . \qquad (66)$$

Note that (66) is identical to (26) with $a = \mu T$. A similar operation on (52b) yields

$$Y_x(\omega) = \lim_{m \to \infty} Y_m(\omega) = \frac{[H(\omega) + H(-\omega)] e^{x[H(\omega) + H(-\omega)]} - e^x}{H(\omega) + H(-\omega) - 1}, \qquad (67)$$

so that the count variance is [see (58)]

$$\mathrm{Var}[N_x(T)] = \mu T \int_{-\infty}^{\infty} Y_x(\omega) \Phi_T(\omega) \frac{d\omega}{2\pi}. \qquad (68)$$

Here $H(\omega)$ is the Fourier transform of $h_0(t)$ (the transfer function of the filtering system), $H(-\omega)$ is the complex conjugate of $H(\omega)$, and the function $\Phi_T(\omega)$ is given in (57b). Using Eqs. (66) and (68), the Fano factor becomes

$$F_x(T) = \int_{-\infty}^{\infty} \Phi_T(\omega) \times \frac{[H(\omega) + H(-\omega)]\, e^{x[H(\omega) + H(-\omega) - 1]} - 1}{H(\omega) + H(-\omega) - 1} \frac{d\omega}{2\pi}. \qquad (69)$$

For the case of instantaneous multiplication, $H(\omega) = 1$ for all $\omega$ so that (68) and (69) reduce to the Poisson-driven Yule–Furry results

$$\mathrm{Var}[N_x(T)] = \mu T e^x (2e^x - 1) \qquad (70)$$

and

$$F_x = 2e^x - 1, \qquad (71)$$

respectively. Of course, (70) and (71) are then identical with (27a) and (27b) with $a = \mu T$.

To assess the effects of the characteristic decay time $\tau_p$ of the filter $h_0(t)$ on the fluctuation properties of the counting process $N_x(T)$, we consider a simple example. We make use of the ideal low-pass filter transfer function

$$H(\omega) = \begin{cases} 1, & |\omega| \leqslant \omega_c, \\ 0, & \text{otherwise}, \end{cases} \qquad (72)$$

where $\omega_c/2 = 1/\tau_p$. It can be shown that (69) then leads to

$$F_x = [2e^x - 1]\, \xi(T/\tau_p) + [1 - \xi(T/\tau_p)], \qquad (73a)$$

where

$$\xi(\beta) = (2/\pi)\,\mathrm{Si}(\beta/2) - (4/\beta)[1 - \cos(\beta/2)] \qquad (73b)$$

and

$$\mathrm{Si}(\beta) = \int_0^\beta \frac{\sin(y)}{y} dy. \qquad (73c)$$

In Fig. 5 we plot the Fano factor $F_x(T)$ as a function of the branching parameter $x$, with the ratio $\beta = T/\tau_p$ as a parameter. In the limit $T \gg \tau_p$, $\xi(T/\tau_p) \to 1$, and

$$F_x = 2e^x - 1 \quad \text{for } T \gg \tau_p, \qquad (73d)$$

in accord with the (instantaneous) results presented in (71). In the opposite limit ($T \ll \tau_p$), $\xi(T/\tau_p) \to 2T/\tau_p$, corresponding to a reduced Fano factor

$$F_x = [2e^x - 1](2T/\tau_p) + 1 - (2T/\tau_p) \quad \text{for } T \ll \tau_p. \qquad (74)$$

It is apparent from (74) and from Fig. 5 that as $T/\tau_p$ decreases, the Fano factor, and therefore the degree of fluctuation, decreases. The reason for this, once again, is the cutting apart of the particlelike clusters of multiplied events.

### 4. Autocorrelation function and power spectral density

The autocorrelation function and power spectral density for the process $dN_x(t)$ may be determined by taking the limit of (52a) and (54), respectively. The results are

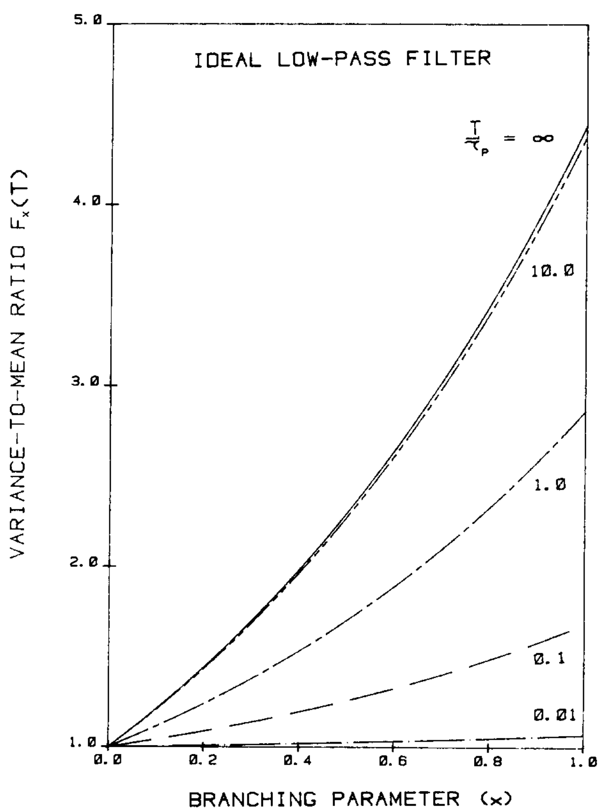


FIG. 5. Fano factor $F_x(T)$ as a function of the branching parameter $x$, with $T/\tau_p$ as a parameter. In this example of continuous branching, the time dependence of the process is represented by an impulse-response function whose Fourier transform is an ideal low-pass filter.

$$r_x(\tau) = \mu^2 e^{2x} + \mu \int_{-\infty}^{\infty} Y_x(\omega)\, e^{j\omega\tau} \frac{d\omega}{2\pi} \tag{75}$$

and

$$s_x(\omega) = 2\pi\mu^2 e^{2x}\delta(\omega) + \mu Y_x(\omega), \tag{76}$$

where $Y_x(\omega)$ is given by (67).

The autocorrelation function of the counts $N_x(T)$ for the infinite branching case is obtained from (75) by using (56). This provides

$$R_x(\tau) = (\mu T)^2 e^{2x} + \mu T \int_{-\infty}^{\infty} Y_x(\omega)\Phi_T(\omega)\, e^{j\omega\tau} \frac{d\omega}{2\pi}. \tag{77}$$

The power spectral density in this case is

$$S_x(\omega) = 2\pi(\mu T)^2 e^{2x}\delta(\omega) + \mu T Y_x(\omega)\Phi_T(\omega), \tag{78}$$

corresponding to (59).

In Fig. 6, we present the power spectral density for the Poisson branching point process $s_m(\omega\tau_p)$ versus normalized frequency $\omega\tau_p$ [see (54) and (76)] with $m$ as a parameter. For the purposes of illustration, we have chosen an exponential impulse response function [see (49)] and ignored the delta function at $\omega\tau_p = 0$. The product $m\alpha = x$ was maintained constant for each plot [$m\alpha = x = 1.0$ in Fig. 6(a); $m\alpha = x = 4.0$ in Fig. 6(b)]. This enables us to follow the behavior of $s_m(\omega\tau_p)$ as $m$ increases toward the continuous limit ($m = \infty$). The driving rate was adjusted in all cases to be $\mu = (1+\alpha)^{-m} = (1 + 1/m)^{-m}$ so that the rate of the final point processes is unity. For the parameters shown, it is evident that the curves are of very similar shape, although their absolute and relative magnitudes are strongly dependent on $m$ and on $m\alpha = x$.

Finally, we note that while we generally think of $x$ as position in a continuum of branching stages, and $t$ as time, it may be more appropriate in some applications to regard the variable $x$ as time along which branching progresses, and $t$ as position. In such an interpretation, $h(t)$ will indicate diffusion or migration of particles in space, and $N_x(T)$ the number of particles in the space $[0,T]$ at the time $x$.

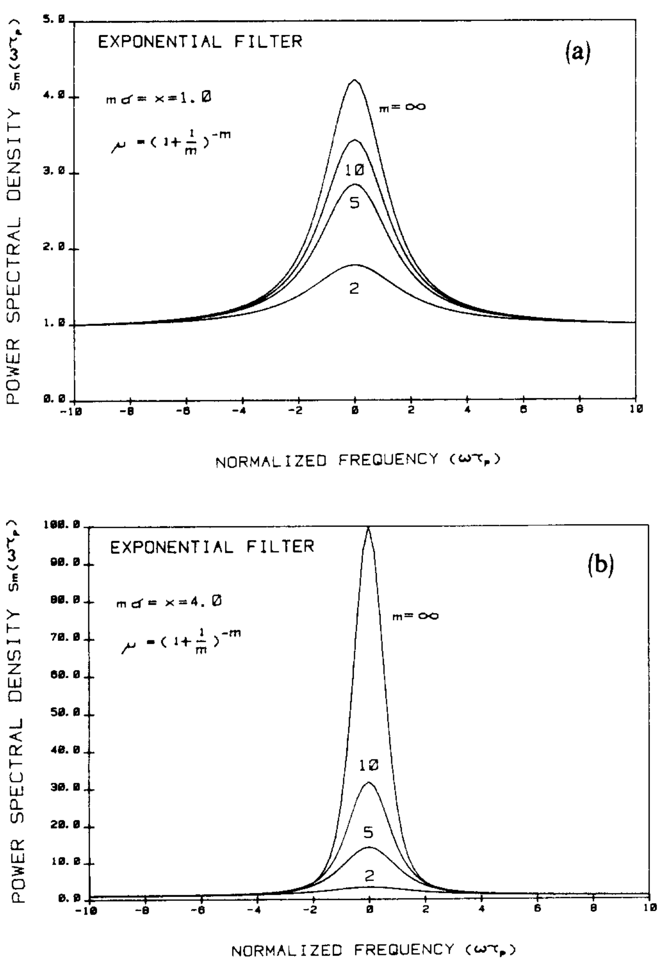


FIG. 6. Power spectral density for the Poisson branching point process $s_m(\omega\tau_p)$ vs normalized frequency $\omega\tau_p$, with $m$ as a parameter. For the purposes of this illustration, we have chosen an exponential impulse response function, and eliminated the delta function at $\omega\tau_p = 0$. The driving rate $\mu = (1+\alpha)^{-m} = (1+1/m)^{-m}$ in all cases. (a) $m\alpha = x = 1.0$; (b) $m\alpha = x = 4.0$.


## ACKNOWLEDGMENTS

This work was supported by the Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Contract DAAG29-82-K-0080 and by the National Science Foundation under Grant ECS-82-19636.


## APPENDIX: DERIVATION OF THE CORRELATION FUNCTION $r_m(\tau)$ FOR THE POISSON BRANCHING POINT PROCESS

Differentiating (37) with respect to $s_1$ and $s_2$, and setting $s_1 = s_2 = 0$, provides

$$\begin{aligned} r_m(\tau) &= \lim_{\Delta t \to 0} \frac{1}{(\Delta t)^2} \langle \Delta N_m(t) \Delta N_m(t+\tau) \rangle \\ &= \mu^2 \int_{-\infty}^{\infty} \Phi_m^{(1)}(t)dt \int_{-\infty}^{\infty} \Phi_m^{(2)}(t)dt \\ &\quad + \mu \int_{-\infty}^{\infty} \Phi_m^{(3)}(t)dt, \end{aligned} \tag{A1}$$

where

$$\begin{aligned} \Phi_m^{(1)}(t) &= \Phi_1^{(1)}(t) + h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(1)}(t), \\ \Phi_m^{(2)}(t) &= \Phi_1^{(2)}(t) + h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(2)}(t), \\ \Phi_m^{(3)}(t) &= \Phi_1^{(3)}(t) + \Phi_1^{(1)}(t)\Big[h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(2)}(t)\Big] \\ &\quad + \Phi_1^{(2)}(t)\Big[h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(1)}(t)\Big] \\ &\quad + h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(3)}(t) \\ &\quad + \Big[h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(1)}(t)\Big] \\ &\quad \times \Big[h(-t) * \sum_{j=1}^{m-1} \Phi_j^{(2)}(t)\Big], \end{aligned} \tag{A2}$$

with the initial conditions

$$\begin{aligned} \Phi_1^{(1)}(t) &= -\delta(t), \\ \Phi_1^{(2)}(t) &= -\delta(t-\tau), \\ \Phi_1^{(3)}(t) &= \delta(t)\delta(\tau). \end{aligned} \tag{A3}$$

Taking the Fourier transform of (A2) and (A3) to obtain the frequency-domain equivalent of Eq. (A1) provides

$$r_m(\tau) = \mu^2 \widetilde{\Phi}_m^{(1)}(0)\widetilde{\Phi}_m^{(2)}(0) + \mu\widetilde{\Phi}_m^{(3)}(0)\,, \tag{A4}$$

where $\widetilde{\Phi}_m^{(i)}(0)$ is the Fourier transform of $\Phi_m^{(i)}(t)$ evaluated at $\omega = 0$. A simple calculation shows that the first term in (A4) is

$$\mu^2\widetilde{\Phi}_m^{(1)}(0)\widetilde{\Phi}_m^{(2)}(0) = \{\mu(1+\alpha)^{m-1}\}^2\,, \tag{A5}$$

whereas the second term of (A4) is

$$\mu\widetilde{\Phi}_m^{(3)}(0) = \mu\int_{-\infty}^{\infty} Y_m(\omega)e^{j\omega\tau}\frac{d\omega}{2\pi}\,, \tag{A6}$$

with $Y_m(\omega)$ as given in (52b).


[1]T. E. Harris, *The Theory of Branching Processes* (Springer, New York, 1963).
[2]W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1968), 3rd ed., Vol. I, Chap. XII.
[3]B. E. A. Saleh and M. C. Teich, "Multiplied-Poisson noise in pulse, particle, and photon detection," Proc. IEEE **70**, 229–245 (1982).
[4]M. C. Teich and B. E. A. Saleh, "Interevent-time statistics for shot-noise driven self-exciting point processes in photon detection," J. Opt. Soc. Am. **71**, 771–776 (1981).
[5]B. E. A. Saleh and M. C. Teich, "Statistical properties of a nonstationary Neyman–Scott cluster process," IEEE Trans. Inform. Theory **IT-29**, 939–941 (1983).
[6]K. Matsuo, M. C. Teich, and B. E. A. Saleh, "Thomas point process in pulse, particle, and photon detection," Appl. Opt. **22**, 1898–1909 (1983).
[7]K. Matsuo, B. E. A. Saleh, and M. C. Teich, "Cascaded Poisson Processes," J. Math. Phys. **23**, 2353–2364 (1982).
[8]M. Thomas, "A generalization of Poisson's binomial limit for use in ecology," Biometrika **36**, 18–25 (1949).
[9]M. C. Teich, "Role of the doubly stochastic Neyman Type-A and Thomas counting distributions in photon detection," Appl. Opt. **20**, 2457–2467 (1981).
[10]E. Parzen, *Stochastic Processes* (Holden-Day, San Francisco, 1962).
[11]B. E. A. Saleh, *Photoelectron Statistics* (Springer, New York, 1978).
[12]A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).
[13]D. L. Snyder, *Random Point Processes* (Wiley, New York, 1975).

# Orthogonal polynomials with exponential weight in a finite interval and application to the optical model

R. Mach
*Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Rez, Czechoslovakia*

(Received 28 December 1982; accepted for publication 18 November 1983)

A quadrature procedure is developed which makes the construction of momentum-space meson-nucleus optical potentials more accurate. We deal with numerical evaluation of integrals with finite $t$-integration range which contain $\exp(Dt)$ explicitly, where $D$ is a parameter. The Gaussian rule is used with abscissas determined as roots of orthogonal polynomials with exponential weight function in the interval $[-1,1]$. Recurrence relations and inequalities for these polynomials are obtained. A nonlinear recursion is derived, which permits the evaluation of abscissas and weights without accumulation of roundoff error. The nonlinear recursion is solved by means of an iterative procedure, the convergence properties of which are established. The quadrature procedure is summarized as an easily implementable algorithm. The rate of convergence is demonstrated for several test integrals.

PACS numbers: 02.60.Jh, 02.30.Bi

## I. INTRODUCTION

Meson-nucleus scattering at medium energies is currently studied in the framework of multiple scattering theory. The meson-nucleus scattering amplitude is obtained as a solution of Lippmann–Schwinger or coupled-channel equations, where the optical potential (or potential matrix) is typically of the form[1]

$$V_l(p',p,E)=(2l+1)A\int_{-1}^{1}P_l(\cos\vartheta)\times t(\mathbf{p}',\mathbf{p},E)\,F(q)\,d(\cos\vartheta). \qquad (1.1)$$

Here, $P_l(\cos\vartheta)$ are Legendre polynomials, $\cos\vartheta=\mathbf{p}'\cdot\mathbf{p}/(p'p)$ and $l$ labels the meson-nucleus partial waves. The elementary meson-nucleon amplitude is usually given in terms of the partial wave decomposition

$$t(\mathbf{p}',\mathbf{p},E)=\sum_{\lambda=0}(2\lambda+1)t_\lambda(p',p,E)\,P_\lambda(\cos\vartheta) \qquad (1.2)$$

and the nuclear form factor can be represented as

$$F(q)=\exp\left(-\frac{a^2q^2}{4}\right)Q_n(q^2),\quad q=|\mathbf{p}'-\mathbf{p}|. \qquad (1.3)$$

Here, $Q_n(q^2)$ is a polynomial and $a$ is related to the nuclear radius. Since $q^2=p'^2+p^2-2p'p\cos\vartheta$, from Eqs. (1.1)–(1.3) we have

$$V_l(p',p,E)\sim\int_{-1}^{1}e^{Dt}\overline{Q}_m(p',p,E;t)dt, \qquad (1.4)$$

where $D=0.5\,p'pa^2$ and $\overline{Q}_m=\overline{Q}_m(p',p,E;t)$ is a polynomial in the variable $t=\cos\vartheta$. The degree of the polynomial increases with the increasing mass number $A$ and the energy $E$. In typical medium energy calculations it does not exceed ten or twenty. Relativistic and Fermi motion corrections spoil somewhat the polynomial behavior of $\overline{Q}_m$; however, their role at intermediate energies is not of crucial importance.[2]

The angular integration indicated in (1.4) is to be performed with high accuracy, since the optical potential $V_l(p',p,E)$ enters the kernel of Lippmann–Schwinger or coupled-channel equations and it is necessary to ensure that the resulting meson-nucleus amplitudes are not biased by numerical uncertainty and that they reflect actual physical assumptions made in constructing the optical potential.[3] With increasing $D$, the function $\exp(Dt)\,\overline{Q}_m$ represents a more and more narrow peak in the vicinity of $t=1$. Therefore, the usual methods of evaluating (1.4), e.g., Gauss–Legendre quadrature, are rather awkward for momenta $p'$ and $p$ higher than typical nuclear values ($\sim 1/a$), since either only few abscissas fall into the region, where $\exp(Dt)\overline{Q}_m$ is actually concentrated, or the number of prints in the quadrature rule becomes impractically large.

The aim of the present paper is to develop an efficient and numerically stable procedure for evaluation of the integrals

$$I=\frac{1}{2}\int_{-1}^{1}\omega(t)f(t)dt, \qquad (1.5)$$

where $\omega(t)=\exp(Dt)$ is the weight function, $D$ is a real parameter, and $f(t)$ is a function which can be approximated to good accuracy by a polynomial. The Gauss quadrature rule will be applied to Eq. (1.5), i.e., the integral $I$ is approximated by $I_N$, where

$$I_N=\sum_{i=1}^{N}\lambda_i f(t_i). \qquad (1.6)$$

The method is based on the existence (for any $\omega(t)>0$) of a sequence of polynomials $\{S_n(t)\}_{n=0}^{\infty}$ which are orthogonal with respect to $\omega(t)$ and in which $S_n(t)$ is of exact degree $n$ so that

$$(S_n,S_m)=\frac{1}{2}\int_{-1}^{1}\omega(t)S_n(t)S_m(t)dt=h_n \quad \text{when } n=m$$
$$=0 \quad \text{when } n\neq m. \qquad (1.7)$$

The polynomial $S_N(t)=k_N\Pi_{i=1}^{N}(t-t_i)$, $k_N>0$, has $N$ real roots $-1<t_1<t_2<\cdots<t_N<1$. Further, the weights are given by

$$\lambda_i=-\frac{k_{N+1}h_N}{k_NS'_N(t_i)S_{N+1}(t_i)},\quad i=1,2,\ldots,N, \qquad (1.8)$$

 0022-2488/84/072186-08$02.50  

where $S'_N(t_i) = (dS(t)/dt)_{t=t_i}$. Note that $t_i$ and $\lambda_i$ depend on $N$, as well as $t_i, \lambda_i, h_N$, and $S_N(t)$ in our case depend on $D$. However, the dependence on $N$ and $D$ has been suppressed here to simplify the notation.

It can be shown that for $f(t) \in C^{2N}[-1,1]$

$$I = I_N + \frac{f^{(2N)}(\xi)h_N}{(2N)!k_N^2} \text{ and } -1<\xi<1 \tag{1.9}$$

holds, thus the Gaussian rule is exact for all polynomials of degree $\leqslant 2N-1$. Proofs of the statements (1.7)–(1.9) can be found in Ref. 4.

The abscissas $t_i$ can be, of course, chosen also in a different manner.[5] However it is for the property of the highest algebraic accuracy (1.9) that we prefer to use the Gauss quadrature. The property enables one to minimize the number of usually time-consuming evaluations of the integrand in (1.4).

The existence of the three term recurrence relation [for any $\omega(t)>0$]

$$S_{n+1}(t) = (\alpha_n t + \beta_n)S_n(t) - \gamma_n S_{n-1}(t), n = 0,1,...,N-1 \tag{1.10}$$

with $\alpha_n > 0,\ \gamma_n > 0,\ S_{-1}(t) = 0$, and $S_0(t) = 1$

makes it possible[6] to determine the roots $t_i$ and weights $\lambda_i$ by solving an eigenvalue problem provided that the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$ are known. The method is briefly reviewed in Sec. II.

A numerically stable algorithm is not known for evaluation of the coefficients of the three term recurrence relation (1.10) in the case of an arbitrary weight $\omega(t)>0$. This represents a serious difficulty in generating $t_i$ and $\lambda_i$. With the aim of developing a method for computation of the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$ in the case $\omega(t) = \exp(Dt)$, the properties of the corresponding orthogonal polynomials are investigated in Sec. III. Relations between the polynomials [we call them $P_n(D,t)$] and their derivatives are obtained. The links are established between $P_n(D,t)$ and Legendre and Laguerre polynomials. Further, we succeeded in finding a nonlinear recursion among the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$, which turned out to be very useful for practical purposes.

The nonlinear recursion can be solved by an iterative procedure, the convergence of which is proved in Sec. IV. An algorithm is given, which permits an easy and numerically stable evaluation of the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$ and, hence, of $t_i$ and $\lambda_i$, too. The rate of convergence of the quadrature rule is shown in Sec. V and compared with that of Gauss–Legendre rule in the case of several test integrals.

Section VI contains a summary and conclusions.

## II. GENERATING ABSCISSAS AND WEIGHTS

It was established more than twenty years ago[6] that a very powerful method for generating roots of orthogonal polynomials consists in rewriting the condition $S_N(t) = 0$ into the matrix form

$$T\,S(t) = t\,S(t). \tag{2.1}$$

Here, the three term recurrence relation (1.10) was used, $S^T(t) = (S_0(t), S_1(t),...,S_{N-1}(t))$ and $T$ is the tridiagonal matrix with the diagonal elements $t_{nn} = -\beta_{n-1}/\alpha_{n-1}$, $n = 1,...,N$, and the off-diagonal elements $t_{n,n+1} = 1/\alpha_{n-1}$ and $t_{n+1,n} = \gamma_n/\alpha_n$ for $n = 1,...,N-1$. Thus $S_N(t_i) = 0$ holds if and only if $t_i$ is an eigenvalue of the matrix $T$. Further, it can be shown[6,7] that $T$ is symmetric if the polynomials $S_n(t)$ are orthonormal. If $T$ is not symmetric, then a diagonal similarity transformation is to be performed, which yields the orthonormal set of polynomials $\overline{S}(t) = ZS(t)$ and the symmetric tridiagonal matrix $J = ZTZ^{-1}$. Eigenvalues of the matrix $J$ are abscissas of the Gauss rule. Calculating the eigenvectors $\overline{S}(t_i)$, associated with the eigenvalue $t_i$, one can obtain the weights $\lambda_i$ from

$$\lambda_i [\overline{S}(t_i)]^T \overline{S}(t_i) = 1, \quad i = 1,...,N, \tag{2.2}$$

which is a consequence of Christoffel–Darboux identity.[8]

Therefore, the crucial point in generating abscissas and weights is the evaluation of the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$, which form the elements of the matrix $J$. The polynomials can be expressed in terms of the moments

$$R_j = \frac{1}{2}\int_{-1}^{1} \omega(t) t^j\, dt, \quad j = 0,...,2N-1 \tag{2.3}$$

as

$$S_n(t) = \frac{k_n}{B_n^{(n)}} \begin{vmatrix} R_0 & R_1 & \cdots & R_n \\ R_1 & R_2 & \cdots & R_{n+1} \\ \vdots & \vdots & & \vdots \\ R_{n-1} & R_n & & R_{2n-1} \\ 1 & t & & t^n \end{vmatrix}$$

$$= \frac{k_n}{B_n^{(n)}} \sum_{i=0}^{n} B_i^{(n)} t^i, \tag{2.4}$$

where $B_0^{(0)} = 1$,

$$B_n^{(n)} = \mathrm{Det}(B_{ij}) > 0, \quad B_{ij} = R_{i+j-2}$$
$$\text{for } 1 \leqslant i \leqslant n-1, \ 1 \leqslant j \leqslant n-1, \tag{2.5}$$

$k_n \neq 0$ is arbitrary and the remaining coefficients $B_i^{(n)}$ can be inferred from (2.4). It is tempting to express the coefficients $\{\alpha_n, \beta_n, \gamma_n\}$ in terms of the moments (2.3), which can be easily calculated in the case of $\omega(t) = \exp(Dt)$. Such a procedure consists of two steps.[7]

(i) The norm of the polynomials $S_n(t)$ is $h_n = k_n^2 B_{n+1}^{(n+1)}/B_n^{(n)}$ and the three term recurrence relation (1.10) takes the form

$$\sqrt{b_{n+1}}\,\overline{S}_{n+1}(t) = (a_n - a_{n+1} + t)\overline{S}_n(t)$$
$$- \sqrt{b_n}\,\overline{S}_{n-1}(t), \quad n = 0,...,N-1 \tag{2.6}$$

with $\overline{S}_{-1}(t) = 0$ and $\overline{S}_0(t) = 1/\sqrt{R_0}$

for orthonomal polynomials $\overline{S}_n(t) = S_n(t)/\sqrt{h_n}$, where

$$a_n = -B_{n-1}^{(n)}/B_n^{(n)}, \ b_n = B_{n+1}^{(n+1)}B_{n-1}^{(n-1)}/[B_n^{(n)}]^2 > 0 \tag{2.7}$$

and $a_0 = b_0 = 0$.

The matrix $J$, which is to be diagonalized, has the following nonzero elements: $J_{i,i} = a_i - a_{i-1}$ for $i = 1,...,N$ and $J_{i,i+1} = J_{i+1,i} = \sqrt{b_i}$ for $i = 1,...,N-1$.

(ii) The matrix $B = \{B_{ij}\}$, where $B_{ij}$ are defined in (2.5), is symmetric and positive definite. Such a matrix can be decomposed as $B = F^T F$, where $F$ is an upper tridiagonal matrix with elements

$$F_{ik} = \left(R_{i+k-2} - \sum_{j=1}^{i-1} F_{ji} F_{jk}\right)/F_{ii} \quad i\leqslant k,\ i = 1,...,N+1. \tag{2.8}$$

Golub and Welsch have shown[7] that

$$a_i = F_{i,i+1}/F_{i,i} \quad b_i = (F_{i+1,i+1}/F_{i,i})^2. \tag{2.9}$$

The decomposition (2.8) represents a straightforward method for obtaining all the coefficients $a_i$ and $b_i$ necessary for constructing the $J$ matrix.

Unfortunately, in the case of weight $\omega(t) = \exp(Dt)$, the method (ii) gives numerically unstable results[9] for all values $D$ and for as small a degree as $N = 10$. It would be desirable to obtain a recursion among the coefficients $a_i$ and $b_i$, which is more transparent than Eq. (2.8) and does not contain redundant elements $F_{i,j+i}, j = 2,3,...,N+1$. This is the reason why properties of orthogonal polynomials with exponential weight are studied in some detail in the next section.

## III. ORTHOGONAL POLYNOMIALS WITH EXPONENTIAL WEIGHT IN [ − 1,1]

In this section, the weight is specified as

$$\omega(t) = \exp(Dt), \tag{3.0.1}$$

where $D$ is a real parameter. The properties of the polynomials

$$P_n(D,t) = S_n(D,t)/k_n = t^n - a_n(D)t^{n-1} + \cdots \tag{3.0.2}$$

are studied, since most of the expressions obtained have a simpler form for $P_n(D,t)$ than for $S_n(D,t)$ or $\overline{S}_n(D,t)$. Whenever the quantities under consideration depend on the parameter $D$ [e.g., $a_n(D)$ and $b_n(D)$ as defined in (2.7)], it will be shown explicitly in this section.

### III.1 Moments

The following recurrence relations hold for the moments (2.3)

$$R_{2k}(D) = \frac{\sinh(D)}{D} - \frac{2k}{D} R_{2k-1}(D), \quad R_{-1}(D) = 0, \tag{3.1.1}$$

$$R_{2k+1}(D) = \frac{\cosh(D)}{D} - \frac{2k+1}{D} R_{2k}(D), \quad k = 0,1,\ldots .$$

Another obvious relation

$$R_{k+1}(D) = d(R_k(D))/dD \tag{3.1.2}$$

can be obtained from Eq. (2.4).

### III.2 Symmetry properties

It follows from (3.1.1) that $R_{2k}(D) = R_{2k}(-D)$ and $R_{2k+1}(D) = -R_{2k+1}(-D)$. Using Eq. (2.5), we have after simple manipulations

$$B_n^{(n)}(D) = B_n^{(n)}(-D) \tag{3.2.1}$$

and

$$P_n(D,t) = (-1)^n P_n(-D,-t). \tag{3.2.2}$$

This is the reason why we restrict ourselves to nonnegative values $D$ in our further investigation. Particularly,

$$a_n(D) = -a_n(-D) \quad b_n(D) = b_n(-D). \tag{3.2.3}$$

### III.3 Explicit expressions for $P_n(D,t)$ with $n\leqslant 2$

It is instructive to evaluate the lowest degree polynomials $P_n(D,t)$ using moments (2.3) and Eq. (2.4). We have

$$P_0(D,t) = 1, \quad P_1(D,t) = t + 1/D - \coth(D)$$

and

$$P_2(D,t) = t^2 + \frac{2t}{D}\left[1 + \frac{1 - D\coth(D)}{1 + D^2(1-\coth^2(D))}\right] - 1 - \frac{2}{D^2} + \frac{2}{D^2}\frac{D^2 - 2D\coth(D) + 2}{1 + D^2(1-\coth^2(D))}. \tag{3.3.1}$$

With increasing degree $n$, the coefficients of $P_n(D,t)$ contain higher and higher powers of coth($D$) and $D$.

### III.4 Limiting cases

It is evident from (2.3) that in the case $D = 0$ ($\omega(t) = 1$), the polynomials $P_n(D,t)$ go over to the Legendre polynomials $P_n(t)$. We have

$$\lim_{D\to 0} P_n(D,t) = \frac{n!}{(2n-1)!!} P_n(t). \tag{3.4.1}$$

If the values

$$\lim_{D\to 0} R_{2k}(D) = \frac{1}{2k+1},$$

$$\lim_{D\to 0} R_{2k+1}(D) = 0, \quad k = 0,...,2n-2 \tag{3.4.2}$$

are substituted in (2.4), we are left with

$$\lim_{D\to 0} B_n^{(n)}(D) = \prod_{k=0}^{n-1} \frac{1}{2k+1}\left[\frac{k!}{(2k-1)!!}\right]^2. \tag{3.4.3}$$

Using Eqs. (3.4.1) and (3.4.3) we have

$$\lim_{D\to 0} a_n(D) = 0 \quad \lim_{D\to 0} b_n(D) = \frac{n^2}{4n^2-1}. \tag{3.4.4}$$

In investigating the asymptotic region of large $D$, we begin with

$$R_k(D) = \frac{e^D}{2D}\sum_{j=0}^{k}\frac{k!}{(k-j)!}\left(-\frac{1}{D}\right)^j + O(e^{-D}). \tag{3.4.5}$$

After substituting (3.4.5) into Eqs. (2.4) and (2.5), we have for fixed $n$

$$B_n^{(n)}(D) = \frac{\exp(nD)}{2^n D^{n^2}} \prod_{k=0}^{n-1} (k!)^2 + O(e^{(n-2)D}) \tag{3.4.6}$$

and

$$P_n(D,t) = \frac{n!}{D^n} L_n[(1-t)D] + O(e^{-2D}), \tag{3.4.7}$$

respectively. Here,

$$L_n(z) = \sum_{j=0}^{n}\binom{n}{j}\frac{(-z)^j}{j!} \tag{3.4.8}$$

are Laguerre polynomials. Finally, we obtain

$$a_n(D) = n - n^2/D + O(e^{-2D})$$
$$b_n(D) = n^2/D^2 + O(e^{-2D}). \tag{3.4.9}$$

It can be concluded that the quadrature procedure studied here turns out to be the Gauss–Legendre rule for $D = 0$ and asymptotically goes over to the Gauss–Laguerre rule for large $D$.

### III.5 Relations among polynomials and their derivatives

We begin with the observation that the three term recurrence relation (2.6) can be rewritten for the polynomials $P_n(D,t)$ as

$$P_{n+1}(D,t) - (a_n(D) - a_{n+1}(D) + t)P_n(D,t) + b_n(D)P_{n-1}(D,t) = 0 \quad (3.5.1)$$

with $P_{-1}(D,t) = 0$ and $P_0(D,t) = 1$. Except for Eq. (3.5.1), all other relations derived in this subsection reflect the special properties

$$\frac{d\omega(t)}{dt} = D\omega(t) \text{ and } \frac{d\omega(t)}{dD} = t\omega(t) \quad (3.5.2)$$

of our weight function (3.0.1).

It can be seen from Eq. (3.0.2) that $dP_n(D,t)/dD$ is a linear combination of polynomials $P_i(D,t)$ with the highest possible degree $i = n-1$,

$$\frac{dP_n(D,t)}{dD} = \sum_{i=0}^{n-1} \delta_i P_i(D,t). \quad (3.5.3)$$

Constructing now the expressions

$$\frac{d}{dD}(P_n(D,t), P_i(D,t)) = (P_n(D,t), t\,P_i(D,t)) + \left(\frac{d}{dD}P_n(D,t), P_i(D,t)\right) + \left(P_n(D,t), \frac{d}{dD}P_i(D,t)\right) \quad i = 0,\ldots,n-1 \quad (3.5.4)$$

and using

$$\begin{aligned}(P_i(D,t), P_j(D,t)) &= B_{i+1}^{(i+1)}(D)/B_i^{(i)}(D) \quad \text{for } i=j \\ &= 0 \quad \text{for } i \neq j\end{aligned} \quad (3.5.5)$$

we arrive at the conclusion that $\delta_i = 0$ for $i = 0,\ldots,n-2$ and $\delta_{n-1} = -b_n(D)$, so that we are left with the important relation

$$d\,P_n(D,t)/dD + b_n(D)\,P_{n-1}(D,t) = 0. \quad (3.5.6)$$

It follows from the comparison of the coefficients at $t^{n-1}$ in Eq. (3.5.6) that

$$da_n(D)/dD = b_n(D) > 0. \quad (3.5.7)$$

Therefore, $a_n(D)$ is an increasing function of $D$, positive [see Eq. (3.4.1)] for $D > 0$.

Another class of relations involves derivatives $d\,P_n(D,t)/dt$. The relation

$$(t^2-1)\frac{d}{dt}P_n(D,t) = \sum_{i=n-2}^{n+1} \epsilon_i\, P_i(D,t) \quad (3.5.8)$$

holds, since

$$\left(P_i(D,t)(t^2-1)\frac{d}{dt}P_n(D,t)\right) = -\left(P_n(D,t), \frac{1}{\omega(t)}\frac{d}{dt}[(t^2-1)\omega(t)\,P_i(D,t)]\right) = 0 \quad (3.5.9)$$

for $i < n-2$. The coefficients $\epsilon_i$ can be determined from Eqs. (3.5.8), (3.0.2), and the orthogonality relation (3.5.5). Further, using the three term recurrence relation (3.5.1), we have

$$\begin{aligned}(t^2-1)\frac{d\,P_n(D,t)}{dt} &= (nt + a_n(D) + Db_n(D))\,P_n(D,t) \\ &\quad - b_n(D)[2n+1+D(t+a_{n+1}(D) - a_n(D))] \\ &\quad \times P_{n-1}(D,t).\end{aligned} \quad (3.5.10)$$

From similar considerations we also obtain

$$\begin{aligned}\frac{t^2-1}{\omega(t)}\frac{d}{dt}[\omega(t)\,P_{n-1}(D,t)] &= [2n-1+D(t+a_n(D)-a_{n-1}(D))]\,P_n(D,t) \\ &\quad - (nt + a_n(D) + Db_n(D))\,P_{n-1}(D,t).\end{aligned} \quad (3.5.11)$$

The last two equations serve as a starting point in deriving the relations among the coefficients $a_n(D)$ and $b_n(D)$ as well as the differential equation for $P_n(D,t)$.

### III.6 Relations among $a_n(D)$ and $b_n(D)$

At the points $t = 1$ and $t = -1$, Eqs. (3.5.10) and (3.5.11) reach an especially simple form. Denoting

$$X_i(D) = 2i + 1 + D(a_{i+1}(D) - a_i(D)), \quad i = 0,1,\ldots,n. \quad (3.6.1)$$

we have for $t = 1$

$$\begin{aligned}(n + a_n(D) + Db_n(D))\,P_n(D,1) &= b_n(D)(X_n(D)+D)\,P_{n-1}(D,1), \\ (X_{n-1}(D)+D)\,P_n(D,1) &= (n + a_n(D) + Db_n(D))\,P_{n-1}(D,1).\end{aligned} \quad (3.6.2)$$

Since $P_i(D,1) \neq 0$ for $i = 0,\ldots,n$, we have

$$n + a_n(D) + Db_n(D) = \sqrt{b_n(D)}\,\sqrt{X_n(D)+D}\,\sqrt{X_{n-1}(D)+D}. \quad (3.6.3)$$

Here, we used the inequalities $a_n(D) \geqslant 0$ and $b_n(D) > 0$ for $D \geqslant 0$, which have been proved in previous subsections and which imply [see Eqs. (3.6.2)] that $(X_i(D)+D)$ does not change the sign for $D \geqslant 0$. It follows from (3.6.1) and (3.4.4) that $(X_i(D)+D) > 0$ holds for $D \geqslant 0$. Further, $P_i(D,1) > 0$ follows from Eq. (3.6.2).

An analogous derivation can be performed also for $t = -1$. Taking into account that $P_i(D,-1) = (-1)^i|\,P_i(D,-1)| \neq 0$, $i = 0,\ldots,n$, and $X_0(D) - D = D\coth(D) > 0$ [see Eq. (3.3.1)], we obtain

$$n - a_n(D) - Db_n(D) = \sqrt{b_n(D)}\sqrt{X_n(D)-D}\,\sqrt{X_{n-1}(D)-D}, \quad (3.6.4)$$

where

$$n > a_n(D) > n - n^2/D, \quad b_n(D) < (n/D)^2, \text{ and } X_n(D) > D \text{ for all } D > 0. \quad (3.6.5)$$

The system of two equations (3.6.3) and (3.6.4) can be treated as a recursion for $a_n(D)$, $n = 0,1,\ldots,N$, and $b_n(D)$, $n = 1,\ldots,N-1$, with starting values $a_0(D) = 0$ and $a_1(D)$

$= \coth(D) - 1/D$. Technical aspects associated with the evaluation of $a_n(D)$ and $b_n(D)$ will be discussed in the next section. Here, we give two alternative formulations of the recursion, which are useful in practical applications.

It is easy to verify that the system (3.6.3) and (3.6.4) is equivalent to the following one:

$$\sqrt{b_n(D)} = 2n/\left[\sqrt{X_n(D)+D}\,\sqrt{X_{n-1}(D)+D} + \sqrt{X_n(D)-D}\,\sqrt{X_{n-1}(D)-D}\right], \quad (3.6.6)$$

$$2n\,a_n(D) = b_n(D)D^2(a_{n+1}(D) - a_{n-1}(D) + 2n/D). \quad (3.6.7)$$

The second formulation is obtained when one introduces $\alpha_i(D) > 0, i = 1,...,N+1$, by $X_i(D)/D = \cosh(2\alpha_i(D))$ and rewrites Eqs. (3.6.6) and (3.6.7) as

$$a_n(D) = n\frac{\cosh(\alpha_n(D) - \alpha_{n-1}(D))}{\cosh(\alpha_n(D) + \alpha_{n-1}(D))} - \frac{n^2}{D}\frac{1}{\cosh^2(\alpha_n(D) + \alpha_{n-1}(D))}. \quad (3.6.8)$$

Using an analogous expression for $a_{n+1}(D)$, one obtains after simple manipulations a very instructive recursion

$$\sinh(2\alpha_n(D)) = (n/D)\tanh(\alpha_n(D) + \alpha_{n-1}(D)) + ((n+1)/D)\tanh(\alpha_{n+1}(D) + \alpha_n(D)) \quad (3.6.9)$$

for $\alpha_n(D)$ with starting values $\alpha_{-1}(D) = 0$ and

$$\alpha_0(D) = -\tfrac{1}{2}\ln(\tanh D/2). \quad (3.6.10)$$

Taking into account Eq. (3.5.7), we can obtain from (3.6.3) and (3.6.4) also

$$2\frac{d}{dD}\alpha_n(D) = \frac{n}{D}\tanh(\alpha_n(D) + \alpha_{n-1}(D)) - \frac{n+1}{D}\tanh(\alpha_{n+1}(D) + \alpha_n(D)). \quad (3.6.11)$$

Equations (3.6.6)–(3.6.11) provide a solid basis for numerical evaluation of $a_n(D)$ and $b_n(D)$.

## IV. ALGORITHM FOR EVALUATING $a_n(D)$ AND $b_n(D)$

The recursions derived in the previous section are not very transparent and are to be investigated in some detail before using them for computation of $a_n(D)$ and $b_n(D)$. Our objective is to generate the sequences $a_n(D)$ and $b_n(D)$, $n = 1,...,N+1$, for fixed value $D$.

Let us start with the case of large $D$. It is advantageous to rewrite Eqs. (3.6.6) and (3.6.7) as

$$2D^2b_n = D^2 - X_{n-1}^2 + 2nX_{n-1} - 2a_nD - (X_{n-1}^2 - D)^{1/2} \times [(X_{n-1} - 2n)^2 + 4Da_n - D^2]^{1/2}, \quad (4.1.1)$$

$$a_{n+1} = a_{n-1} + 2na_n/D^2b_n - 2n/D, \quad n = 1,...,N+1$$

with $a_0 = 0$ and $a_1 = \coth(D) - 1/D$. In what follows, the dependence of $a_n$, $b_n$, and $X_n$ on $D$ is suppressed to simplify the notation. The expression for $X_n$ is given by Eq. (3.6.1). It can easily be verified that the asymptotic expression (3.4.9) for $a_n$ and $b_n$ provide an exact solution to (4.1.1) for any $D > 0$. This causes some difficulty when $a_n$ and $b_n$ are evaluated numerically, since the starting value $a_1 = \coth(D) - 1/D$ is very close to $1 - 1/D$ for $D \gg 1$ and the asymptotic rather than desired solution is generated by (4.1.1) for $n > 1$.

The problem can be solved by introducing

$$a_n = n - n^2/D + 2g_n \quad \text{and} \quad b_n = (n/D)^2 - 2d_n \quad (4.1.2)$$

and rewriting Eqs. (4.1.1) in the form

$$d_n = (g_n - g_{n-1})\left(1 - \frac{n}{D}\right) + \frac{g_n}{D} + (g_n - g_{n-1})^2 + [(g_n - g_{n-1})^2 + (g_n - g_{n-1})]^{1/2} \times \left[(g_n - g_{n-1})^2 + (g_n - g_{n-1})\left(1 - \frac{2n}{D}\right) + \frac{2g_n}{D}\right]^{1/2} \quad (4.1.3)$$

$$g_{n+1} = g_{n-1} + 2[ng_n + D(D-n)d_n]/(n^2 - 2D^2d_n), \quad n = 1,...,N+1,$$

with $g_0 = 0$ and $g_1 = \exp(-2D)$.

It should be noted that $g_{n+1} > g_n > 0$ and $d_n > 0$ follows from Eqs. (3.6.5) for $D > 0$. Therefore, no cancellation occurs in (4.1.3) if $d_n$ and $g_{n+1}$ are calculated for $n < \min(N+1, D/2)$. The error in determining $d_n$ and $g_{n+1}$ is not larger than approximately $10^{-\delta}$, where $\delta$ is the number of digits carried in the calculation. We have verified by computer calculation that still for $n < \min(N+1, D)$ only two or three decimal digits are lost if $N \leqslant 40$, which is quite acceptable for practical purposes. On the contrary, the recursion (4.1.3) quickly breaks down for $n > D$ due to enormous cancellation, which occurs especially in the expression for $d_n$.

To complete the algorithm for evaluating $a_n$ and $b_n$ we need a method which works in the interval $N+1 \geqslant n \geqslant D$. In this "small $D$" region, we encounter the following difficulty. Let us represent $a_n$ and $a_{n-1}$ as

$$a_n = D\sum_{i=0}^{I} c_i^{(n)}D^{2i} + O(D^{2I+2}),$$

$$a_{n-1} = D\sum_{i=0}^{I} c_i^{(n-1)}D^{2i} + O(D^{2I+2}). \quad (4.1.4)$$

This can be always done in a disk on the complex $D$-plane with the center at $D = 0$ and with a finite diameter, since $a_n$ is an analytic function of $D$ in the vicinity of the origin [see Eqs. (2.4), (2.9), and (3.4.4)]. Now we evaluate $a_{n+1}$ using Eqs. (4.1.1). The error of this quantity will be of the order of $O(D^{2I})$—larger than the error of the input values. We can conclude that in the "small $D$" region the errors are accumulated when we move in the recursion (4.1.1) from small to large values $n$.

Unfortunately, the same is true when we move in (4.1.1) from large values $n$ towards small ones. This property of the recursion remains unchanged also in the other formulations derived in the previous section. This is the reason why we prefer to solve the recursion in the $D < n$ region by iterations. The method is based on the following theorem.

**Theorem:** Let us consider a set $S$ of sequences $\{\alpha_n^{(i)}\}$, where $\alpha_{-1}^{(i)} = 0$ for $i = 1,2,...,\alpha_n^{(1)}$ are arbitrary real numbers such that $\alpha_n^{(1)} \geqslant 0$ holds for $n = 0,1,...$, and $\alpha_n^{(1)} > 0$ holds at least for one $n$. Finally, the elements $\alpha_n^{(i+1)}$ are defined by

$$\sinh(2\alpha_n^{(i+1)}) = \frac{n+1}{D}\tanh(\alpha_{n+1}^{(i)} + \alpha_n^{(i)}) + \frac{n}{D}\tanh(\alpha_n^{(i)} + \alpha_{n-1}^{(i)}) \quad (4.1.5)$$

for $i = 1,2,...;\ n = 0,1,...$

Then the limits $0 \leqslant \lim_{i\to\infty} \alpha_n^{(i)} = \alpha_n < \infty$ exist for all $D > 0$, $n = 0,1,...,$ and define uniquely the sequence $\{\alpha_n\}$, which satisfies the recursion (3.6.9). There exists only one sequence $\{a_n\}$, $\alpha_{-1} = 0$ and $\alpha_n > 0$, $n = 0,1,...,$ that satisfies (3.6.9) and its starting value is $\alpha_0 = -0.5\ln(\tanh D/2)$.

*Proof*: We start with several simple observations. For a sequence $\{\alpha_n\}$, $\alpha_n > 0$, which satisfies Eq. (3.6.9), the inequalities

$$\sinh(2\alpha_n) > \frac{2n+1}{D}\tanh(\alpha_n) \text{ or } \cosh^2(\alpha_n) > \frac{2n+1}{2D} \quad (4.1.6)$$

hold for $n = 0,1,...$ Further, we have

$$\lim_{n\to\infty} \alpha_n = \infty, \quad \lim_{n\to\infty}\left[\frac{1}{n}\sinh(2\alpha_n)\right] = \frac{2}{D} \quad (D > 0) \quad (4.1.7)$$

and

$$\lim_{n\to\infty}\{n^2[\tanh(\alpha_n + \alpha_{n-1}) - 1]\} = \frac{D^2}{8}.$$

Consider now a sequence $\{\alpha_n^{(i)}\} \in S$. There exist two sequences $\{\overline{\alpha}_n^{(i)}\} \in S$ and $\{\overline{\alpha}_n^{(i)}\} \in S$ with the following properties:

(i) Let $\overline{\alpha}_n^{(1)} = \infty$, then $\overline{\alpha}_n^{(i)} > \overline{\alpha}_n^{(i+1)}$ and $\overline{\alpha}_n^{(i)} > \alpha_n^{(i)}$ hold for $n = 0,1,...;\ i = 1,2,...$ .

(ii) Let $\alpha_n^{(1)} > 0$ be the first nonzero element from $\alpha_n^{(1)}$, $n = 0,1,...$ .

We define $\overline{\alpha}_k^{(1)} = \alpha_k^{(1)}$ if $(2k+1) < 2D$ and $\alpha_k^{(1)} = \min(\alpha_k^{(1)}$, $\operatorname{arcosh}[((2k+1)/2D)^{1/2}]$ otherwise. Further, we put $\overline{\alpha}_n^{(1)} = 0$ for $n \neq k$. Then $\overline{\alpha}_n^{(i)} \leqslant \overline{\alpha}_n^{(i+1)}$ [see Eq. (4.1.6)] and $\overline{\alpha}_n^{(i)} \leqslant \alpha_n^{(i)}$ hold for $n = 0,1,...;\ i = 1,2,...$ . The limits $\lim_{i\to\infty}\overline{\alpha}_n^{(i)} = \overline{\alpha}_n$ and $\lim_{i\to\infty}\overline{\alpha}_n^{(i)} = \overline{\alpha}_n$ obviously exist, $0 \leqslant \overline{\alpha}_n \leqslant \overline{\alpha}_n$ holds, and $\overline{\alpha}_n$ and $\overline{\alpha}_n$ satisfy (3.6.9).

Further, it can be shown from (3.6.1) and (3.6.8) that for any sequence $\{\alpha_n\}$, $\alpha_n > 0$ which satisfies (3.6.9),

$$\sum_{i=0}^{n-1}\cosh(2\alpha_i) = \frac{n^2}{D}\tanh^2(\alpha_n + \alpha_{n-1}) + n\frac{\cosh(\alpha_n - \alpha_{n-1})}{\cosh(\alpha_n + \alpha_{n-1})} = \frac{n^2}{D} + \frac{D}{4} + \eta_n \quad (4.1.8)$$

holds, where $\lim_{n\to\infty}\eta_n = 0$. In deriving the last relation, Eqs. (3.6.1), (3.6.8), and (4.1.7) were used. Therefore,

$$\lim_{n\to\infty}\sum_{i=0}^{n}[\cosh(2\overline{\alpha}_i) - \cosh(2\overline{\alpha}_i)] = 0 \quad (4.1.9)$$

and $\{\overline{\alpha}_n\} = \{\overline{\alpha}_n\}$. It means that $\lim_{i\to\infty}\alpha_n^{(i)}$ exist and define uniquely the sequence $\{\alpha_n\} = \{\overline{\alpha}_n\} = \{\overline{\alpha}_n\}$, which satisfies (3.6.9).

Let us have a sequence $\{\beta_n\}$, $\beta_{-1} = 0$, and $\beta_n > 0$, $n = 0,1,...,$ that satisfies Eq. (3.6.9). Since $\lim_{i\to\infty}\alpha_n^{(i)} = \alpha_n$, $n = 0,1,...,$ hold for $\{\alpha_n^{(i)}\} \in S$ when $\alpha_n^{(1)} > 0$ is chosen arbitrarily, the same must be true when $\alpha_n^{(1)} = \beta_n$, $n = 0,1,...$ .

Therefore, we have $\{\alpha_n\} = \{\beta_n\}$ and the sequence $\{\alpha_n\}$ is the only one with $\alpha_{-1} = 0$ and $\alpha_n > 0$, $n = 0,1,...,$ that satisfies (3.6.9). The sequence with such properties was obtained already in Sec. III.6 and its starting value is $\alpha_0 = -0.5\ln(\tanh D/2)$.

For practical purposes, the iterations (4.1.5) can be reformulated in terms of $a_n^{(i)}$, which are defined by

$$X_n^{(i)}/D = \frac{2n+1}{D} + a_{n+1}^{(i)} - a_n^{(i)} = \cosh(2\alpha_n^{(i)}), \quad a_0^{(i)} = 0 \quad (4.1.10)$$

for $n = 0,1,...;\quad i = 1,2,...$ .

The resulting expression is

$$a_{n+1}^{(i+1)} - a_n^{(i+1)} = D\frac{1 - 2(2n+1)A_n^{(i)} + (DA_n^{(i)})^2}{2n+1+\sqrt{D^2 + (2n+1-D^2A_n^{(i)})^2}}, \quad (4.1.11)$$

where

$$\frac{A_n^{(i)}}{2} = \frac{n+1}{D^2 + [D\exp(\alpha_{n+1}^{(i)} + \alpha_n^{(i)})]^2} + \frac{n}{D^2 + [D\exp(\alpha_n^{(i)} + \alpha_{n-1}^{(i)})]^2}$$

and $D\exp(2\alpha_n^{(i)}) = 0.5(\sqrt{X_n^{(i)} + D} + \sqrt{X_n^{(i)} - D})^2$ for $n = 0,1,...$ As opposed to (4.1.5), the iterations (4.1.11) yield a finite result also for $D = 0$, and the expressions $\lim_{i\to\infty}(a_{n+1}^{(i)} - a_n^{(i)}) = (a_{n+1} - a_n)$, the existence and uniqueness of which is guaranteed by the theorem, enter directly the $J$ matrix, which was defined in Sec. II.

In concluding this section, we would like to summarize the algorithm for obtaining abscissas and weights.

(i) For $n \leqslant N_0 = \min([D], N+1)$, the coefficients $a_n$ and $b_n$, $n = 0,1,...,N_0$, are evaluated according to Eqs. (4.1.2)–(4.1.3).

(ii) If $N + 1 > [D]$, the coefficients $a_n$, $n = N_0 + 1,...,N+1$, are obtained using the iterative procedure (4.1.11). In such a case, we put $X_n^{(i)} = 2n + 1 + D(a_{n+1} - a_n)$ for all $i = 0,1,...,$ and $n = 0,1,...,N_0 - 1$, where $a_n$ are those as obtained in (i). Further, the starting values $X_n^{(1)} = ((2n+1)^2 + D^2)^{1/2}$ are chosen for $N_0 \leqslant n \leqslant N + 10$ and $X_n^{(1)} = 0$ for $n > N + 10$. The rate of convergence of (4.1.11) was checked for all

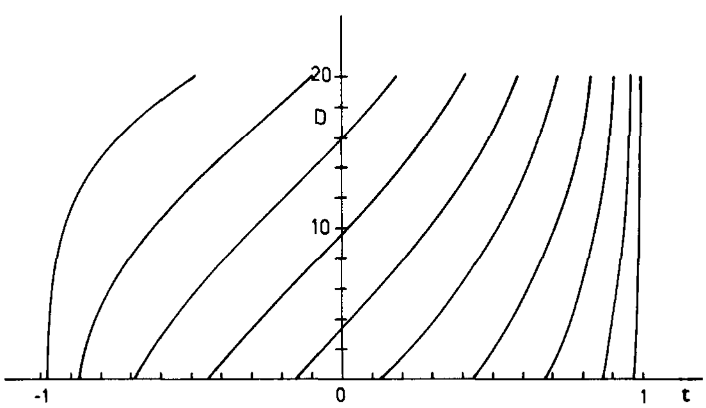


FIG. 1. Roots of the polynomial $P_{10}(D,t)$.

TABLE I. Results for moments $R_D(D)$. Underlined figures are those which disagree with the exact result.

| | | $N=2$ | $N=4$ | $N=8$ | $N=16$ |
|---|---|---|---|---|---|
| $R_2(2)$ | $L$ | 0.581370827(00)[a] | 0.838444018(00) | 0.839047460(00) | 0.839047460(00) |
| | $P$ | 0.839047460(00) | 0.839047460(00) | 0.839047460(00) | 0.839047460(00) |
| $R_4(4)$ | $L$ | 0.564872965(00) | 0.301756323(01) | 0.319174000(01) | 0.319174117(01) |
| | $P$ | 0.310853243(01) | 0.319174117(01) | 0.319174117(01) | 0.319174117(01) |
| $R_8(8)$ | $L$ | 0.625817255(00) | 0.516232310(02) | 0.900447566(02) | 0.901570490(02) |
| | $P$ | 0.868697594(02) | 0.901516209(02) | 0.901570490(02) | 0.901570490(02) |
| $R_{16}(16)$ | $L$ | 0.783142057(00) | 0.153216831(05) | 0.125596652(06) | 0.136642696(06) |
| | $P$ | 0.131375034(06) | 0.136595213(06) | 0.136642762(06) | 0.136642762(06) |
| $R_{32}(32)$ | $L$ | 0.122662294(01) | 0.134972374(10) | 0.306701218(12) | 0.610801305(12) |
| | $P$ | 0.587945344(12) | 0.611726995(12) | 0.612041268(12) | 0.612041282(12) |

[a] $(02)=10^2$ etc.

$N_0 \leqslant n \leqslant N+1$. In fact, not more than four or five iterations were needed in order to achieve the results accurate up to ten decimal digits for $N \leqslant 40$. Finally, the coefficients $b_n$ are obtained for $N_0 \leqslant n \leqslant N+1$ from Eq. (3.6.6).

(iii) The matrix $J$ is constructed and diagonalized. The eigenvalues represent abscissas of Gauss rule and the weights are deduced from corresponding eigenvectors.

## V. APPLICATIONS

Now we apply the quadrature rule to several test integrals. Our aim is to examine the rate of convergence of the method as $N$ (the number of abscissas) increases. A comparison is made with the convergence rate of Gauss–Legendre rule. The weights and abscissas needed were generated using the algorithm given at the end of the preceding section. Double-precision arithmetic (15 decimal digits) were used throughout. The dependence of abscissas on the parameter $D$ is demonstrated in Fig. 1, where all roots of the polynomial $P_{10}(D,t)$ are shown.

In Table I we present the results obtained for the moments

$$R_D(D)=\frac{1}{2}\int_{-1}^{1}\exp(Dt)t^D\,dt, \tag{5.1}$$

$D=2,4,8$, and 16, using the Gauss quadrature formulas (1.5)–(1.6) with $N=2,4,8,16$, and 32 and with $\omega(t)=1$ and $\omega(t)=\exp(Dt)$, respectively. It can be seen that the Gauss–Legendre quadrature ($\omega(t)=1$) converges much more slowly than the quadrature associated with the polynomials $P_n(D,t)$ (henceforth referred to as Gauss-$P$ quadrature). Corresponding results are denoted in Table I as $L$ and $P$, respectively. Further, the Gauss-$P$ quadrature yields results accurate up to ten decimal digits for $D \leqslant 2N-1$. This is a useful check on the consistency of the abscissas and weights.

Table I demonstrates in the same time how useful is the Gauss-$P$ quadrature in evaluating the optical potentials. The classical Kisslinger potential[10] or the "potential with the Laplacian"[11] for pion-nucleus scattering are in fact linear combinations of the moments $R_N(D)$, where $N \leqslant 10 \div 20$.

To test the rate of convergence of our method, we must choose integrands more complicated than Eq. (5.1). The results obtained for

TABLE II. Results for $I_1(D)$. Underlined figures are those which disagree with the exact result.

| | | $N=2$ | $N=4$ | $N=8$ | $N=16$ |
|---|---|---|---|---|---|
| $I_1(2)$ | $L$ | − 0.227157369( 0) | − 0.188794514( 0) | − 0.188646335( 0) | − 0.188644849( 0) |
| | $P$ | − 0.188711893( 0) | − 0.188684568( 0) | − 0.188647342( 0) | − 0.188644883( 0) |
| $I_1(4)$ | $L$ | − 0.144553191( 0) | − 0.111718436( 0) | − 0.109380274( 0) | − 0.109380244( 0) |
| | $P$ | − 0.109296871( 0) | − 0.109381250( 0) | − 0.109380312( 0) | − 0.109380245( 0) |
| $I_1(8)$ | $L$ | − 0.465429405( − 1)[a] | − 0.704524333( − 1) | − 0.585985648( − 1) | − 0.585937504( − 1) |
| | $P$ | − 0.585874532( − 1) | − 0.585937479( − 1) | − 0.585937505( − 1) | − 0.585937504( − 1) |
| $I_1(16)$ | $L$ | − 0.294881876( − 2) | − 0.401779288( − 1) | − 0.307437964( − 1) | − 0.302734375( − 1) |
| | $P$ | − 0.302730848( − 1) | − 0.302734375( − 1) | − 0.302734375( − 1) | − 0.302734375( − 1) |

[a] $(-1)=10^{-1}$ etc.

TABLE III. Results for $I_2(D)$. Underlined figures are those which disagree with the exact result.

| | | $N=16$ | $N=32$ | $N=40$ | exact |
|---|---|---|---|---|---|
| $I_2(2)$ | L | 0.314300451(01)[a] | 0.314177425(01) | 0.314168630(01) | 0.314159265(01) |
| | P | 0.314278144(01) | 0.314175900(01) | 0.314167992(01) | |
| $I_2(4)$ | L | 0.630446957(01) | 0.628584626(01) | 0.628455291(01) | 0.628318531(01) |
| | P | 0.629808233(01) | 0.628540596(01) | 0.628436775(01) | |
| $I_2(8)$ | L | 0.152239169(02) | 0.128660592(02) | 0.127186574(02) | 0.125663706(02) |
| | P | 0.138511482(02) | 0.127758143(02) | 0.126805656(02) | |
| $I_2(16)$ | L | 0.315315443(05) | 0.203651013(04) | 0.100273150(04) | 0.251327412(02) |
| | P | 0.542248404(04) | 0.100228118(04) | 0.576163054(03) | |

[a] $(01) = 10^1$ etc.

$$I_1(D) = \frac{e^{-2D}}{4}\int_0^2 y(Dy+4)\ln\left(\frac{y}{2}\right)e^{Dy}\,dy = \frac{1}{4D^2}(1-2D-e^{-2D}) \qquad (5.2)$$

are displayed in Table II. The Gauss-$P$ quadrature gives again much better results than the Gauss–Legendre one especially for $D=16$ and 32. The convergence is rather slow for smaller $D$ even using the Gauss-$P$ quadrature. Here, the exponential does not dominate and the integrand exhibits nonpolynomial behavior.

Typical corrections to the optical potentials (e.g., the nonlocal $\Delta_{33}$-propagation or relativistic corrections) have also monotone or slowly oscillating nonpolynomial behavior and Table II provides us with some idea about the efficiency of the Gauss-$L$ and -$P$ quadratures in such cases.

Finally, the limitations of our method are demonstrated in Table III, where the results are shown as obtained for the integral

$$I_2(D) = \int_{-1}^{1} e^{Dt}\sin(D\sqrt{1-t^2})dt = \frac{\pi D}{2}. \qquad (5.3)$$

Although the Gauss-$P$ quadrature works somewhat better than the Gauss–Legendre one, the convergence is poor in both cases especially for large $D$. The reason is that the integrand contains a rapidly oscillating function, the behavior of which is substantially nonpolynomial.

## VI. SUMMARY

The quadrature procedure was developed for integrals with finite integration range that contain the weight function $\exp(Dt)$. The procedure is based on the Gauss rule, the abscissas being determined as roots of orthogonal polynomials with exponential weight in a finite interval. Properties of the polynomials were studied in some detail. A recursion was found for the coefficients of the three term recurrence relation which holds among the orthogonal polynomials. The recursion can be solved by iterations without accumulating roundoff errors, therefore the abscissas and weights are obtained (with the help of the matrix diagonalization technique) with high precision. The rate of convergence of our quadrature procedure is very rapid for integrands that contain a polynomial-like function in addition to the exponential. Such integrals are encountered in various physical applications, e.g., in constructing the optical model.

## ACKNOWLEDGMENT

It is a pleasure to thank Dr. M. Znojil for his advice and helpful comments.

[1]A. W. Thomas and R. H. Landau, Phys. Rep. **58**, 123 (1980).
[2]R. Mach, Nucl. Phys. A **205**, 56 (1973); A **258**, 513 (1976).
[3]D. J. Ernst and G. A. Miller, Phys. Rev. C **12**, 1962 (1975).
[4]P. J. Davis and P. Rabinowitz, *Numerical Integration* (Blaisdell, Waltham, Mass., 1967).
[5]I. H. Sloan, J. Math. Phys. **21**, 1032 (1980), and references therein.
[6]H. Wilf, *Mathematics and the Physical Sciences* (Wiley, New York, 1962).
[7]G. H. Golub and J. H. Welsch, Math. Comp. **23**, 221 (1969).
[8]In fact only the first components of the orthonormal eigenvectors of the $J$ matrix are needed in evaluating the weights (see Refs. 6 and 7).
[9]The observation is made in the original paper (Ref. 7) that the method (ii) yields numerically unstable results for the coefficients $a_i$ and $b_i$ in the case of many orthogonal polynomials.
[10]L. S. Kisslinger, Phys. Rev. **98**, 761 (1955).
[11]H. K. Lee and H. McManus, Nucl. Phys. A **167**, 257 (1971).

# Two-dimensional time-dependent Hamiltonian systems with an exact invariant

B. Grammaticos and B. Dorizzi
*Département de Mathématiques Appliquées (MTI), Centre National d'Etudes des Télécommunications, 92131 Issy les Moulineaux, France*

(Received 21 July 1983; accepted for publication 3 February 1984)

We present a direct approach to investigate the existence of an exact invariant for two-dimensional Hamiltonians, in which the potential depends explicitly on time. The method is based on an expansion of the invariant in the velocities. The problem is solved completely for invariants linear and quadratic in the momenta. Our results contain as a particular case the results of Lewis and Leach on one-dimensional systems.

PACS numbers: 03.20. + i, 02.30. + g

## I. INTRODUCTION

The theoretical description of nonstationary physical phenomena often leads to time-dependent Hamiltonians. An example of historical importance is the description of the motion of a charged particle moving in an electromagnetic field. The Hamiltonian of the system can, in some cases at least, be reduced to the Hamiltonian of a harmonic oscillator, the frequency of which depends on time: $H = \frac{1}{2}[\dot{x} + \omega^2(t)x^2]$. Lewis[1] has shown that an exact invariant, i.e., a conserved quantity, can be constructed for this problem:

$$C = \tfrac{1}{2}[x^2/\rho^2 + (\rho\dot{x} - \dot{\rho}x)^2],$$

in terms of an auxiliary function $\rho(t)$ which is the solution of the equation $\ddot{\rho} + \omega^2(t)\rho = 1/\rho^3$. The derivation of the invariant can be traced back to Ermakov[2] who derived it in 1880. Gambier,[3] in 1910, has also analyzed the equation for $\rho$, or rather for $\Psi = \rho^2$, from the point of view of the Painlevé property. He has integrated it by reducing it to a linear equation, which is exactly the equation for the harmonic oscillator $\ddot{x} + \omega^2(t)x = 0$, and obtained the invariant in the course of his analysis. The importance of the result of Lewis stems from the fact that he used the invariant in order to construct the solution of the quantum time-dependent oscillator,[4] thus reducing the solution of a PDE (the Schrödinger equation) to the solution of an ODE (the equation for $\rho$).

The interest in time-dependent systems has increased appreciably these last years. Several methods have been devised for the derivation of the Lewis invariant, which was originally obtained through an application of the asymptotic theory of Kruskal[5] in closed form: Leach[6] has obtained the same result using a time-dependent canonical transformation. Lutzky's[7] derivation was based on Noether's theorem. Ray and Reid[8] have resurrected the old Ermakov technique, and were able to obtain the existence of a Lewis-type invariant for the case of two coupled nonlinear equations of motion:

$$\ddot{x} + \omega^2(t)x = (1/x^2\rho)g(\rho/x),$$

$$\ddot{\rho} + \omega^2(t)\rho = (1/\rho^2 x)f(x/\rho),$$

namely

$$C = \frac{1}{2}(x\dot{\rho} - \rho\dot{x})^2 + \int^{x/\rho} f(\eta)d\eta + \int^{\rho/x} g(\eta)d\eta.$$

In a series of papers, Ray, Reid, and Lutzky[9–16] have extended further the class of nonlinear equations which possess an exact invariant. They have shown how the same results can be reached using Noether's theorem and demonstrated that there exists a general, nonlinear superposition law for the systems they studied.

A particularly simple analysis, which provides an insight into the results of Ray, Reid, and Lutzky, has been given by Sarlet[17] who has related the existence of the invariant to the integrability of a certain differential one-form. In a more recent paper, Sarlet and Ray[18] have provided a classification scheme for Ermakov-type differential systems, thus establishing some unity into the multitude of examples of time-dependent systems with an exact invariant treated in the literature.

In the Ermakov methodology, one derives the invariant starting from a set of given equations, i.e., the auxiliary equation must be known in advance. However, when one starts from an explicitly time-dependent equation of motion, there is no simple way to guess even the existence of such an auxiliary equation, let alone its form. Because of this and the fact that the method of symmetry transformations, based for example on Noether's theorem, can be roundabout, Lewis and Leach[19] have presented a direct approach for the determination of the invariant of the system with a Hamiltonian of the form $H = \frac{1}{2}p^2 + V(x,t)$.

The extension of the above results to several spatial dimensions presents, of course, a great interest. Some results exist in this direction, although not as ample as in the case of one dimension due to complexity of the problem. Günther and Leach[20] have derived a tensor invariant for an $N$-dimensional time-dependent isotropic harmonic oscillator. Ray and Reid[21] as well as Lutzky[9] have given a brief discussion concerning the extension of their method to several spatial dimensions. In a more recent work, Sarlet and Cantrijn[22] have presented a generalization of this method which, in principle, deals with systems of $n + 1$ second-order differential equations with $n$ first integrals quadratic in the velocities. As in the case of the Ermakov systems, one of the equations plays the role of the auxiliary equation.

In the present work, we will present a study of two-dimensional time-dependent Hamiltonian systems from the point of view of the existence of an exact invariant. The method used is a natural extension of our previous work on

 0022-2488/84/072194-06$02.50  

completely integrable (time-independent) Hamiltonian systems in two dimensions.[23] However, the explicit time dependence of the potential will modify the calculations appreciably. We will use a direct approach for the construction of an invariant polynomial in the velocities of degree 1 or 2. From this respect, our work constitutes an extension in two dimensions of the work of Lewis and Leach on one-dimensional time-dependent Hamiltonians.

In the second and third sections of this paper, we present the construction of linear and quadratic invariants, respectively. In the fourth section, a comparison with previous results is presented, together with our conclusion.

## II. CONSTANTS LINEAR IN THE VELOCITIES

We will consider a Hamiltonian of the form

$$H = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2) + V(x,y,t). \tag{1}$$

The equations of motion associated to this system are simply

$$\ddot{x} = -V_x, \quad \ddot{y} = -V_y, \tag{2}$$

and we can notice that, as the potential $V$ depends explicitly on the time $t$, $H$ is not a constant of the motion. We will first concentrate on the search of an invariant linear in the velocities. It has the general form

$$C = g^0\dot{x} + g^1\dot{y} + h, \tag{3}$$

where $g^0$, $g^1$, and $h$ are functions of $x$, $y$, and $t$.

The condition $dC/dt = 0$ leads to the following polynomial identity in terms of $\dot{x}$ and $\dot{y}$:

$$g^0_x\dot{x}^2 + g^1_y\dot{y}^2 + (g^0_y + g^1_x)\dot{x}\dot{y} + (g^0_t + h_x)\dot{x} + (g^1_t + h_y)\dot{y} + h_t + g^0\ddot{x} + g^1\ddot{y} = 0. \tag{4}$$

This is equivalent to equating to zero the coefficient of each distinct monomial in $\dot{x}$ and $\dot{y}$ and leads to

$$g^0_x = 0, \quad g^0_y + g^1_x = 0, \quad g^1_y = 0, \tag{5}$$

$$g^0_t + h_x = 0, \quad g^1_t + h_y = 0, \tag{6}$$

$$h_t + g^0\ddot{x} + g^1\ddot{y} = 0. \tag{7}$$

The integration of the system (5) is straightforward and reads

$$g^0 = \alpha(t)y + \beta(t), \quad g^1 = -\alpha(t)x + \gamma(t). \tag{8}$$

The system (6) leads to a compatibility condition (9) which ensures the existence of the function $h$:

$$g^0_{t,y} = g^1_{t,x}. \tag{9}$$

In terms of $\alpha, \beta, \gamma$ we get:

$$2\alpha'(t) = 0.$$

Thus, since $\alpha$ is time-independent, we can easily integrate Eqs. (6) and obtain for $h$:

$$h = -x\beta'(t) - y\gamma'(t) + \epsilon(t). \tag{10}$$

The last relation (7) reads, in terms of $\alpha, \beta, \gamma, \epsilon$,

$$x\beta''(t) + y\gamma''(t) - \epsilon'(t) + (\alpha y + \beta)V_x - (\alpha x - \gamma)V_y = 0. \tag{11}$$

Equation (11) is the linear PDE that the potential $V$ must satisfy for the system to possess an invariant linear in the velocities.

We will distinguish two cases.

(a) $\alpha = 0$. Equation (11) reduces to

$$\beta V_x + \gamma V_y + \beta''x + \gamma''y - \epsilon' = 0,$$

or, equivalently,

$$V_\xi + V_\eta + \beta''\beta\xi + \gamma''\gamma\eta - \epsilon' = 0, \tag{12}$$

with $\xi = x/\beta$, $\eta = y/\gamma$.

The integration of the homogeneous equation $V_\xi + V_\eta = 0$ is straightforward and reads

$$V = F(\xi - \eta, t). \tag{13}$$

We need now a particular solution of Eq. (12). For this, we introduce the variables

$$u = \xi - \eta, \quad v = \eta + \xi,$$

which lead to the following form of Eq. (12):

$$2V_v - Bv - \Gamma u - \epsilon' = 0, \tag{14}$$

with

$$\Gamma = -\tfrac{1}{2}(\beta''\beta - \gamma''\gamma), \quad B = -\tfrac{1}{2}(\beta''\beta + \gamma''\gamma).$$

It is immediate to check that $V = \frac{1}{2}(\Gamma uv + \frac{1}{2}Bv^2 + \epsilon' v)$ is a solution of (14), and thus the general solution of Eq. (12) reads

$$V = F(\xi - \eta, t) + \tfrac{1}{2}[(\Gamma + \tfrac{1}{2}B)\xi^2 - (\Gamma - \tfrac{1}{2}B)\eta^2 + B\xi\eta + \epsilon'\xi + \epsilon'\eta]. \tag{15}$$

Let us now examine the second case.

(b) $\alpha \neq 0$. In terms of the variables

$$\xi = x - \gamma/\alpha, \quad \eta = y + \beta/\alpha,$$

Eq. (11) becomes

$$\eta V_\xi - \xi V_\eta + \frac{\beta''}{\alpha}\xi + \frac{\gamma''}{\alpha}\eta + \frac{\beta''\gamma}{\alpha^2} - \frac{\beta\gamma''}{\alpha^2} - \frac{\epsilon'}{\alpha} = 0. \tag{16}$$

Transforming into polar coordinates, $\xi = \rho\cos\varphi$, $\eta = \rho\sin\varphi$, it takes the simpler form

$$V_\varphi = \frac{\beta''}{\alpha}\rho\cos\varphi + \frac{\gamma''}{\alpha}\rho\sin\varphi + \frac{\beta''\gamma - \beta\gamma''}{\alpha^2} - \frac{\epsilon'}{\alpha}.$$

Its general solution is thus

$$V = F(\rho,t) + (\beta''/\alpha)\rho\sin\varphi - (\gamma''/\alpha)\rho\cos\varphi + A\varphi,$$

or, in terms of $\xi$ and $\eta$,

$$V = \frac{\beta''}{\alpha}\eta - \frac{\gamma''}{\alpha}\xi + A\arctan\left(\frac{\eta}{\xi}\right) + F(\xi^2 + \eta^2, t), \tag{17}$$

with

$$A = (\beta''\gamma - \beta\gamma'')/\alpha^2 - \epsilon'/\alpha.$$

Formulas (15) and (17) exhaust all the possible forms of potential for which a constant linear in the velocities exists.

## III. CONSTANTS QUADRATIC IN THE VELOCITIES

The general form of such a constant is

$$C = f^0\dot{x}^2 + f^1\dot{x}\dot{y} + f^2\dot{y}^2 + g^0\dot{x} + g^1\dot{y} + h. \tag{18}$$

Following the method of Sec. II, we write $dC/dt$ as a polynomial of degree 3 in $\dot{x}$ and $\dot{y}$:

$$\frac{dC}{dt} = f^0_x\dot{x}^3 + (f^0_y + f^1_x)\dot{x}^2\dot{y} + (f^1_y + f^2_x)\dot{x}\dot{y}^2 + f^2_y\dot{y}^3 + (g^0_x + f^0_t)\dot{x}^2 + (g^0_y + f^1_t + g^1_x)\dot{x}\dot{y} + (f^2_t + g^1_y)\dot{y}^2 + (g^0_t + h_x + 2f^0\ddot{x} + f^1\ddot{y})\dot{x} + (g^1_t + h_y + f^1\ddot{x} + 2f^2\ddot{y})\dot{y} + (h_t + g^0\ddot{x} + g^1\ddot{y}). \quad (19)$$

We are thus led to systems of partial differential equations for the $f_i$, $g_i$, and $h$.

The first set of equations for the $f_i$'s can be easily integrated:

$$f_0 = \alpha y^2 + \beta y + \gamma,$$
$$f_1 = -2\alpha xy - \beta x - \delta y - \epsilon, \quad (20)$$
$$f_2 = \alpha x^2 + \delta x + \zeta.$$

The functions $f_i$ have the same quadratic dependence in $x$ and $y$ as in the case of a time-independent Hamiltonian.[21] The main difference, here, stems from the fact that the coefficients depend explicitly on time. The constant $C$ (18) can also be written in terms of the angular momentum $L = x\dot{y} - y\dot{x}$:

$$C = \alpha L^2 - \beta\dot{x}L + \delta\dot{y}L + \gamma\dot{x}^2 - \epsilon\dot{x}\dot{y} + \zeta\dot{y}^2 + g^0\dot{x} + g^1\dot{y} + h. \quad (21)$$

The remaining equations have the form

$$g^0_x + f^0_t = 0, \quad g^0_y + f^1_t + g^1_x = 0, \quad g^1_y + f^2_t = 0; \quad (22)$$
$$g^0_t + 2f^0\ddot{x} + f^1\ddot{y} + h_x = 0, \quad g^1_t + f^1\ddot{x} + 2f^2\ddot{y} + h_y = 0; \quad (23)$$
$$h_t + g^0\ddot{x} + g^1\ddot{y} = 0. \quad (24)$$

From the knowledge of the functions $f_i$, system (22) allows the calculation of the function $g_i$, providing the following compatibility condition is satisfied:

$$(f^0_{yy} - f^1_{xy} + f^2_{xx})_t = 0.$$

Due to the special form (20) of the functions $f_i$, this condition reduces to $\alpha'(t) = 0$. Once the $g_i$'s are known, a second compatibility condition, which allows the calculation of $h$, results from the system (23)

$$\frac{\partial}{\partial y}(g^0_t + 2f^0\ddot{x} + f^1\ddot{y}) = \frac{\partial}{\partial x}(g^1_t + f^1\ddot{x} + 2f^2\ddot{y}). \quad (25)$$

That is

$$f_1(V_{yy} - V_{xx}) + 2(f_0 - f_2)V_{xy} + (2f^0_y - f^1_x)V_x - (2f^2_x - f^1_y)V_y = g^0_{ty} - g^1_{tx}. \quad (26)$$

This last relation is quite similar to the one obtained in the search of a time-independent potential $V$ that admits a constant of motion quadratic in the velocities.[21] The difference is only in the existence of a nonhomogeneous part in this linear PDE. This remark will lead us to the same classification as in the autonomous case. Before proceeding further, let us point out that, once $V$ is determined satisfying (26), there remains a last relation (24) to check for the system to possess a quadratic invariant. This was not the case for time-independent potentials and, as we will see further, this relation strongly reduces the admissible forms of potentials.

We will distinguish three distinct cases, according to the value of the highest power of the angular momentum $L$ that appears in the constant (21). In each case, we will reduce the form of the invariant by translations and rotations of coordinates. One can note that a rotation does not change $L$, while a translation keeps $\dot{x}$ and $\dot{y}$ invariant.

*Case* (*a*): $\alpha = \beta = \delta = 0$. This is the separable case. There is no dependence on $L$ in the constant $C$.

By an adequate rotation, the coefficient of $\dot{x}\dot{y}$ can be set to zero unless $(\gamma - \zeta)^2 + \epsilon^2 = 0$.

*Case* (*b*): $\alpha = 0$, $\beta$ (or $\delta$) nonvanishing. Translations of $x$ and $y$ allow us to eliminate $\gamma - \zeta$ and $\epsilon$ and an adequate rotation of coordinates allows the choice $\delta = 0$, unless $\beta^2 + \delta^2 = 0$ as can be easily seen from the form (21) of the constant $C$.

*Case* (*c*): $\alpha \neq 0$. Translation of $x$ and $y$ allow the elimination of all the linear in $L$ terms in (21). ($\beta = \delta = 0$.) Then, by a rotation, $\zeta$ can be set equal to zero unless $\epsilon^2 + (\gamma - \zeta)^2 = 0$.

We will now proceed with the integration of Eqs. (22)–(24) in each of the distinct reduced cases (a), (b), and (c).

*Case* (*a*): $f_0 = \gamma$, $f_1 = 0$, $f_2 = \zeta$. The integration of Eq. (22) leads to

$$g_0 = -\gamma' x + \theta y + \lambda, \quad g_1 = -\zeta' y - \theta x + \kappa. \quad (27)$$

In the following, we will choose $\lambda = \kappa = 0$; it corresponds to an adequate translation of coordinates. However, with $\theta \neq 0$, there exists no solution to the system other than trivial harmonic oscillator: $V = \frac{1}{2}\Phi(t)x^2 + \frac{1}{2}\Psi(t)y^2$. Apparently, the condition $\theta \neq 0$ imposes severe constraints on the potential. For this reason, we will look for solutions with $\theta = 0$.

The condition (25) writes

$$2(\gamma - \zeta)V_{xy} = 0,$$

or

$$V = F(x,t) + G(y,t). \quad (28)$$

Integration of Eq. (23) for $h$ is straightforward and leads to

$$h = 2\gamma F + 2\zeta G + \tfrac{1}{2}\gamma'' x^2 + \tfrac{1}{2}\zeta'' y^2. \quad (29)$$

And finally, Eq. (24) takes the following form:

$$2\gamma' F + 2\zeta' G + 2\gamma F_t + 2\zeta G_t + \tfrac{1}{2}\gamma''' x^2 + \tfrac{1}{2}\zeta''' y^2 = -\gamma' x F_x - \zeta' y G_y.$$

This equation separates (up to a function of time, to be included in $F$ or $G$) into the system

$$2\gamma' F + 2\gamma F_t + \gamma' x F x = -\tfrac{1}{2}\gamma''' x^2, \quad (30)$$
$$2\zeta' G + 2\zeta G_t + \zeta' y G_y = -\tfrac{1}{2}\zeta''' y^2.$$

In order to solve the equation for $F$, we look for a particular solution of the form $F(x) = x^2\Psi(t)$. It leads to

$$\Psi' + 2(\gamma'/\gamma)\Psi = -\tfrac{1}{4}(\gamma'''/\gamma),$$

whose solution writes

$$\Psi(t) = -\frac{1}{4}\left[\frac{\gamma''}{\gamma} - \frac{1}{2}\left(\frac{\gamma'}{\gamma}\right)^2\right] = -\frac{1}{2}\frac{\sigma''}{\sigma},$$

with $\gamma = \sigma^2$.

The general solution of (30) is now easy to obtain; it writes

$$F(x,t) = (1/\gamma)\chi(x/\sqrt{\gamma}) + x^2\Psi(t). \quad (31)$$

The expression of $G$ is similar in terms of $\zeta$ and $y$, instead of $\gamma$ and $x$.

We will now investigate the particular case $\epsilon^2 + (\gamma - \zeta)^2 = 0$, where the rotation is not possible.

One obtains for $f_i$ and $g_i$

$$f_0 = \gamma, \quad f_1 = i(\gamma - \zeta), \quad f_2 = \zeta,$$
$$g^0 = -\gamma' z, \quad g^1 = i\zeta' z$$

[with the integration constants taken equal to zero as in the case (a) above], with $z = x + iy$.

The partial differential equation for $V$ (26) has the form

$$\epsilon(V_{yy} - V_{xx}) - 2i\epsilon V_{xy} = -i\sigma'',$$

with $\sigma = \zeta + \gamma, \epsilon = i(\gamma - \zeta)$, or in terms of $z$ and $\bar{z} = x - iy$,

$$-4i\epsilon V_{\bar{z}\bar{z}} = \sigma''.$$

The integration for $V$ is thus trivial;

$$V = i(\sigma''/8\epsilon)\bar{z}^2 + \bar{z}F_z(z,t) + G(z,t). \tag{32}$$

The system of PDE for $h$, in terms of $z$ and $\bar{z}$, reads

$$\begin{aligned} h_z &= \tfrac{1}{2}[-g_t^0 + ig_t^1 + V_z(2f_0 + 2f_2) \\ &\quad + V_{\bar{z}}(2f_0 - 2f_2 - 2if_1)], \\ h_{\bar{z}} &= \tfrac{1}{2}[-g_t^0 - ig_t^1 \\ &\quad + V_z(2f_0 - 2f_2 + 2if_1) + V_{\bar{z}}(2f_0 + 2f_2)]. \end{aligned} \tag{33}$$

In this precise case, due to the values of $g$ and $f$, it reduces to

$$h_z = -i\epsilon''(z/2) + \sigma V_z - 2i\epsilon V_{\bar{z}}, \quad h_{\bar{z}} = \sigma''(z/2) - \sigma V_{\bar{z}},$$

leading thus to

$$h = \sigma V + (z\bar{z}/2)\sigma'' - i\epsilon''(z^2/4) - 2i\epsilon F. \tag{34}$$

The last relation is now the following:

$$h_t = -\sigma' z(\bar{z}F_z + G_z) - \mu' z[(\sigma''/4\mu)\bar{z} + F].$$

Separating the different terms according to their dependence on $\bar{z}$ in this last relation, we obtain

$$(\sigma\sigma''/\epsilon)' = 0 \quad (\text{terms in } \bar{z}^2),$$

hence $\sigma\sigma''/\epsilon = 4\nu$ (const); (35)

$$\sigma' F_z + \sigma F_{zt} + \sigma' z F_{zz} = -\frac{z}{4}\left(2\sigma''' + \frac{\epsilon'\sigma''}{\epsilon}\right) \quad (\text{terms in } \bar{z}^1); \tag{36}$$

$$\sigma' G + \sigma G_t + \sigma' z G_z = -\epsilon'''(z^2/4) - 2\epsilon' F - 2\epsilon F_t - \epsilon' z F_z$$

(terms in $\bar{z}^0$). (37)

One recognizes readily in (36) and (37) the same left-hand side as in Eq. (30). Integrating, we obtain for $F$:

$$F = \Phi(z/\sigma) - \tfrac{1}{2}mz^2, \tag{38}$$

where $m = \frac{1}{2}\sigma''/\sigma - \frac{1}{4}(\sigma'/\sigma)^2 + \nu\epsilon/\sigma^2$.

Once $F$ is known, it is easy to check that the right-hand side of (37) is of the form $nz^2 - 2\epsilon'\Phi + (1/\sigma)(2\epsilon\sigma'/\sigma^2 - \epsilon')z\Phi'$, where $n = -\mu'''/4 - 2\mu' m - \mu m'$.

The solution of (37) is thus the following:

$$G = \frac{1}{\sigma}\Psi\left(\frac{z}{\sigma}\right) + pz^2 - \frac{2\epsilon}{\sigma}\Phi(qz), \tag{39}$$

where,

$$p = -\sigma^{-3}\int \eta\sigma^2\, dt, \quad q = \frac{1}{\sigma}\ln\frac{\sqrt{\epsilon}}{\sigma}.$$

(The case of power-like $\Phi$ should in principle be treated apart).

Relations (32), (38), and (39) determine the precise form of a potential $V$ for which a constant of that kind exists.

Let us proceed now with the search of a constant linear in $L$.

*Case* (*b*): In this case, we obtain for $f_i$ and $g_i$:

$$f^0 = \beta y + \gamma, \quad f^1 = -\beta x, \quad f^2 = \gamma; \tag{40}$$

$$g^0 = -(\beta' y + \gamma')x + \theta y + \lambda,$$
$$g^1 = -\gamma' y + \beta' x^2 - \theta x + \kappa. \tag{41}$$

The analysis of the corresponding case for time-independent Hamiltonians has shown that the adequate variables were

$$u = \rho + \eta, \quad v = \rho - \eta,$$

where

$$\rho = \sqrt{x^2 + y^2}, \quad \eta = y.$$

Moreover, a detailed analysis has shown that there is no solution with nonvanishing $\lambda$ and $\kappa$ and we will thus put them to zero in order to alleviate the presentation. In terms of $u$ and $v$, Eqs. (23) take the following form:

$$\begin{aligned} h_u &= (-\beta v + 2\gamma)V_u + \frac{1}{4}(-\beta'' v + \gamma'')(u + v) \\ &\quad + \frac{1}{4}\frac{\theta' v(u + v)}{\sqrt{uv}}, \\ h_u &= (\beta u + 2\gamma)V_u + \frac{1}{4}(\beta'' u + \gamma'')(u + v) \\ &\quad - \frac{1}{4}\frac{\theta' u(u + v)}{\sqrt{uv}}. \end{aligned} \tag{42}$$

The compatibility equation resulting from (42) leads to a PDE for $V$ in terms of the independent variables $u$ and $v$. Its solution is straightforward:

$$V = \frac{F(u) + G(v)}{u + v} - \frac{3}{8}\frac{\beta''}{\beta}uv + \frac{2}{3}\frac{\theta'}{\beta}\sqrt{uv}, \tag{43}$$

as is the integration of (42):

$$\begin{aligned} h &= \beta\frac{uG(v) - vF(u)}{u + v} + 2\gamma\frac{F(u) + G(v)}{u + v} + \frac{1}{8}\beta' uv(v - u) \\ &\quad + \frac{1}{8}\gamma''(u + v)^2 - \frac{3}{4}\frac{\gamma\beta'' uv}{\beta} + \frac{4\gamma\theta'}{3\beta}\sqrt{uv} \\ &\quad + \frac{1}{6}\theta'(u - v)\sqrt{uv}. \end{aligned} \tag{44}$$

There remains now Eq. (24) to be checked. This relation will impose constraints on the form of the potential $V$ and on the time-dependent functions $\beta, \gamma, \theta$. Three different kinds of terms, functions of $u$ alone and $v$ alone, as well as terms where $u$ and $v$ are mixed together, are involved. These three families of functions will give three distinct relations. After some lengthy manipulations, it results, from the mixed term relation, that $\theta$ and $\beta''$ must vanish. In this case, this last relation reduces to

$$G + (\beta/\beta')G_t + vG_v = 0, \quad F + (\beta/\beta')F_t + uF_u = 0. \tag{45}$$

On the other hand, the terms that depend only on $u$ in the relation (24) lead to the following constraint:

$$F + (2\gamma/\gamma')F_t + uF_u = 0.$$

Similarly, we have for $G$:

$$G + (2\gamma/\gamma')G_t + vG_v = 0.$$

All these equations are compatible only if $\gamma = \beta^2$.

We recognize in (45), Eq. (30) up to the right-hand side and we thus obtain the solution of (45):

$$F(u,t) = (1/\beta)\Phi(u/\beta), \quad G(v,t) = (1/\beta)\Psi(v/\beta).$$

In conclusion, the potential

$$V = \frac{\Phi\,[(y + \sqrt{x^2+y^2})/\beta\,] + \Psi\,[(\sqrt{x^2+y^2} - y)/\beta\,]}{\beta\sqrt{x^2+y^2}},$$

$$\beta = c_1 + c_2 t \tag{46}$$

is the general form of potential for which a constant of motion linear in the angular momentum exists. The particular case where $\delta = i\beta$ must be treated apart.

One obtains, in this case, for $f_i$ and $g_i$:

$$f_0 = \beta y + \gamma, \quad f_1 = -\beta x - i\beta y = -\beta z,$$

$$f_2 = i\beta x + \gamma;$$

$$g_0 = -\beta' y\bar{z} - \gamma' x, \quad g_1 = \beta' x\bar{z} - \gamma' y;$$

with $z = x + iy$, $\bar{z} = x - iy$.

The partial differential equation for $V$ [(26)] in terms of $z$ and $\bar{z}$ reduces to

$$2zV_{zz} + 3V_z = -\tfrac{3}{2}(\beta''/\beta)\bar{z}. \tag{47}$$

Integrating (47), one finds

$$V = z^{-1/2}F(\bar{z},t) + G_{\bar{z}}(\bar{z},t) - \tfrac{1}{2}(\beta''/\beta)z\bar{z}. \tag{48}$$

It is now easy to obtain from (33) the equations for $h$:

$$h_z = \tfrac{1}{2}(\gamma'' + i\beta''\bar{z})\bar{z} + (i\beta\bar{z} + 2\gamma)V_z,$$

$$h_{\bar{z}} = \tfrac{1}{2}(\gamma'' - i\beta''\bar{z})z - 2i\beta z V_z + (i\beta\bar{z} + 2\gamma)V_{\bar{z}}.$$

From these we deduce the value of $h$:

$$h = (i\beta\bar{z} + 2\gamma)V + \tfrac{1}{2}(\gamma'' + i\beta''\bar{z})z\bar{z} - i\beta G(\bar{z},t). \tag{49}$$

As before, there remains a last relation (24) to be verified. It involves three distinct and independent families of terms; namely functions of $\bar{z}$ that multiply either $z^{-1/2}$, $z$, or 1. We thus obtain the following three equations:

$$2\gamma'\frac{\beta''}{\beta} + \gamma\left(\frac{\beta''}{\beta}\right)' - \frac{1}{2}\gamma''' = 0,$$

$$\frac{3}{2}(i\beta'\bar{z} + \gamma')F + (i\beta\bar{z} + 2\gamma)F_t + (\gamma' + i\beta'\bar{z})\bar{z}F_{\bar{z}} = 0,$$

$$G_{\bar{z}t}(i\beta\bar{z} + 2\gamma) + G_{\bar{z}}(i\beta'\bar{z} + 2\gamma') + (i\beta'\bar{z} + \gamma')\bar{z}G_{\bar{z}\bar{z}}$$

$$= i\beta' G + i\beta G_t.$$

By analogy to the general case (b), we will look for solutions with $\gamma = 0$. Indeed in that case, equations reduce to

$$\frac{3}{2}\beta' F + \beta F_t + \beta'\bar{z}F_{\bar{z}} = 0, \tag{50}$$

$$\bar{z}(\beta G_{t\bar{z}} + \beta' G_{\bar{z}} + \beta'\bar{z}G_{\bar{z}\bar{z}}) = \beta' G + \beta G_t, \tag{51}$$

and we recognize the form of the Eq. (45) for Eq. (50).

The general solution for $F$ thus reads

$$F = (1/\beta^{3/2})\chi(\bar{z}/\beta). \tag{52}$$

Concerning the equation for $G$, the change of the dependent variable

$$H = \bar{z}G_{\bar{z}} - G$$

leads to the expression

$$\beta' H + \beta H_t + \beta'\bar{z}H_{\bar{z}} = 0,$$

the solution of which reads

$$H = (1/\beta)\Omega(\bar{z}/\beta),$$

or, equivalently,

$$H = \frac{\bar{z}^2}{\beta}\frac{d}{d\bar{z}}\left[\frac{1}{\bar{z}}\Phi\left(\frac{\bar{z}}{\beta}\right)\right], \tag{53}$$

and, finally,

$$G = (1/\beta)\Phi(\bar{z}/\beta) + \bar{z}\Psi(t).$$

In conclusion, provided $V$ is given in terms of $z$ and $\bar{z}$ by formulas (48), (52), and (53), a constant of the motion, linear in $L$ exists with the condition $\beta = i\delta$.

*Case* (*c*): The last case to examine is the case where the constant $C$ has a quadratic dependence in $L$, namely, after the necessary reductions, the constant (21) assumes the form

$$C = \alpha L^2 + \gamma\dot{x}^2 + \zeta\dot{y}^2 + g_0\dot{x} + g_1\dot{y} + h.$$

As we have seen before, [(9)], $\alpha$ must be time-independent and will be taken equal to 1.

We first examine the case $\gamma = \zeta$.

The corresponding values of the $f_i$'s and $g_i$'s are

$$f_0 = y^2 + \gamma, \quad f_1 = -2xy, \quad f_2 = x^2 + \gamma;$$

$$g_0 = -\gamma' x + \lambda y, \quad g_1 = -\gamma' y - \lambda x.$$

The PDE for $V$ is easily solved using polar coordinates and leads to the following form for $V$:

$$V = F(\theta,t)/\rho^2 + G(\rho,t) - \theta\lambda'(t)/2, \tag{54}$$

$$\theta = \arctan(y/x), \quad \rho^2 = x^2 + y^2.$$

We write $h$ as

$$h = 2\gamma(F/\rho^2) + 2F + 2\gamma G - \gamma\lambda'\theta + \gamma''\rho^2/2. \tag{55}$$

The last compatibility relation, in terms of $\rho$, $\theta$, and $t$ reads

$$h_t = -\gamma'_\rho V_\rho - \lambda V_\theta.$$

Again, three different kinds of terms appear which must vanish separately; this leads to the following relations:

$$2F_t - \theta\gamma'\lambda' - \gamma\theta\lambda'' = \lambda\lambda'/2,$$

$$(\gamma'''/2)\rho^2 + 2\gamma' G + 2\gamma G_t = -\gamma'\rho G_\rho,$$

$$2\gamma F_t = -\lambda F_\theta.$$

The equation for $G$ has been encountered previously. Its solutions is

$$G(\rho,t) = \frac{1}{\gamma}\Phi\left(\frac{\rho^2}{\gamma}\right) - \frac{\rho^2}{4}\left[\frac{\gamma''}{\gamma} - \frac{1}{2}\left(\frac{\gamma'}{\gamma}\right)^2\right]. \tag{56}$$

Now, the two distinct equations for $F$ put constraints on $\gamma$ and $\lambda$.

A solution

$$F = \lambda^2/8 - \theta k/2 \tag{57}$$

exists whenever $k = \gamma\lambda'$ is time independent.

The general solution $V$ is thus the following:

$$V = \frac{1}{\gamma}\Phi\left(\frac{x^2+y^2}{\gamma}\right) - \frac{x^2+y^2}{4}\left[\frac{\gamma''}{\gamma} - \frac{1}{2}\left(\frac{\gamma'}{\gamma}\right)^2\right] + \frac{1}{x^2+y^2}\left(\frac{\lambda^2}{8} - \frac{k}{2}\arctan\frac{y}{x}\right) - \frac{k}{2\gamma}\arctan\frac{y}{x}, \tag{58}$$

with

$$\lambda = k\int_0^t \frac{d\tau}{\gamma}.$$

When $\lambda = 0$, the solution is slightly different, because $F$ can be any time-independent function of $\theta$. If $\gamma = 0$, one has

$$V = \frac{F(\theta)}{\rho^2} + G(\rho,t). \tag{59}$$

If $\gamma \neq 0$, one has

$$V = \frac{F(\theta)}{\rho^2} + \frac{1}{\gamma}\Phi\left(\frac{\rho^2}{\gamma}\right) - \frac{\rho^2}{4}\left[\frac{\gamma''}{\gamma} - \frac{1}{2}\left(\frac{\gamma'}{\gamma}\right)^2\right]. \tag{60}$$

We have also examined in detail the case $\gamma \neq \zeta$, but it failed to yield any solution. Thus we will not exhibit any calculations here. This completes the study of the time-dependent Hamiltonian with an exact invariant quadratic in the velocities.

## IV. COMPARISON WITH OTHER RESULTS

As was stated in the Introduction, there exist numerous studies on one-dimensional time-dependent systems but significantly fewer on two- (or more-) dimensional ones. Moreover, as the Ermakov approach is most often employed, the equations of motion are usually not Hamiltonian. Still some comparisons with existing results can be made.

To start with, our separable case should encompass the results for the one-dimensional systems of Lewis and Leach.[19] They dealt with time-dependent Hamiltonian systems in one dimension $H = \frac{1}{2}p^2 + V(x,t)$, and gave, in particular, conditions for which a potential $V(x,t)$ possesses a constant quadratic in the velocity $p$. These conditions determine the precise form of $V$ in terms of arbitrary functions of time, namely,

$$V(x,t) = -F(t)x + \frac{1}{2}\Omega^2(t)x^2 + \frac{1}{\rho^2}U\left(\frac{x-v}{\rho}\right). \tag{61}$$

Here, $U$ is an arbitrary function of its argument and $F, \Omega^2, \rho$, and $v$ are arbitrary functions submitted to the following constraints:

$$\ddot\rho + \Omega^2(t)\rho - k/\rho^3 = 0, \quad \ddot v + \Omega^2(t)v = F(t).$$

In Sec. II, we examined separable potentials $V = F(x,t) + G(y,t)$. Thus, the two directions $x$ and $y$ decouple and we have in fact two one-dimensional Hamiltonians. The following form was found for $F(x,t)$ [and similarly for $G(y,t)$]:

$$F(x,t) = \frac{1}{\gamma}\Phi\left(\frac{x}{\sqrt{\gamma}}\right) - \frac{x^2}{4}\left[\frac{\gamma''}{\gamma} - \frac{1}{2}\left(\frac{\gamma'}{\gamma}\right)^2\right].$$

This form is identical to (61) up to a translation of $x$, which would correspond to a choice of $\lambda \neq 0$ in $g_0$.

We turn now to genuine two-dimensional systems. In a recent publication, Lutzky[9] proved that the quantity

$$C = \frac{1}{2}(x\dot y - \dot x y)^2 + \int_0^{x/y}\Psi(\lambda)\frac{d\lambda}{\lambda}$$

is conserved by the motion described by the equations

$$\ddot x + \omega^2(t)x = f_1(x,y), \quad \ddot y + \omega^2(t)y = f_2(x,y) \tag{62}$$

provided that $f_1$ and $f_2$ satisfy the relation

$$xf_2 - yf_1 = (1/xy)\Psi(x/y). \tag{63}$$

This constant $C$ is quadratic in $L$ and corresponds to the case (c) we introduced in Sec. III. Moreover we have $\gamma = \zeta = 0$. We have found that the only form of the potential $V$ in that case was given by (59). In order to compare our results to Lutzky's, we must examine under which conditions his equations derive from a Hamiltonian. This happens if there exists a function $W$ such that $f_1 = W_x, f_2 = W_y$. Relation (62) is thus equivalent to

$$xW_y - yW_x = \Psi(x/y)/xy,$$

which in polar coordinates reads

$$\frac{\partial W}{\partial\theta} = \frac{F'(\theta)}{\rho^2},$$

$F$ determined in terms of $\Psi$. That is, $W = F(\theta)/\rho^2 + G(\rho,t)$.

This result coincides with (59).

So we conclude that Lutzky's result is identical to ours whenever the equations of motion (62) are Hamiltonian.

## V. CONCLUSION

In this paper, we have presented an investigation of time-dependent Hamiltonians in two space dimensions from the point of view of the existence of an exact invariant. The method we have used was the direct computation of the invariant, which was employed in our previous work on 2-D time-independent Hamiltonians, as well as in the work of Lewis and Leach on 1-D systems. We were able to identify the forms of the time-dependent potential for which an invariant linear or quadratic in the velocities exists. Thus our results extend the results obtained previously by various groups to the case of Hamiltonian systems. In particular, they contain as a special case the Hamiltonians of Lewis and Leach and have a nonzero overlap with Lutzky's results.


[1]H. R. Lewis, J. Math. Phys. **9**, 1976 (1968).
[2]V. P. Ermakov, Unit Izv. Kiev **20**, (9) 1 (1880).
[3]B. Gambier, Acta Math. **33**, 45 (1910).
[4]H. R. Lewis, Jr. and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[5]M. D. Kruskal, J. Math. Phys. **3**, 806 (1962).
[6]P. G. L. Leach, SIAM J. Appl. Math. **34**, 496 (1978).
[7]M. Lutzky, Phys. Lett. A **68**, 3 (1978); see also P. G. L. Leach, Phys. Lett. **84**, 161 (1981).
[8]J. Ray and J. L. Reid, Phys. Lett. A **71**, 317 (1979).
[9]M. Lutzky, Phys. Lett. A **78**, 301 (1980).
[10]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979).
[11]M. Lutzky, J. Math. Phys. **21**, 1370 (1980).
[12]J. L. Reid and J. R. Ray, J. Math. Phys. **21**, 1583 (1980).
[13]J. R. Ray, Phys. Lett. A **78**, 4 (1980).
[14]J. R. Ray, J. L. Reid, and M. Lutzky, Phys. Lett. A **84**, 42 (1981).
[15]J. L. Reid and J. R. Ray, J. Math. Phys. **23**, 503 (1982).
[16]J. R. Ray and J. L. Reid, J. Math. Phys. **22**, 91 (1981).
[17]W. Sarlet, Phys. Lett. A **82**, 161 (1981).
[18]W. Sarlet and J. P. Ray, J. Math. Phys. **22**, 2504 (1981).
[19]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).
[20]N. J. Günther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).
[21]J. R. Ray and J. L. Reid, Phys. Lett. A **74**, 23 (1979).
[22]W. Sarlet and F. Cantrijn, Phys. Lett. A **88**, 383 (1982).
[23]B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **24**, 2282 (1983).

# Integrals of motion for Toda systems with unequal masses

B. Dorizzi, B. Grammaticos, R. Padjen, and V. Papageorgiou
*CNET/PAA/TIM/MTI, 38-40, Avenue du Général Leclerc, 92131 Issy-les-Moulineaux, France*

(Received 5 April 1983; accepted for publication 30 December 1983)

We present new integrals of motion for the Toda lattice (chain of particles in one dimension with exponential interaction) for two special cases of boundary conditions: the free-end lattice with three non-equal-mass particles and the fixed-end lattice for two particles. In both cases, we use two distinct approaches in order to identify the integrable cases: direct search of the integral of motion and group theoretical methods. Our results are in agreement with the predictions of Painlevé analysis.

PACS numbers: 03.20. + i, 02.30. + g, 02.20. + b

## I. INTRODUCTION

The Toda lattice[1] is a one-dimensional system of equal mass particles interacting via nonlinear forces: the interactions occur only between nearest neighbors and are of exponential type. With this lattice, we are in presence of a "small miracle": The system is integrable for any number of particles in the chain. (Integrability, in the case of Hamiltonian systems of $N$ degrees of freedom, is synonymous to existence of $N$ analytical, single-valued integrals of the motion, time independent and in involution.) The integrability for periodic boundary conditions has been shown independently by Hénon,[2] Flaschka,[3] and Manakov.[4] The first has explicitly calculated the integrals of the motion, while the two others deduced the integrability from group theoretical methods.

Moreover, this system is also integrable for other boundary conditions. The integrals in the case of the fixed-end lattice (two ends of the chain are set permanently equal to zero) can be easily deduced from the periodic case as shown by Hénon.[2] The free-end lattice (the beginning of the chain is set to $-\infty$, the end is set to $+\infty$) has been discussed by Moser.[5] He has shown that the system admits the Lax-pair representation and has used the latter to calculate the $N$ integrals of the motion.

The aim of this work is to use this twofold approach, i.e., direct computation of the constant of motion and group theoretical methods (i.e., search for Lax representations), in order to study other cases of integrability in low-dimensional and unequal-mass systems. Indeed, the great number of integrability conditions to be satisfied in the general $N$ body case compels us to deal with particles of equal masses and equal ranges of interaction. However, one can reasonably hope that for systems of two or three particles, the constraints of equal masses can be relaxed. A very useful tool for the identification of integrability candidates is the Painlevé criterion as introduced by the work of Ablowitz, Ramani, and Segur.[6] They have conjectured that integrability is intimately related to the analytic properties of the solutions of the equations of motion. Namely, whenever the solutions possess the Painlevé property, i.e., their only movable singularities on the complex-time plane are poles, the system is integrable. The reciprocal is also true and has been verified for the known integrable systems. However recent results of Ramani, Dorizzi, and Grammaticos[7] have shown that for two-dimensional Hamiltonian systems, integrability can sometimes be associated to some weakened Painlevé property. In the case of the Toda system at hand, such a generalization is unnecessary, and the known integrable cases possess the full Painlevé property.

In this paper, we will concentrate on two particular forms of the Toda system.

(1) The free-end lattice with three masses:

$$H=\frac{p_1^2}{2m_1}+\frac{p_2^2}{2m_2}+\frac{p_3^2}{2}+e^{\epsilon(q_1-q_2)}+e^{q_2-q_3}.$$

The integrability of this system cannot be verified numerically by the surfaces of section method since the above Hamiltonian describes a scattering problem.

In a recent publication,[8] Bountis, Segur, and Vivaldi have found that the Painlevé property is satisfied for three values of the parameters:

(a) $m_1=\dfrac{\epsilon(2\epsilon-1)}{2-\epsilon}, \quad m_2=2\epsilon-1, \quad \dfrac{1}{2}<\epsilon<2,$

(b) $m_1=\dfrac{\epsilon(\epsilon-1)}{2-\epsilon}, \quad m_2=\epsilon-1, \quad 1<\epsilon<2,$

(c) $m_1=\dfrac{3\epsilon(2\epsilon-1)}{2-3\epsilon}, \quad m_2=2\epsilon-1, \quad \dfrac{1}{2}<\epsilon<\dfrac{2}{3}.$

The integrability of case (a) has been proved rigorously by Moser[5] and Bogoyavlenski.[9] Cases (b) and (c) can be deduced from the theorem of Bogoyavlenski[9] and are explicitely studied in Ref. 10.

In Sec. II, we will present a direct computation of the constants of motion for all the three cases (a), (b), and (c) above.

(2) The fixed-end lattice with two masses:

$$H=\frac{p_1^2}{2m_1}+\frac{p_2^2}{2m_2}+e^{-\delta q_1}+e^{\epsilon(q_1-q_2)}+e^{q_2}.$$

Casati and Ford predicted integrability for $\epsilon=\delta=1$, $m_1/m_2=1$, based on a numerical study.

The Painlevé analysis of Bountis *et al.* suggests three cases of integrability:

(a) $m_1/m_2=1, \quad \delta=\epsilon=1,$

(b) $m_1/m_2=1, \quad \delta=1, \quad \epsilon=\frac{1}{2},$

(c) $m_1/m_2=\frac{1}{3}, \quad \delta=1, \quad \epsilon=\frac{1}{2}.$

On the other hand, the Lie algebra study of Bogoyavlensky yields two integrability candidates: case (b) above and

 0022-2488/84/072200-12$02.50  

(d) $m_1/m_2 = \frac{1}{3}, \quad \delta = \frac{1}{3}, \quad \epsilon = \frac{1}{2}$.

In fact, we find five integrable combinations of masses and ranges: cases (a), (b), (c), (d) listed above and

(e) $m_1/m_2 = 1, \quad \delta = \frac{1}{2}, \quad \epsilon = \frac{1}{2}$.

Actually all five integrable cases can be predicted by the Painlevé analysis, as was shown by Ramani.[11]

Sec. IV deals exclusively with group theoretical methods. This approach leads to the identification of the integrable cases listed above as can be shown in several recent works.[9,10,12] However their approaches are general and rather abstract. Our aim in this section is to explicit these results in the various cases of integrability. In particular we exhibit the Cartan matrix for the three different cases for which the free-end Toda lattice admits the Lax-pair representation. This allows the calculation of the integrals in each case. The same procedure is applied to the fixed-end case.

## II. FREE-END LATTICE

Let us consider a free-end lattice with three, non-equal-mass, particles. In that case, the Hamiltonian governing the system reads

$$H = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{p_3^2}{2m_3} + e^{\epsilon(q_1 - q_2)} + e^{q_2 - q_3}. \tag{2.1}$$

In order to explicit the motion of the center of mass $z$ of the system, we introduce the following change of variable:

$$x = \epsilon(q_1 - q_2), \quad y = q_2 - q_3 \quad z = m_1 q_1 + m_2 q_2 + m_3 q_3. \tag{2.2}$$

The equations of motion associated to the system are derived from the Lagrange equations:

$$\frac{\partial}{\partial t}(m_1 \dot q_1) = -\epsilon X, \quad \frac{\partial}{\partial t}(m_2 \dot q_2) = \epsilon X - Y, \quad \frac{\partial}{\partial t}(m_3 \dot q_3) = Y, \tag{2.3}$$

with

$$X = e^{\epsilon(q_1 - q_2)} = e^x, \quad Y = e^{q_2 - q_3} = e^y.$$

From Eqs. (2.2), it is obvious that

$$\ddot z = m_1 \ddot q_1 + m_2 \ddot q_2 + m_3 \ddot q_3 = 0. \tag{2.4}$$

One can also obtain, from (2.2), the equations of the motion of $x$ and $y$:

$$\ddot x = \epsilon(\ddot q_1 - \ddot q_2) = \frac{\epsilon}{m_2}\left(Y - \epsilon \frac{m_1 + m_2}{m_1} X\right), \quad \ddot y = \ddot q_2 - \ddot q_3 = \frac{\epsilon}{m_2}\left(X - \frac{m_2 + m_3}{\epsilon m_3} Y\right), \tag{2.5}$$

which, after a scaling in time, read

$$\ddot x = Y - \alpha X, \quad \ddot y = X - \beta Y, \tag{2.6}$$

with

$$\alpha = \epsilon(m_1 + m_2)/m_1, \quad \beta = (m_2 + m_3)/\epsilon m_3.$$

The equations for $(x, y)$ and $z$ are separated. So, for the system (2.3) to be completely integrable it is sufficient that the system (2.6) itself be integrable. It already possesses a first integral of motion, namely the energy, obtained by subtraction of the center of mass energy from the Hamiltonian $H$ (written in terms of $x$ and $y$)

$$E = \frac{m_3(m_1 + m_2)}{2\alpha(m_1 + m_2 + m_3)}(\beta \dot x^2 + 2\dot x \dot y + \alpha \dot y^2) + X + Y. \tag{2.7}$$

We will then look for a second integral polynomial in $\dot x$ and $\dot y$, i.e., of the form

$$C = \sum_{n=0}^{N} \sum_{k=0}^{n} f_k^n \dot x^{n-k} \dot y^k, \tag{2.8}$$

where $f_k^n$ are general functions of $x, y$. (Only powers of the same parity in the velocities will appear in the sum due to the time reversal invariance of the Hamiltonian.) We will exhaust all the possible constants of the form (2.8) up to order $N = 6$.

This method has already proved to be a valuable tool in the study of integrable dynamical systems with polynomial potentials. In particular, in the case of polynomial potentials of degree 3, it allows the calculation of the integrals of motion in *all* the Painlevé cases.[13] This will be the case for the Toda lattice as well.

The complete details for the search of a constant of orders 2, 3, and 4 with a general potential are exposed in Ref. 13. In this paper, we will just present the calculation in the particular case of the Toda potentials.

A calculation at order 2 does not give any result.

At order 3, the form of the integral is

$$C = f_0 \dot x^3 + f_1 \dot x^2 \dot y + f_2 \dot x \dot y^2 + f_3 \dot y^3 + g_0 \dot x + g_1 \dot y. \tag{2.9}$$

A direct computation shows that the $f_i$'s must be constant.

The condition $dC/dt = 0$ leads to a system of partial differential equations for the $g_i$'s:

$$3 f_0 \ddot x + f_1 \ddot y + \frac{\partial g_0}{\partial x} = 0,$$

$$2 f_1 \ddot x + 2 f_2 \ddot y + \frac{\partial g_0}{\partial y} + \frac{\partial g_1}{\partial x} = 0,$$

$$f_2 \ddot x + 3 f_3 \ddot y + \frac{\partial g_1}{\partial y} = 0.$$

The compatibility condition, necessary for the integration of the equations for the $g_i$'s,

$$\frac{\partial^2}{\partial x^2}(f_2 \ddot x + 3 f_3 \ddot y) - \frac{\partial}{\partial x \, \partial y}(2 f_1 \ddot x + 2 f_2 \ddot y) + \frac{\partial^2}{\partial y^2}(3 f_0 \ddot x + f_1 \ddot y) = 0, \tag{2.10}$$

is here considerably simplified thanks to the form of the Eqs. (2.6) for $\ddot x$ and $\ddot y$. Indeed, as the functions $X$ and $Y$ depend only on $x$ and $y$, respectively, we are led to the conditions

$$3 f_0 = \beta f_1, \quad 3 f_3 = \alpha f_2. \tag{2.11}$$

One can then easily integrate the equations for the $g_i$'s:

$$g_0 = (\alpha\beta - 1) f_1 X + 2(\beta f_2 - f_1) Y, \quad g_1 = 2(\alpha f_1 - f_2) X + (\alpha\beta - 1) f_2 Y. \tag{2.12}$$

It remains just to satisfy the condition

$$g_0 \ddot x + g_1 \ddot y = 0.$$

This expression is an identity in terms of the independent functions of $x$ and $y$: $X^2, XY, Y^2$. Thus the coefficients $f_1$ and $f_2$ must satisfy three linear equations:

$$\alpha f_1(3-\alpha\beta)+2f_2=0,\quad -2f_1+\beta f_2(3-\alpha\beta)=0,$$
$$f_1(2\alpha-\alpha\beta-1)+f_2(2\beta-\alpha\beta-1)=0. \tag{2.13}$$

This system has a nontrivial solution $(f_1,f_2)$ whenever $\alpha$ and $\beta$ satisfy the two equations

$$\alpha(3-\alpha\beta)(2\beta-\alpha\beta-1)+2(2\alpha-\alpha\beta-1)=0,$$
$$\beta(3-\alpha\beta)(2\alpha-\alpha\beta-1)+2(2\beta-\alpha\beta-1)=0, \tag{2.14}$$

or equivalently

$$(\alpha-\beta)(\alpha\beta-1)^2=0.$$

The condition $\alpha\beta=1$ is to be discarded because, in this case, the change of variable (2.2) is not defined.

Thus $\alpha=\beta\,(\neq 1)$, the first equation (2.14) then reads

$$(\alpha-1)^2(\alpha+1)^2(\alpha-2)=0.$$

When $\alpha=\beta=2$ $(f_2=-f_1)$ or, in terms of $\epsilon$,

$$m_1=\frac{\epsilon(2\epsilon-1)}{2-\epsilon},\quad m_2=2\epsilon-1,\quad m_3=1,\quad \frac{1}{2}<\epsilon<2,$$

the system (2.5) is integrable and admits apart from the energy a second constant of motion cubic in the velocities

$$C_1=2\dot x^3+3\dot x^2\dot y-3\dot x\dot y^2-2\dot y^3$$
$$+9(e^x-2e^y)\dot x+9(2e^x-e^y)\dot y. \tag{2.15}$$

The associated free-end Toda lattice is then also integrable. We recover thus the first case quoted by Bountis *et al.* and treated independently by Moser and Bogoyavlenski.

Let us consider now an integral of order 4 in the velocities

$$C=f_0\dot x^4+f_1\dot x^3\dot y+f_2\dot x^2\dot y^2+f_3\dot x\dot y^3+f_4\dot y^4$$
$$+g_0\dot x^2+g_1\dot x\dot y+g_2\dot y^2+h(x,y). \tag{2.16}$$

As in the preceding case, we can restrict ourselves to constant $f_i$'s and $f_4$ can be taken equal to zero by adding to $C$ the suitable multiple of the square of the Hamiltonian.

We recall the form of the compatibility condition for the integrability of the $g_i$'s:

$$\frac{\partial^3}{\partial y^3}(4f_0\ddot x+f_1\ddot y)-\frac{\partial^3}{\partial y^2\,\partial x}(3f_1\ddot x+2f_2\ddot y)$$
$$+\frac{\partial^3}{\partial y\,\partial x^2}(2f_2\ddot x+3f_3\ddot y)-\frac{\partial^3}{\partial x^3}(f_3\ddot x+4f_4\ddot y)=0. \tag{2.17}$$

In the particular case where $\ddot x$ and $\ddot y$ are given by Eq. (2.5), this last equation reduces to

$$\frac{\partial^3}{\partial y^3}(4f_0Y-\beta f_1Y)+\frac{\partial^3}{\partial x^3}(f_3\alpha X)=0;$$

thus

$$f_3=0,\quad 4f_0=\beta f_1. \tag{2.18}$$

Using (2.17), the equations for the $g_i$'s can be integrated to give

$$g_0=f_1(\alpha\beta-1)X+(2\beta f_2-3f_1)Y,$$
$$g_1=(3\alpha f_1-2f_2)X-2f_2Y, \tag{2.19}$$
$$g_2=2\alpha f_2X.$$

The second compatibility condition for the integrability of $h$,

$$\frac{\partial}{\partial y}(2g_0\ddot x+g_1\ddot y)=\frac{\partial}{\partial x}(g_1\ddot x+2g_2\ddot y), \tag{2.20}$$

is an identity in terms of the independent functions $X^2, XY, Y^2$, as in the preceeding case. We obtain thus a system in terms of $f_1, f_2, \alpha, \beta$:

$$\alpha f_1-2f_2=0,\quad f_1-\beta f_2=0,$$
$$f_1(3\alpha-\alpha\beta-2)+2f_2(\beta-\alpha)=0. \tag{2.21}$$

If the conditions

$$\alpha\beta=2,\quad (1-\alpha)(\alpha-2)=0$$

hold, the system (2.21) will have a nontrivial solution $(f_1,f_2)$. We have thus obtained that whenever

$$\text{(a)}\quad \alpha=1,\beta=2\quad \text{or}\quad \text{(b)}\quad \alpha=2,\beta=1, \tag{2.22}$$

the system (2.6) is integrable and possesses a second integral quartic in the velocities.

The two cases (a) and (b) are related by changing $\epsilon$ in $2\epsilon$, i.e., a scaling on $x$ ($x$ in $x/2$). The case (b) writes, in terms of $m_1$, $m_2$, $m_3$, and $\epsilon$,

$$m_1=\frac{\epsilon(\epsilon-1)}{2-\epsilon},\quad m_2=\epsilon-1,\quad m_3=1,\quad 1<\epsilon<2. \tag{2.23}$$

In that case, the constant $C_2$ is given by

$$C_2=\dot x^4+4\dot x^3\dot y+4\dot x^2\dot y^2+4(e^x-e^y)\dot x^2$$
$$+8(2e^x-e^y)\dot x\dot y+16e^x\dot y^2+4e^{2y}. \tag{2.24}$$

So the Toda lattice related to that case is integrable. The values of the parameters $m_i$ and $\epsilon$ correspond to the second case provided by the Painlevé analysis (cf. Ref. 8).

The case of a constant of motion quintic in the velocities has been examined but does not yield any positive result.

Let us now consider the case of a constant of order 6 in the velocities. The computations are similar but more complicated.

The form of a sixth-order constant is

$$C=e_0\dot x^6+e_1\dot x^5\dot y+e_2\dot x^4\dot y^2+e_3\dot x^3\dot y^3+e_4\dot x^2\dot y^4$$
$$+e_5\dot x\dot y^5+e_6\dot y^6+f_0\dot x^4+f_1\dot x^3\dot y+f_2\dot x^2\dot y^2+f_3\dot x\dot y^3$$
$$+f_4\dot y^4+g_0\dot x^2+g_1\dot x\dot y+g_2\dot y^2+h. \tag{2.25}$$

As in the preceding cases, the choice of constant $e_i$'s emerges naturally in the computation ($e_6$ can be taken equal to zero by adding to $C$ a multiple of the cube of the Hamiltonian). Now, in order to equate to zero the coefficients of order 5 in $dC/dt=0$, we obtain a system of partial differential equations for the $f_i$, which leads to the new compatibility condition (2.26). As soon as this last condition is satisfied, one can calculate the functions $f_i$. The problem is then reduced to the search of the $g_i$'s and $h$ from relations that read exactly the same as in the previous case of constants of order 4 in the velocities.

The new compatibility condition gives

$$\frac{\partial^5}{\partial y^5}(6e_0\ddot x+e_1\ddot y)-\frac{\partial^5}{\partial y^4\,\partial x}(5e_1\ddot x+2e_2\ddot y)$$
$$+\frac{\partial^5}{\partial x^2\,\partial y^3}(4e_2\ddot x+3e_3\ddot y)-\frac{\partial^5}{\partial x^3\,\partial y^2}(3e_3\ddot x+4e_4\ddot y)$$
$$+\frac{\partial^5}{\partial x^4\,\partial y}(2e_4\ddot x+5e_5\ddot y)-\frac{\partial^5}{\partial x^5}(e_5\ddot x)=0, \tag{2.26}$$

which immediately gives

$$6e_0 = e_1\beta, \quad e_5 = 0. \tag{2.27}$$

The integration for the $f_i$ is straightforward:

$$\begin{aligned}
f_0 &= e_1(\alpha\beta - 1)X + (2e_2\beta - 5e_1)Y &&\equiv A_0X + B_0Y,\\
f_1 &= (5e_1\alpha - 2e_2)X + (3e_3\beta - 4e_2)Y &&\equiv A_1X + B_1Y,\\
f_2 &= (4e_2\alpha - 3e_3)X + (4e_4\beta - 3e_3)Y &&\equiv A_2X + B_2Y,\\
f_3 &= (3e_3\alpha - 4e_4)X - 2e_4Y &&\equiv A_3X + B_3Y,\\
f_4 &= 2e_4\alpha X &&\equiv A_4X.
\end{aligned} \tag{2.28}$$

The next compatibility condition (2.17) reduces to

$$\begin{aligned}
&\beta B_1 - 4B_0 = 0,\\
&\alpha A_3 - 4A_4 = 0,\\
&2(B_2\alpha - A_2) + 3(A_3\alpha - B_3) + 4(B_0\alpha - A_0) + (A_1\alpha - B_1)\\
&\quad = 2(A_2\beta - B_2) + 3(B_1\alpha - A_1) + 4(A_4\beta - B_4)\\
&\qquad + (B_3\alpha - A_3).
\end{aligned} \tag{2.29}$$

If $\alpha$, $\beta$, and the constants $e_i$ satisfy Eqs. (2.29), it is then possible to calculate the functions $g_i$:

$$\begin{aligned}
g_0 &= \tfrac{1}{2}(4A_0\alpha - A_1)X^2 + [4(B_0\alpha - A_0) + (A_1\beta - B_1)]XY\\
&\quad + \tfrac{1}{2}[2(B_2\beta - B_1)]Y^2 \equiv C_0X^2 + D_0XY + E_0Y^2,\\
g_1 &= \tfrac{1}{2}(3A_1\alpha - 2A_2)X^2 + [3(B_1\alpha - A_1) + 2(A_2\beta - B_2)\\
&\quad - 4(B_0\alpha - A_0) - (A_1\beta - B_1)]XY + \tfrac{1}{2}(3B_3\beta - 2B_2)Y^2\\
&\equiv C_1X^2 + D_1XY + E_1Y^2,\\
g_2 &= \tfrac{1}{2}(2A_2\alpha - 3A_3)X^2 + [(B_3\alpha - A_3) + 4(A_4\beta - B_4)]XY\\
&\quad + \tfrac{1}{2}(4B_4\mu - B_3)Y^2 \equiv C_2X^2 + D_2XY + E_2Y^2.
\end{aligned} \tag{2.30}$$

Here $C_i, D_i, E_i$; $i = 0, 1, 2$ are complicated expressions in terms of $\alpha, \beta$ and $A_j, B_j, j = 0, 1, 2, 3, 4$.

The last compatibility relation (2.20) will give four other constraints:

$$\begin{aligned}
&\alpha C_1 - 2C_2 = 0,\\
&2(D_0\alpha - C_0) + (\beta C_1 - D_1) - 2(\alpha D_1 - C_1)\\
&\quad - 4(C_2\beta - D_2) = 0,\\
&2(D_2\beta - E_2) + (\alpha E_1 - D_1) - 2(\beta D_1 - E_1)\\
&\quad - 4(\alpha E_0 - D_0) = 0,\\
&\beta E_1 - 2E_0 = 0.
\end{aligned} \tag{2.31}$$

The nine equations (2.27), (2.29), (2.31) summarize in terms of $\alpha, \beta$, and the $e_i$'s the conditions for which the system possesses an integral of order 6 in the velocities. They form a linear, homogeneous system of nine equations for the six unknown $e_i$. It is possible to show that, in order to get a nontrivial solution to this system, $\alpha$ and $\beta$ have to take the following values (up to permutations of $\alpha$ and $\beta$).

(a) $\alpha = 2, \beta = 2$: The integral is the square of the constant $C_1$ of degree 3 found previously (2.15).

(b) $\alpha = 2, \beta = 1$: The integral is the product of the constant $C_2$ (2.24) of degree 4, found previously, with the Hamiltonian.

(c) $\alpha = \frac{2}{3}, \beta = 2$: This is a new case which corresponds to

$$m_1 = \frac{3\epsilon(2\epsilon - 1)}{2 - 3\epsilon}, \quad m_2 = 2\epsilon - 1, \quad m_3 = 1, \quad \frac{1}{2} < \epsilon < \frac{2}{3}. \tag{2.32}$$

For these values of the parameters, Bountis *et al.* found the system to be Painlevé. It is, in fact, integrable with a nondegenerate constant of the form

$$\begin{aligned}
C &= 4\dot{x}^6 + 12\dot{x}^5\dot{y} + 13\dot{x}^4\dot{y}^2 + 6\dot{x}^3\dot{y}^3 + \dot{x}^2\dot{y}^4\\
&\quad + 4(e^x - 2e^y)\dot{x}^4 + (14e^x - 16e^y)\dot{x}^3\dot{y}\\
&\quad + 10(\tfrac{5}{3}e^x - e^y)\dot{x}^2\dot{y}^2 + 2(4e^x - e^y)\dot{x}\dot{y}^3\\
&\quad + \tfrac{4}{3}e^x\dot{y}^4 + (-\tfrac{5}{3}e^{2x} + \tfrac{20}{3}e^{x+y} + 4e^{2y})\dot{y}^2\\
&\quad + (-\tfrac{8}{3}e^{2x} + 6e^{x+y} + 4e^{2y})\dot{x}\dot{y} + (-\tfrac{8}{9}e^{2x}\\
&\quad + \tfrac{4}{3}e^{x+y} + e^{2y})\dot{y}^2 + \tfrac{4}{27}e^{3x} + \tfrac{44}{9}e^{2x+y}.
\end{aligned} \tag{2.33}$$

So, every case of integrability predicted by the Painlevé analysis[8] was indeed recovered by the direct approach for the calculation of the integrals of motion.

## III. THE FIXED-END LATTICE FOR TWO PARTICLES

Let us consider a fixed-end lattice with two nonequal masses and nonequal interactions. The form of the Hamiltonian governing the system is then

$$H = \frac{p_x^2}{2m_1} + \frac{p_y^2}{2m_2} + e^{-\delta x} + e^{\epsilon(x-y)} + e^y. \tag{3.1}$$

In order to alleviate the notations, we put

$$X = e^{-\delta x}, \quad D = e^{\epsilon(x-y)}, \quad Y = e^y. \tag{3.2}$$

The equations of motion read (up to a scaling in time)

$$\ddot{x} = \delta X - \epsilon D, \quad \ddot{y} = a(\epsilon D - Y), \quad a = m_1/m_2. \tag{3.3}$$

At this point, we remark that a change

$$a' = 1/a, \quad \delta' = 1/\delta, \quad \epsilon' = \epsilon/\delta$$

leads to the same form of the equations of motion for $\xi = -\delta y, \eta = -\delta x$. These cases are then equivalent up to an $x, y$ permutation and a scaling.

As in the preceeding section, we will systematically look for a second constant of motion polynomial in the velocities. We will not burden the presentation by exposing the computations at orders 2 and 3; they did not yield any positive result.

Let us then begin with a quartic constant [form (2.16)]. As previously, it is sufficient to deal with constant coefficients $f_i$ ($f_4 = 0$).

The first compatibility condition (2.17) reduces to

$$f_1 = f_3 = 0, \tag{3.4a}$$

$$f_2(1 - a) + 2f_0 = 0. \tag{3.4b}$$

To obtain this result we use again explicitly the fact that the functions $X$ and $Y$ depend only, respectively, on $x$ and $y$.

We integrate and find the functions $g_i$:

$$\begin{aligned}
g_0 &= -4f_0(X + D) - 2af_2Y \equiv F_0(X + D) + F_2Y,\\
g_1 &= -4f_0D + 2af_2D \equiv (F_0 - F_2)D,\\
g_2 &= -2f_2X \equiv (F_2/a)X.
\end{aligned} \tag{3.5}$$

Once the relations (3.4a) and (3.4b) are fulfilled, it suffices to satisfy the second compatibility condition (2.20) for the system to possess an integral quartic in velocities. In this relation appear the independent functions of $(x, y)$ $XD$, $XY$, $DY$, $D^2$. Thus $(F_0, F_2)$ must be a nontrivial solution of the system

$(1 - a)F_2 + aF_0 = 0$ [transcription of (3.4b)],

and

$$(1-\epsilon)(aF_0-F_2)(a-2\epsilon)=0,$$
$$(3-a)F_0+(a-1)F_2=0, \qquad (3.6)$$
$$(\delta^2+2\epsilon^2-3\delta\epsilon)(F_0-F_2)=0$$

(the coefficient of $XY$ is always zero).

It is obvious that the solutions are

$$F_0=0, \quad F_2\neq 0;$$

$$a=1, \quad \epsilon=1, \quad \text{or } \epsilon=\tfrac{1}{2};$$

and

$$\delta/\epsilon=2 \quad \text{or} \quad \delta/\epsilon=1.$$

We finally find three distinct cases for which a quartic integral exists:

(a) $\dfrac{m_1}{m_2}=1, \quad \delta=\epsilon=1,$ (3.7)

(b) $\dfrac{m_1}{m_2}=1, \quad \delta=1, \quad \epsilon=\dfrac{1}{2},$ (3.8)

(e) $\dfrac{m_1}{m_2}=1, \quad \delta=\dfrac{1}{2}, \quad \epsilon=\dfrac{1}{2}$

(equivalent to $\delta=2$, $\epsilon=1$). (3.9)

[The classifications (a), (b), and (e) are those of the Introduction.]

The explicit values of the constants are

(a) $C=\dot{x}^2\dot{y}^2/2+e^y\dot{x}^2-e^{x-y}\dot{x}\dot{y}+e^{-x}\dot{y}^2$
$\qquad +e^{2(x-y)}/2+e^x+2e^{y-x}+e^{-y},$ (3.10)

(b) $C=\dot{x}^2\dot{y}^2/2+e^y\dot{x}^2-e^{(x-y)/2}\dot{x}\dot{y}+e^{-x}\dot{y}^2$
$\qquad -e^{x-y}/2+2e^{y-x},$ (3.11)

(e) $C=\dot{x}^2\dot{y}^2/2+e^y\dot{x}^2-e^{(x-y)/2}\dot{x}\dot{y}+e^{-x/2}\dot{y}^2$
$\qquad +e^{x-y}/2+2e^{y-x/2}+e^{-y/2}.$ (3.12)

In order to find the integrals for the other candidates of integrability [cases (c) and (d) of the Introduction], it is necessary, as in the preceding section, to perform the calculations at order 6. (Recall $e_6=0$.)

The first compatibility condition (2.25) easily gives a first relation between the $e_i$'s:

$$e_1=e_5=0,$$
$$E_0+(2-a)E_2+(1-a)E_3+(1-2a)E_4=0 \qquad (3.13)$$
$$(E_0=6e_0,\ E_2=2e_2,\ E_3=3e_3, E_4=2e_4).$$

The integration of the equations for the $f_i$'s is then straightforward:

$$f_0=E_0X+aE_2Y+E_0D, \quad f_1=aE_3Y+(E_0-aE_2)D,$$
$$f_2=2E_2X+[E_0+(2-a)E_2-aE_3]D+2aE_4y, \qquad (3.14)$$
$$f_3=E_3X-E_4D,\ f_4=E_4X.$$

The second compatibility condition (2.17) gives relations on the coefficients $E_i$:

$$3E_0(a-3)+E_2(9a-4-3a^2)+E_4(1-3a)=0, \qquad (3.15)$$
$$(\epsilon-1)[E_0(1-\epsilon)(1-3\epsilon)+E_2(1-\epsilon)(3a\epsilon-a-4\epsilon^2)$$
$$+\epsilon^2E_4(4a-4a\epsilon-3+4\epsilon]=0,$$

and

$$\{E_0\epsilon(5\epsilon\delta-2\delta^2-4\epsilon^2)+a\epsilon E_2[-5\epsilon\delta a+4\epsilon^2a+2a\delta^2$$
$$-4(\delta-\epsilon)^2]+a(\epsilon-\delta)^2E_4(4\epsilon a-\delta)\}(\epsilon-\delta)=0. \qquad (3.16)$$

This system [(3.13), (3.15), (3.16)] is satisfied for

$$a=1, \quad \epsilon=1, \quad \delta=1,$$
$$a=1, \quad \epsilon=\tfrac{1}{2}, \quad \delta=\tfrac{1}{2},$$
$$a=1, \quad \epsilon=1, \quad \delta=2,$$

which corresponds to the preceding cases (3.7)–(3.9).

For $a=\frac{1}{3}$, $\epsilon=\frac{1}{2}$, Eqs. (3.13) and (3.15) are satisfied and yield

$$E_2=-6E_0, \quad E_4=27E_0.$$

In order to satisfy Eq. (3.16) for these values of $a$ and $\epsilon$, $\delta$ must take one of the following values:

$$\delta=1, \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{2}.$$

Now, once the system [(3.13), (3.15), (3.16)] is satisfied, the explicit calculation of the $g_i$'s reads

$$g_0=2E_0X^2+4E_0XD+4aE_2XY+(YD/\epsilon)(4aE_2\epsilon-a^2E_2+aE_0)+(D^2/2)(4E_0-aE_0+a^2E_2)+2a^2E_4Y^2, \qquad (3.17)$$
$$g_1=\frac{D^2}{2}(7E_0-3aE_0-7aE_2+3a^2E_2)+\frac{3\delta-4\epsilon}{\epsilon-\delta}XD(aE_2-E_0)$$
$$+\frac{YD}{\epsilon}\left(-4\epsilon a^2E_4+3aE_0-3a^2E_2+4a\epsilon E_2-\frac{aE_0}{\epsilon}+\frac{a^2E_2}{\epsilon}\right),$$
$$g_2=\frac{E_4}{2}D^2-\frac{XD}{\epsilon}(E_4\delta-4a\epsilon E_4)+4aE_4XY+2E_2X^2.$$

At this point, there remains a last relation (2.20) to ensure the existence of the function $h$ of (2.25) and thus the existence of the integral at this order. We can now explicit the values of $g_i$ in (3.17) for particular cases of values of $a$, $\epsilon$, and $\delta$ solutions of (3.13), (3.15), and (3.16) and check whether relation (2.20) holds or not.

(1) $a=\frac{1}{3}$, $\epsilon=\frac{1}{2}$, $\delta=1$ [case (c) in the Introduction]. The $g_i$ are

$$g_0=\frac{2X^2}{3}+\frac{4XD}{3}-\frac{8XD}{3}-2YD+\frac{D^2}{2}+2Y^2,$$
$$g_1=3D^2+2XD-6YD, \qquad (3.18)$$
$$g_2=9D^2/2-6XD+12XY-4X^2 \quad (\text{with } 3E_0=1).$$

Relation (2.20) is verified and leads to the following value of the constant:

$$\begin{aligned}C = {} & \dot{x}^6/18 - 2\dot{x}^4\dot{y}^2 + 9\dot{x}^2\dot{y}^4 + (e^{-x} - 2e^{y} + e^{(x-y)/2})\dot{x}^4/3 + e^{(x-y)/2}\dot{x}^3\dot{y} + (6e^{y} - 3e^{(x-y)/2} - 4e^{-x})\dot{x}^2\dot{y}^2 \\ & - 9e^{(x-y)/2}\dot{x}\dot{y}^3 + 9e^{-x}\dot{y}^4 + (2e^{-2x} + 4e^{-(x+y)/2} - 8e^{y-x} - 6e^{(x+y)/2} + \tfrac{3}{2}e^{x-y} + 6e^{2y})(\dot{x}^2/3) + (3e^{x-y} \\ & + 2e^{-(x+y)/2} - 6e^{(x+y)/2})\dot{x}\dot{y} + (\tfrac{9}{2}e^{x-y} - 6e^{-(x+y)/2} + 12e^{y-x} - 4e^{-2x})\dot{y}^2 - 12e^{(y-x)/2} \\ & + 4e^{-(y+3x)/2} - 8e^{y-2x} - \tfrac{4}{3}e^{-3x} + 12e^{2y-x}. \end{aligned} \tag{3.19}$$

(2) $a = \frac{1}{3}$, $\epsilon = \frac{1}{2}$, $\delta = \frac{1}{3}$ [case (d)]. The $g_i$ are

$$\begin{aligned} g_0 &= \frac{2X^2}{3} + \frac{4XD}{3} - \frac{8XY}{3} - 2YD + \frac{D^2}{2} + 2Y^2, \\ g_1 &= 3D^2 + 6XD - 6YD, \\ g_2 &= 9D^2/2 + 6XD + 12YX - 4X^2. \end{aligned} \tag{3.20}$$

In this case also, one can easily check that relation (2.20) is verified and then compute the constant

$$\begin{aligned}C = {} & \dot{x}^6/18 - 2\dot{x}^4\dot{y}^2 + 9\dot{x}^2\dot{y}^4 + (e^{-x/3} - 2e^{y} + e^{(x-y)/2})\dot{x}^4/3 + e^{(x-y)/2}\dot{x}^3\dot{y} + (6e^{y} - 3e^{(x-y)/2} - 4e^{-x/3})\dot{x}^2\dot{y}^2 \\ & - 9e^{(x-y)/2}\dot{x}\dot{y}^3 + 9e^{-x/3}\dot{y}^4 + (2e^{-2x/3} + 4e^{x/6-y/2} - 8e^{y-x/3} - 6e^{(x+y)/2} + \tfrac{3}{2}e^{x-y} + 6e^{2y})\dot{x}^2/3 \\ & + (3e^{x-y} + 6e^{x/6-y/2} - 6e^{(x+y)/2})\dot{x}\dot{y} + (\tfrac{9}{2}e^{x-y} + 6e^{x/6-y/2} + 12e^{y-x/3} - 4e^{-2x/3})\dot{y}^2 + e^{x} - \tfrac{4}{9}e^{-x} - 4e^{2y-x/3} \\ & - 4e^{x/6+y/2} + (8e^{-2x/3+y})/3 - \tfrac{4}{3}e^{-x/6-y/2}. \end{aligned} \tag{3.21}$$

(3) A precise analysis of the values

$$a = \tfrac{1}{3}, \quad \epsilon = \tfrac{1}{2}, \quad \delta = \tfrac{2}{3},$$

and

$$a = \tfrac{1}{3}, \quad \epsilon = \tfrac{1}{2}, \quad \delta = \tfrac{1}{2}$$

shows that they do not yield any integrable case.

Exactly as in the cases of the free-end lattice, we have been able to calculate the constants of motion for the five distinct cases for which the fixed-end Toda lattice was found to possess the Painlevé property. At this point, we can remark that the leading power of the velocities of each constant of motion can be predicted from group theoretical considerations (see Sec. III). It goes without saying that such a knowledge is most helpful for the direct calculation of the integrals of motion.

## IV. INTEGRABILITY OF THE TODA CHAIN BY LIE ALGEBRA TECHNIQUES

### A. The free-end case

We consider the Hamiltonian system in the six-dimensional phase space $p_i$, $q_i$:

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i},$$

$$H = \frac{p_1^2}{2m_1} + \frac{p_2^2}{2m_2} + \frac{p_3^2}{2} + \exp[\epsilon(q_1 - q_2)] + \exp(q_2 - q_3).$$

After an appropriate canonical transformation we cast it in a form suitable for the search of integrable cases by Lie algebra theory.[9] The cases predicted by the Painlevé analysis and already found by direct calculation of the integrals will appear as the ones corresponding to simple Lie algebras of rank 2: $A_2$, $B_2$, $G_2$.

We consider the transformation, as in Sec. II,

$$q_1' = \epsilon(q_1 - q_2), \quad q_2' = q_2 - q_3, \quad q_3' = m_1 q_1 + m_2 q_2 + q_3,$$

[which we complete to a canonical one by introducing the momenta $p_1', p_2', p_3'$ satisfying $\{q_i', p_j'\} = \delta ij$ ($\{\cdots\}$ denotes the Poisson bracket relative to $q_i, p_i$ coordinates), i.e.,

$$p_1' = (1/\epsilon M)(M p_1 - m_1 P), \quad p_2' = (1/M)(P - M p_3),$$

$$p_3' = (P/\sqrt{M}),$$

where

$$P = p_1 + p_2 + p_3, \quad M = m_1 + m_2 + 1.$$

Then the reduced Hamiltonian system in the four-dimensional phase space reads

$$p_i' = -\frac{\partial H'}{\partial q_i'}, \quad q_i' = \frac{\partial H'}{\partial p_i'},$$

$$H' = \frac{1}{2}\left(\sum_{\substack{i=1,2\\ j=1,2}} a_{ij} p_i' p_j'\right) + \exp q_1' + \exp q_2',$$

where

$$a_{11} = \frac{\epsilon^2(M-1)}{m_1 m_2}, \quad a_{12} = a_{21} = -\frac{\epsilon}{m_2}, \quad a_{22} = \frac{1+m_2}{m_2}.$$

From now on, we will drop the primes in order to alleviate the notations.

The above Hamiltonian system has a Lax form representation furnishing the integrals and providing complete integrability, once the matrix $A = (a_{ij})$ is a positive scalar multiple of the matrix $((h_i, h_j))$, where $h_1$, $h_2$ constitute a root system base of a simple Lie algebra of rank 2 and $(.,.)$ denotes the scalar product determined by the Killing form. In other words, the matrix $C = (c_{ij} = 2a_{ij}/a_{jj})$ must be a rank 2 Cartan matrix.[9]

So, one has to test the values of $m_1$, $m_2$ for $C$ to become a rank 2 Cartan matrix, i.e., one of the following matrices or their transpose[14]:

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & -2 \\ -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}.$$

In each case $\epsilon$ disappears with time scaling. The cases corresponding to transpose matrices are equivalent within

TABLE I.

| Lie algebra | Cartan matrix | $m_1, m_2$ | Hamiltonian ($\epsilon$-free) | Squared time-scaling factor |
|---|---|---|---|---|
| $A_2$ | $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ | $m_1 = \frac{\epsilon(2\epsilon-1)}{2-\epsilon}$, $m_2 = 2\epsilon - 1$ | $2p_1^2 - 2p_1p_2 + 2p_2^2 + \exp q_1 + \exp q_2$ | $\frac{\epsilon}{2(2\epsilon-1)}$ |
| $B_2$ | $\begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix}$ | $m_1 = \frac{\epsilon(2\epsilon-1)}{1-\epsilon}$, $m_2 = 2\epsilon - 1$ | $p_1^2 - 2p_1p_2 + 2p_2^2 + \exp q_1 + \exp q_2$ | $\frac{\epsilon}{2(2\epsilon-1)}$ |
| $G_2$ | $\begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$ | $m_1 = \frac{3\epsilon(2\epsilon-1)}{2-3\epsilon}$, $m_2 = 2\epsilon - 1$ | $2p_1^2 - 6p_1p_2 + 6p_2^2 + \exp q_1 + \exp q_2$ | $\frac{\epsilon}{6(2\epsilon-1)}$ |

scaling of $\epsilon$ and interchanging $p_1$ with $p_2$ and $q_1$ with $q_2$. In fact, three distinct integrable cases appear (see Table I).

The Lax-pair form representation is obtained through an exact representation of minimal dimension of the corresponding Lie algebra, i.e., the Hamiltonian system is presented in the form $\dot{L} = [L, P]$, where $[L, P] = LP - PL$, $L$, $P$ belong to the representation, and they are functions of $p_i$, $q_i$.

It follows that the Hamiltonian system has integrals: $I_k = \mathrm{Tr}(L^k)$, $k = 1, 2, \ldots$ .

However we have to perform a subtle search in order to find the algebraically independent ones. In fact, the theory of polynomial invariants[15] gives the values of $k$: In the case of the Lie algebra $A_2$ the algebraically independent integrals are $I_2$ and $I_3$, for $B_2$, $I_2$, and $I_4$ and for $G_2$, $I_2$, and $I_6$. In every case $I_2$ is the Hamiltonian (up to a multiplicative constant).

For each algebra $A_2$, $B_2$, $G_2$, a suitable representation can be obtained through their correspondence to sl(3, $C$), so(5), and so(7) (see Ref. 14).

The only problem that remains is to exhibit the specific form of the Lax pair $(L, P)$ in these different cases. We will follow for this the method exposed in Ref. 9.

*Case $A_2$* (the classical Toda chain). After changing the variables $q_1$, $q_2$ to

$$l_1 = \exp \tfrac{1}{2}q_1, \quad l_2 = \exp \tfrac{1}{2}q_2,$$

the equations of motion read

$$\dot{l}_1 = l_1(2p_1 - p_2), \quad \dot{p}_1 = -l_1^2,$$
$$\dot{l}_2 = l_2(2p_2 - p_1), \quad \dot{p}_2 = -l_2^2,$$

which is a suitable form for the search of the Lax pair. Following Ref. 9 we know that the vectors

$$l(t) = l_1(t)(e_{\alpha_1} + e_{-\alpha_1}) + l_2(t)(e_{\alpha_2} + e_{-\alpha_2}) + p_1h_1 + p_2h_2,$$

$$A(l(t)) = l_1(t)(e_{\alpha_1} - e_{-\alpha_1}) + l_2(t)(e_{\alpha_2} - e_{-\alpha_2})$$

satisfy the following relation in the algebra $A_2$: $\dot{l} = [l, A(l)]$ where the vectors $(e_{\alpha_i}, h_k)$ are chosen from the basis of the algebra $A_2$. If we now use the usual matrix representation of $A_2$, namely sl(3, C), we find the form of the Lax pair:

$$L = \begin{bmatrix} p_1 & \tfrac{1}{2}l_1 & 0 \\ l_1 & p_2 - p_1 & \tfrac{1}{2}l_2 \\ 0 & l_2 & -p_2 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & \tfrac{1}{2}l_1 & 0 \\ -l_1 & 0 & \tfrac{1}{2}l_2 \\ 0 & -l_2 & 0 \end{bmatrix}.$$

The algebraically independent integrals are $I_2 = \mathrm{Tr}(L^2) = H$ and

$$I_3 = \mathrm{Tr}(L^3) = 3p_1p_2(p_1 - p_2) + \tfrac{3}{2}(p_2 l_1^2 - p_1 l_2^2).$$

*Case $B_2$*: With the same transformation, as in case $A_2$, the equations of motion are

$$\dot{l}_1 = l_1(p_1 - p_2), \quad \dot{p}_1 = -l_1^2$$

and

$$\dot{l}_2 = l_2(2p_2 - p_1), \quad \dot{p}_2 = -l_2^2.$$

Using a $5 \times 5$ matrix representation of $B_2$ we write the Lax pair:

$$L = \begin{bmatrix} p_2 & \tfrac{1}{2}l_2 & 0 & 0 & 0 \\ l_2 & p_1 - p_2 & \tfrac{1}{2}l_1 & 0 & 0 \\ 0 & l_1 & 0 & \tfrac{1}{2}l_1 & 0 \\ 0 & 0 & -l_1 & p_2 - p_1 & -\tfrac{1}{2}l_2 \\ 0 & 0 & 0 & -l_2 & -p_2 \end{bmatrix},$$

$$P = \begin{bmatrix} 0 & \tfrac{1}{2}l_2 & 0 & 0 & 0 \\ -l_2 & 0 & \tfrac{1}{2}l_1 & 0 & 0 \\ 0 & -l_1 & 0 & -\tfrac{1}{2}l_1 & 0 \\ 0 & 0 & l_1 & 0 & -\tfrac{1}{2}l_2 \\ 0 & 0 & 0 & l_2 & 0 \end{bmatrix}.$$

The algebraically independent integrals are

$$I_2 = \mathrm{Tr}(L^2) = 2H,$$

and

$$I_4 = \mathrm{Tr}(L^4) = 2(p_1 - p_2)^4 + 2p_2^4 + 2l_1^2(p_1 - p_2)^2 + 2p_2^2 l_2^2 + 2p_1^2 l_2^2 + 2l_1^2 l_2^2 + l_2^4.$$

Note that $I_3 = \mathrm{Tr}(L^3) = 0$.

*Case $G_2$:* The Hamiltonian system under the transformation $l_1 = \exp(q_1/2)$, $l_2 = \exp(q_2/2)$ reads

$$\dot{l}_1 = l_1(2p_1 - 3p_2), \quad \dot{p}_1 = -l_1^2,$$
$$\dot{l}_2 = l_2(6p_2 - 3p_1), \quad \dot{p}_2 = -l_2^2,$$

and using the $7 \times 7$ matrix representation of the exceptional Lie algebra $G_2$ we obtain the Lax pair

$$L=\begin{bmatrix} 0 & (\sqrt{2}/2)l_1 & 0 & 0 & -l_1\sqrt{2} & 0 & 0 \\ \sqrt{2}l_1 & 3p_1-2p_2 & 0 & 3l_2 & 0 & 0 & 0 \\ 0 & 0 & p_1 & 0 & 0 & 0 & -l_1/2 \\ 0 & l_2/2 & 0 & p_1-3p_2 & 0 & l_1/2 & 0 \\ -(\sqrt{2}/2)l_1 & 0 & 0 & 0 & 2p_1-3p_2 & 0 & -l_2/2 \\ 0 & 0 & 0 & l_1 & 0 & -p_1 & 0 \\ 0 & 0 & -l_1 & 0 & -3l_2 & 0 & 3p_2-p_1 \end{bmatrix},$$

$$P=\begin{bmatrix} 0 & (\sqrt{2}/2)l_1 & 0 & 0 & 0 & \sqrt{2}l_1 & 0 \\ -\sqrt{2}l_1 & 0 & 0 & 0 & 3l_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -l_1/2 \\ 0 & -l_2/2 & 0 & 0 & 0 & l_1/2 & 0 \\ -(\sqrt{2}/2)l_1 & 0 & 0 & 0 & 0 & 0 & -l_2/2 \\ 0 & 0 & 0 & -l_1 & 0 & 0 & 0 \\ 0 & 0 & l_1 & 0 & -3l_2 & 0 & 0 \end{bmatrix}.$$

The algebraically independent integrals are $I_2 = \mathrm{Tr}(L^2) = 6H$ and $I_6 = \mathrm{Tr}(L^6)$.

### B. The fixed-end case

We consider the Hamiltonian system in the four-dimensional phase space $p_i, q_i$:

$$\dot p_i = -\frac{\partial H}{\partial q_1},\quad \dot q_i = \frac{\partial H}{\partial p_i},$$

$$H=\frac{1}{2}\left(\frac{p_1^2}{m_1}+\frac{p_2^2}{m_2}\right) + \exp(-\delta q_1) + \exp\epsilon(q_1-q_2) + \exp q_2.$$

After the canonical transformation $q_i' = q_i\sqrt{m_i}, p_i' = p_i/\sqrt{m_i}$ and dropping the primes for convenience of notation, the system is written

$$\dot p_i = -\frac{\partial H}{\partial q_i},\quad \dot q_i = \frac{\partial H}{\partial p_i},$$

$$H=\frac{1}{2}(p_1^2+p_2^2) + \exp\left(-\frac{\delta}{\sqrt{m_1}}q_1\right) + \exp\epsilon\left(\frac{q_1}{\sqrt{m_1}}-\frac{q_2}{\sqrt{m_2}}\right) + \exp\frac{q_2}{\sqrt{m_2}}.$$

Considering more recent works,[12] in order to find completely integrable cases via a Lax-pair form, we have to keep in mind an extension of Bogoyavlenski's theorem presented in Ref. 9, within the framework of Kac–Moody algebras.

We will give a brief description of the concept of these algebras based on Ref. 16 so as to provide the necessary tools in order to obtain the Lax-pair forms.

So, consider a complex simple Lie algebra $g$ and an automorphism $\sigma$ of $g$ of finite order $d$ (i.e., the least positive integer $d$ such that $\sigma^d =$ identity) induced by a symmetry of the root system of $g$. The order $d$ can take the following values:

$d = 1$, $\sigma =$ identity, case of all simple Lie algebras,

$d = 2$, case of $A_n$, $n\geqslant 2$, $D_n$, $n > 4$, $E_6$,

$d = 3$, case of $D_4$.

Then one has the following decomposition of $g$ into a direct sum of subspaces: $g = g_0 + g_1 + \cdots + g_{d-1}$ indexed over the integers modulo $d$ [$g_k$ is the eigenspace of $\sigma$ corresponding to the eigenvalue $\epsilon^k$; $\epsilon = \exp(2\pi i/d)$] with the property $[g_i, g_j]\subset g_{i+j}$; $i, j, i+j$ are integers modulo $d$. Especially, $g_0$ is a simple complex Lie algebra (see Table II), and because of the relation $[g_0, g_i]\subseteq g_i$, we have a representation of $g_0$ on each $g_i$ which is irreducible. Then, the generalization of Bogoyavlenski's result consists in considering $-\theta$, the opposite of the highest weight $\theta$ of the representation of $g_0$ on $g_1$ relative to a basis $\alpha_1, ..., \alpha_n$ of roots of $g_0$ with respect to a Cartan subalgebra $\mathbf{h}_0$ of $g_0$ ($n$ is the rank of $g_0$) and the analog of Bogoyavlenski's important "admissible" set of roots is the set $\alpha_1, ..., \alpha_n, \alpha_{n+1} = -\theta$ ($\theta$ being the highest weight of the representation of $g_0$ on $g_1$, when $l_i$ are nonnegative integers with at least one of them nonvanishing, $\theta + l_1\alpha_1 + l_2\alpha_2 + \cdots + l_n\alpha_n$ is no more a weight of this representation). This is indeed a generalization, since, in case $d = 1$, $g = g_0 = g_1$, the representation is the adjoint one and the highest weight is the highest root.

Consider now, as in Ref. 9, $e_{\alpha_i}, e_{-\alpha_i}$; $i = 1, ..., n$ vectors in $g_0$, $e_{-\theta}$ in $g_{-1}$, $e_\theta$ in $g_1$, (there is a duality between the representations of $g_0$ on $g_1$ and $g_{-1}$) and a basis $h_1, ..., h_n$ of $\mathbf{h}_0$, that satisfy, among other relations

$$[e_\lambda, e_{-\lambda}] = \lambda,\quad \lambda = \alpha_1, ..., \alpha_{n+1},$$

$$[h_k, e_\lambda] = (h_k, \lambda)e_\lambda,\ [h_i, h_j] = 0,$$

where the scalar product $(x, y) = \mathrm{Tr}(adx\ ady)$, $adx(z) = [x, z]$, $z\in g$, and

$$[e_{\alpha_i}, e_{-\alpha_j}] = 0,\quad i\neq j,\quad i, j = 1, 2, ..., n+1.$$

TABLE II.

| $g$ | $g$ | $A_{2n}(n\geqslant 1)$ | $A_{2n-1}(n\geqslant 2)$ | $D_{n+1}(n>3)$ | $E_6$ | $D_4$ |
|---|---|---|---|---|---|---|
| $d$ | 1 | 2 | 2 | 2 | 2 | 3 |
| $g_0$ | $g$ | $B_n$ | $C_n$ | $B_n$ | $F_4$ | $G_2$ |

Then Theorem 1 in Ref. 9 holds, i.e., if

$$\alpha_i = \sum_{k=1}^{n} d_{ik} h_k, \quad i = 1, ..., n+1,$$

the Hamiltonian system

$$\dot{p}_i = \frac{-\partial H}{\partial q_i}, \quad \dot{q}_i = \frac{\partial H}{\partial p_i},$$

$$H = \sum_{k,m}^{n} (h_k, h_m) p_k p_m + \sum_{j}^{n+1} \exp\Big(\sum_{k}^{n} d_{jk} q_k\Big)$$

has a Lax-pair form.

We present the equation of the Lax pair into a different form due to Manakov[17]:

$$\dot{l} = [l, A(l)],$$

$$l = \sum_{k}^{n} p_k h_k + \sum_{j}^{n+1} l_j e_{\alpha_j} + \sum_{j}^{n+1} e_{-\alpha_j},$$

$$A(l) = \sum_{j}^{n+1} l_j e_{\alpha_j}, \; l_j = \exp\Big(\sum_{k}^{n} d_{jk} q_k\Big),$$

and any linear representation $T$ of $g$ determines the matrix pair $T(l)$, $T(A(l))$.

Applying this theorem in the fixed-end case of the Toda lattice with two movable particles, we obtain completely integrable cases for those values of the parameters $m_1, m_2, \delta, \epsilon$ for which the vectors

$$\beta_1 = -\frac{\delta}{\sqrt{m_1}} h_1, \quad \beta_2 = \frac{1}{\sqrt{m_2}} h_2, \quad \beta_3 = \epsilon\Big(\frac{h_1}{\sqrt{m_1}} - \frac{h_2}{\sqrt{m_2}}\Big)$$

form the "admissible" set of roots and weights of a rank 2 simple Lie algebra $g_0$ which is the $\sigma$-invariant subalgebra of a simple Lie algebra $g$ with respect to an automorphism $\sigma$ of $g$ ($h_1$, $h_2$, constitute an orthonormal basis of the Euclidean two-dimensional space).

Instead of using directly generalized Cartan matrices or extended Dynkin diagrams, we will find the values of the parameters $m_1, m_2, \delta, \epsilon$ that provide integrable cases by visualizing the "admissible" sets of roots and weights. To begin with, we exclude the $A_2^{(1)}$ system because it contains no pairs of orthogonal vectors as required by the orthogonality of $\beta_1, \beta_2$. We consider the systems $B_2^{(1)}, G_2^{(1)}, A_4^{(2)}, D_4^{(3)}$ (notation of Ref. 16) (Fig. 1). We, then, find the cases listed in Table III. Case 2 is equivalent to 2 bis by the transformation $q_1 \to -3q_2, q_2 \to -3q_1$, plus a scaling in time. Case 3 is equivalent to 3 bis by $q_1 \to -q_2/2, q_2 \to -q_1/2$ plus a scaling in time, and 5 to 5 bis by $q_1 \to -q_2, q_2 \to -q_1$.

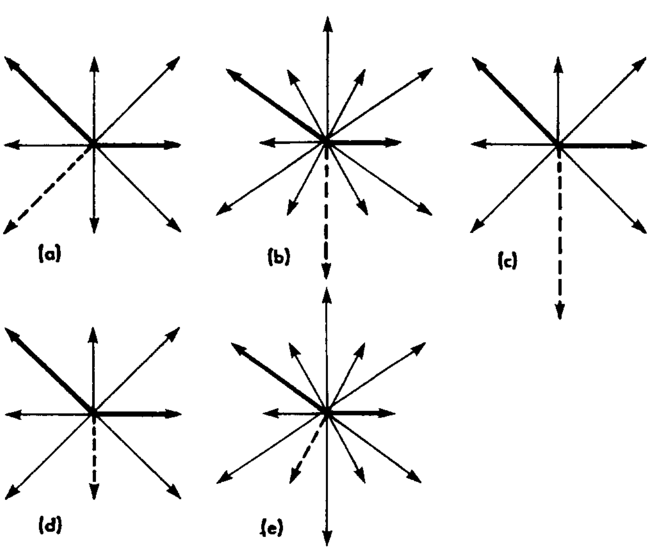


FIG. 1. Identification of a basis of roots (heavy lined arrows) and the opposite of the highest weight (dashed arrow) of the $g_0$ subalgebra of the simple Lie algebra: (a) $B_2$, (b) $G_2$, (c) $A_4$, (d) $D_3$, and (e) $D_4$.

TABLE III.

| Conditions on $\beta_1, \beta_2, \beta_3$ Values of the parameters $\epsilon, \delta, m_1/m_2$ | Hamiltonian |
|---|---|
| 1. $\lvert\beta_1\rvert = \lvert\beta_2\rvert$, $\beta_1 + \beta_2 + 2\beta_3 = 0$ <br> $\frac{1}{2}, 1, 1$ | $\frac{1}{2}(p_1^2 + p_2^2) + e^{-q_1} + e^{(q_1-q_2)/2} + e^{q_2}$ |
| 2. $3\lvert\beta_1\rvert^2 = \lvert\beta_2\rvert^2$, $3\beta_1 + \beta_2 + \beta_3 = 0$ <br> $\frac{1}{2}, \frac{1}{3}, \frac{1}{3}$ | $\frac{1}{2}(p_1^2 + p_{2/3}^2) + e^{-q_1/3} + e^{(q_1-q_2)/2} + e^{q_2}$ |
| 2 bis. $3\lvert\beta_2\rvert^2 = \lvert\beta_1\rvert^2$, $\beta_1 + 3\beta_2 + \beta_3 = 0$ <br> $\frac{3}{2}, 3, 3$ | $\frac{1}{2}(p_1^2/3 + p_2^2) + e^{-3q_1} + e^{3(q_1-q_2)/2} + e^{q_2}$ |
| 3. $2\lvert\beta_1\rvert^2 = \lvert\beta_2\rvert^2$, $2\beta_1 + \beta_2 + \beta_3 = 0$ <br> $\frac{1}{2}, \frac{1}{2}, 1$ | $\frac{1}{2}(p_1^2 + p_2^2) + e^{-q_1/2} + e^{(q_1-q_2)/2} + e^{q_2}$ |
| 3 bis. $2\lvert\beta_2\rvert^2 = \lvert\beta_1\rvert^2$, $\beta_1 + 2\beta_2 + \beta_3 = 0$ <br> $1, 2, 1$ | $\frac{1}{2}(p_1^2 + p_2^2) + e^{-2q_1} + e^{q_1-q_2} + e^{q_2}$ |
| 4. $\lvert\beta_1\rvert = \lvert\beta_2\rvert$, $\beta_1 + \beta_2 + \beta_3 = 0$ <br> $1, 1, 1$ | $\frac{1}{2}(p_1^2 + p_2^2) + e^{-q_1} + e^{q_1-q_2} + e^{q_2}$ |
| 5. $3\lvert\beta_1\rvert^2 = \lvert\beta_2\rvert^2$, $\beta_1 + \beta_2 + 2\beta_3 = 0$ <br> $\frac{1}{2}, 1, 3$ | $\frac{1}{2}(p_1^2/3 + p_2^2) + e^{-q_1} + e^{(q_1-q_2)/2} + e^{q_2}$ |
| 5 bis. $3\lvert\beta_2\rvert^2 = \lvert\beta_1\rvert^2$, $\beta_1 + \beta_2 + 2\beta_3 = 0$ <br> $\frac{1}{2}, 1, \frac{1}{3}$ | $\frac{1}{2}(p_1^2 + p_2^2/2) + e^{-q_1} + e^{(q_1-q_2)/2} + e^{q_2}$ |

Lets now give the Lax-pair form representations for each of the nonequivalent above cases and the indication of the algebraically independent integrals.

*Case $B_2^{(1)}$:* We consider the Hamiltonian

$$H = \tfrac{1}{2}(p_1^2 + p_2^2) + \exp(-2q_1) + \exp(q_1 - q_2) + \exp(2q_2)$$

(after a scaling). The equations of motion after the transformation,

$$l_1 = \exp(-2q_1), \quad l_2 = \exp(2q_2), \quad l_3 = \exp(q_1 - q_2),$$

are written

$$\dot{l}_1 = -2l_1p_1, \qquad \dot{p}_1 = 2l_1 - l_3,$$

$$\dot{l}_2 = 2l_2p_2, \qquad \dot{p}_2 = l_3 - 2l_2,$$

$$\dot{l}_3 = l_3(p_1 - p_2).$$

Using the standard representation so(5) of $B_2$ and, identifying the matrices corresponding to a basis of the roots and the opposite of the highest root, we obtain the Lax pair

$$L = \begin{vmatrix} -(p_1+p_2) & l_1 & & 2 & \\ 2 & p_1 - p_2 & l_3 & & -2 \\ & 2 & & -l_3 & \\ l_2 & & -2 & p_2 - p_1 & -l_1 \\ & -l_2 & & -2 & p_1 + p_2 \end{vmatrix},$$

$$P = \begin{vmatrix} & l_1 & & & \\ & & l_3 & & \\ & & & -l_3 & \\ l_2 & & & & -l_1 \\ & -l_2 & & & \end{vmatrix}.$$

The algebraically independent integrals are $I_2 = \mathrm{Tr}(L^2) = 8H$ and $I_4 = \mathrm{Tr}(L^4)$.

*Case* $G_2^{(1)}$: The Hamiltonian is

$$H = \tfrac{1}{4}(p_1^2/3 + p_2^2) + \exp(-6q_1) + \exp 3(q_1 - q_2) + \exp 2q_2$$

(after a scaling) and with the transformation

$$l_1 = \exp(-6q_1), \quad l_2 = \exp(2q_2), \quad l_3 = \exp 3(q_1 - q_2).$$

The equations of motion are

$$\dot l_1 = -l_1 p_1, \qquad \dot p_1 = 6l_1 - 3l_3,$$
$$\dot l_2 = l_2 p_2, \qquad \dot p_2 = 3l_3 - 2l_2,$$
$$\dot l_3 = \tfrac{1}{2} l_3(p_1 - 3p_2),$$

and the Lax pair, provided by the $7\times 7$ matrix representation of $G_2$, is

$$L = \begin{vmatrix} (p_1-p_2)/2 & l_3 & 3 & & 1 & & \\ 3 & p_2 & & l_2\sqrt{2} & & & \\ l_1 & & -(p_1+p_2)/2 & & & & -1 \\ & \sqrt{2} & & & & -l_2\sqrt{2} & \\ l_2 & & & & (p_1+p_2)/2 & & -3 \\ & & & -\sqrt{2} & & -p_2 & -l_3 \\ & & -l_2 & & -l_1 & & (p_2-p_1)/2 \end{vmatrix},$$

$$P = \begin{vmatrix} & l_3 & & & & & \\ & & & l_2\sqrt{2} & & & \\ l_1 & & & & & & \\ & & & & & -l_2\sqrt{2} & \\ l_2 & & & & & & \\ & & & & & & -l_3 \\ & & -l_2 & & -l_1 & & \end{vmatrix}.$$

The algebraically independent integrals are $I_2 = \mathrm{Tr}(L^2) = \frac{3}{2}H$ and $I_6 = \mathrm{Tr}(L^6)$.

*Case* $A_4^{(2)}$: The Hamiltonian is

$$H = \tfrac{1}{2}(p_1^2 + p_2^2) + \exp(-2q_1) + \exp(q_1 - q_2) + \exp q_2,$$

with the transformation: $l_1 = \exp(-2q_1)$, $l_2 = \exp q_2$, $l_3 = \exp(q_1 - q_2)$ the equations of motion are

$\dot l_1 = 2l_1p_1$, $\dot p_1 = 2l_1 - l_3$, $\dot l_2 = l_2p_2$, $\dot p_2 = l_3 - l_2$, $\dot l_3 = l_3(p_1 - p_2)$.

We will use the standard $5\times 5$ matrix representation sl(5) of $A_4$. An automorphism of order 2 (involution) induced by a symmetry of the root system of $A_4$ is given by

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{vmatrix} \xrightarrow{\sigma} \begin{vmatrix} -a_{55} & a_{45} & -a_{35} & a_{25} & -a_{15} \\ a_{54} & -a_{44} & a_{34} & -a_{24} & a_{14} \\ -a_{33} & a_{43} & -a_{33} & a_{23} & -a_{13} \\ a_{52} & -a_{42} & a_{32} & -a_{22} & a_{12} \\ -a_{51} & a_{41} & -a_{31} & a_{21} & -a_{11} \end{vmatrix},$$

and the matrices belonging to the $B_2$ ($\sigma$-invariant) subalgebra are easily identified as the ones of the form

$$M = \begin{vmatrix} b_{11} & b_{12} & b_{13} & b_{14} & \\ b_{21} & b_{22} & b_{23} & & b_{14} \\ b_{31} & b_{32} & & b_{23} & -b_{13} \\ b_{41} & & b_{32} & -b_{22} & b_{12} \\ & b_{41} & -b_{31} & b_{21} & -b_{11} \end{vmatrix},$$

and the one-dimensional subspace of the matrices corresponding to the opposite of the highest weight of the representation of $g_0$ on $g_1$ are the matrices $(a_{ij})$ $1 < i,j \leqslant 6$ with all entries 0 except $a_{51}$.

So a Lax-pair for the considered Hamiltonian system is

$$L = \begin{vmatrix} p_1 & l_3 & & & 2 \\ 1 & p_2 & l_2 & & \\ & 1 & & l_2 & \\ & & 1 & -p_2 & l_3 \\ l_1 & & & 1 & -p_1 \end{vmatrix},$$

$$P = \begin{vmatrix} & l_3 & & & \\ & & l_2 & & \\ & & & l_2 & \\ & & & & l_3 \\ l_1 & & & & \end{vmatrix}.$$

The algebraically independent integrals are $I_2 = \mathrm{Tr}(L^2) = 4H$ and $I_4 = \mathrm{Tr}(L^4)$.

*Case* $D_3^{(2)}$: The Hamiltonian is

$$H = p_1^2 + p_2^2 + \exp(-q_1) + \exp(q_1 - q_2) + \exp q_2,$$

with

$$l_1 = \exp(-q_1), \quad l_2 = \exp q_2, \quad l_3 = \exp(q_1 - q_2).$$

The equations of motion are

$$\begin{aligned} \dot{l}_1 &= -2l_1 p_1, & \dot{p}_1 &= l_1 - l_3, \\ \dot{l}_2 &= 2l_2 p_2, & \dot{p}_2 &= l_3 - l_2, \\ \dot{l}_3 &= 2l_3(p_1 - p_2). \end{aligned}$$

We use the $4\times 4$ standard representation sl(4) of the $D_3 \simeq A_3$ with the automorphism of order 2,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} \xrightarrow{\sigma} \begin{vmatrix} -a_{44} & a_{34} & -a_{24} & a_{14} \\ a_{43} & -a_{33} & a_{23} & -a_{13} \\ -a_{42} & a_{32} & -a_{22} & a_{12} \\ a_{41} & -a_{31} & a_{21} & -a_{11} \end{vmatrix},$$

to identify the matrices-elements of the $\sigma$-invariant subalgebra of type $B_2$. These matrices have the following form:

$$\begin{vmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & -b_{13} \\ b_{31} & b_{32} & -b_{22} & b_{12} \\ b_{41} & -b_{31} & b_{23} & -b_{11} \end{vmatrix},$$

and the one-dimensional space of matrices corresponding to the opposite of the highest weight are the matrices $(a_{ij})$; $1 \leqslant i, j \leqslant 4$ with the entries vanishing except $a_{31} = \alpha_{42}$.

The Lax pair for the Hamiltonian system is

$$L = \begin{vmatrix} -p_1 - p_2 & l_1 & 1 & \\ 1 & p_1 - p_2 & l_3 & 1 \\ l_2 & 2 & p_2 - p_1 & l_1 \\ & l_2 & 1 & p_1 + p_2 \end{vmatrix},$$

$$P = \begin{vmatrix} & l_1 & & \\ & & l_3 & \\ l_2 & & & l_1 \\ & l_2 & & \end{vmatrix}.$$

The algebraically independent integrals are $I_2 = \mathrm{Tr}(L^2) = 4H$ and $I_4 = \mathrm{Tr}(L^4)$.

*Case* $D_4^{(3)}$: The Hamiltonian is

$$H = \tfrac{1}{4}(3p_1^2 + p_2^2) + \exp(-2q_1) + \exp(q_1 - q_2) + \exp 2q_2,$$

and the equations of motion are

$$\begin{aligned} \dot{l}_1 &= 3l_1 p_1, & \dot{p}_1 &= 2l_1 - l_3, \\ \dot{l}_2 &= l_2 p_2, & \dot{p}_2 &= l_3 - 2l_2, \\ \dot{l}_3 &= \tfrac{1}{2} l_3(3p_1 - p_2). \end{aligned}$$

We consider the $8\times 8$ standard representation so(8) of $D_4$ with the automorphism $\sigma$ of order 3:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & 14 & -b_{11} & -b_{12} & -b_{13} & \\ a_{21} & a_{22} & a_{23} & a_{24} & -b_{21} & -b_{22} & & b_{13} \\ a_{31} & a_{32} & a_{33} & a_{34} & -b_{31} & & b_{22} & b_{12} \\ a_{41} & a_{42} & a_{43} & a_{44} & & b_{31} & b_{21} & b_{11} \\ -c_{11} & -c_{12} & -c_{13} & & -a_{44} & -a_{34} & -a_{24} & -a_{14} \\ -c_{21} & -c_{22} & & c_{13} & -a_{43} & -a_{33} & -a_{23} & -a_{13} \\ -c_{31} & & c_{22} & c_{12} & -a_{42} & -a_{32} & -a_{22} & -a_{12} \\ & c_{31} & c_{21} & c_{11} & -a_{41} & -a_{31} & -a_{21} & -a_{11} \end{vmatrix}$$

$$\xrightarrow{\sigma} \begin{vmatrix} a'_{11} & b_{31} & -b_{21} & b_{11} & -b_{22} & -b_{12} & -b_{13} & \\ c_{13} & a'_{22} & a_{23} & -a_{13} & -a_{24} & -a_{14} & & b_{13} \\ -c_{12} & a_{32} & a'_{33} & a_{12} & -a_{34} & & a_{14} & b_{12} \\ c_{11} & -a_{31} & a_{21} & a'_{44} & & a_{34} & a_{24} & b_{22} \\ -c_{22} & -a_{42} & -a_{43} & & -a'_{44} & -a_{12} & a_{13} & -b_{11} \\ -c_{21} & -a_{41} & & a_{43} & -a_{21} & -a'_{33} & -a_{23} & b_{21} \\ -c_{31} & & a_{41} & a_{12} & a_{31} & -a_{32} & -a'_{22} & -b_{31} \\ & c_{31} & c_{21} & c_{22} & -c_{11} & b_{12} & -b_{13} & -a'_{11} \end{vmatrix},$$

where $a'_{11} = \tfrac{1}{2}(a_{11} + a_{22} + a_{33} - a_{44})$, $a'_{22} = \tfrac{1}{2}(a_{11} + a_{22} - a_{33} + a_{44})$, $a'_{33} = \tfrac{1}{2}(a_{11} - a_{22} + a_{33} + a_{44})$, and $a'_{44} = -\tfrac{1}{2}(-a_{11} + a_{22} + a_{33} + a_{44})$. Following Ref. 16 we exhibit corresponding to the four basic roots of $D_4$, i.e., matrix $X_1$, $X_2$, $X_3$, and $X_4$, with nonvanishing elements of the matrix $a_{12}$, $a_{23}$, $a_{34}$, and $b_{31}$, respectively. Then, the matrices corresponding to basic roots in the $\sigma$-invariant subalgebra of type $G_2$ are $X_2$ and $X_1 + X_3 + X_4$, and the one-dimensional subspace corresponding to the highest weight is spanned by $[X_1,[X_2,X_3]] + \epsilon[X_3,[X_2,X_4]] + \epsilon^2[X_4,[X_2,X_1]]$ (so the space corresponding to the opposite of the highest weight is spanned by the transpose of this matrix) where $\epsilon = \exp(2\, i/3)$. Finally the Lax pair for the Hamiltonian system is

$$L = \left|\begin{array}{cccccccc} -p_2 & l_2 & & 1 & -\epsilon & & & \\ 1 & (-p_2-3p_1)/2 & l_1 & & & -\epsilon^2 & & \\ & 3 & (3p_1-p_2)/2 & l_2 & -l_2 & \epsilon^2 & & \\ l_3 & & 1 & & & l_2 & & \epsilon \\ -\epsilon^2 l_3 & & -1 & & -l_2 & & & -1 \\ & -\epsilon l_3 & & 1 & -1 & (p_2-3p_1)/2 & -l_1 & \\ & & \epsilon l_3 & & & & (p_2+3p_1)/2 & -l_2 \\ & & & \epsilon^2 l_2 & -l_3 & & -1 & p_2 \end{array}\right|,$$

$$P = \left|\begin{array}{cccccccc} & l_2 & & & & & & \\ & & l_1 & & & & & \\ & & & l_2 & -l_2 & & & \\ l_3 & & & & & l_2 & & \\ -\epsilon^2 l_3 & & & & & -l_2 & & \\ & -\epsilon l_3 & & & & & -l_1 & \\ & & -\epsilon l_3 & & & & & -l_2 \\ & & & \epsilon^2 l_3 & -l_3 & & & \end{array}\right|.$$

The algebraically independent integrals are

$$I_2 = \mathrm{Tr}(L^2) = 12\,H \quad \text{and} \quad I_6 = \mathrm{Tr}(L^6).$$

## V. CONCLUSION

In this work, we have presented some new integrable cases for systems which are restrictions of the Toda lattice to chains of two or three particles, interacting exponentially with their nearest neighbors. This restriction of dimensionality leads to new integrable cases as it allows the choice of unequal masses and unequal ranges of interaction. The Painlevé analysis of Bountis *et al.* and of Ramani has identified the different values of parameters for which the above systems could possess a second integral of motion apart from the energy. The integrability of all these cases has been demonstrated in this work through two distinct approaches: direct computation of the integrals of motion and group theoretical methods. The first method, whenever practicable, allows at the same time the identification of the integrable cases and the exact calculation of the integrals of motion. The Lie algebra approach can be used independently from the direct search. It allows easily the identification of integrable cases and can, in principle, be used for the computation of the constant of motion although the latter calculation becomes sometimes pretty involved.

What emerges from our work is the use of different, independent approaches, such as Painlevé analysis, direct search for the constant of motion, and group theoretical methods, can be a most powerful tool for the investigation of integrability of dynamical systems.


## ACKNOWLEDGMENT

The authors are most grateful to Alfred Ramani for numerous fruitful discussions.



[1]M. Toda, J. Phys. Soc. Jpn. **23**, 501 (1967); Progr. Theor. Phys. Suppl. **45**, 174 (1970).

[2]M. Henon, Phys. Rev. B **9**, 1921 (1974).

[3]H. Flaschka, Phys. Rev. B **9**, 1924 (1974); Progr. Theor. Phys. **51**, 703 (1974).

[4]S. V. Manakov, Zh. Eksp. Teor. Fiz. **67**, 543 (1974) [Sov. Phys. JETP **40**, 269 (1975)].

[5]J. Moser, Adv. Math. **16**, 197 (1975).

[6]M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978).

[7]A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 1539 (1982).

[8]T. Bountis, H. Segur, and F. Vivaldi, Phys. Rev. A **25**, 1257 (1982).

[9]O. I. Bogoyavlenski, Commun. Math. Phys. **51**, 201 (1976).

[10]M. A. Olshanetsky and A. M. Perelomov, Invent. Math. **54**, 261 (1979).

[11]A. Ramani (private communication).

[12]M. Adler and P. van Moerbeke, Commun. Math. Phys. **83**, 83 (1982).

[13]B. Grammaticos, B. Dorizzi, and A. Ramani, J. Math. Phys. **24**, 2289 (1979).

[14]J. Humpheys, *Introduction to Lie Algebras and Representation Theory* (Springer, New York, 1972).

[15]C. Chevalley, Am. J. Math. **77**, 778 (1955).

[16]S. Helgason, *Differential Geometry, Lie Groups and Symmetric Spaces* (Academic, New York, 1978).

[17]S. V. Manakov, Funct. Anal. Appl. **10**, 93 (1976).

# The Liouville–Bäcklund transformation for the two-dimensional SU(N) Toda lattice

Ziemowit Popowicz[a)]
*Department of Physics, Brookhaven National Laboratory, Upton, New York 11973*

(Received 25 October 1983; accepted for publication 20 January 1984)

We describe the Liouville–Bäcklund transformation for the two-dimensional SU($N$) Toda lattice with free end points. Integration of this transformation gives us the general solution of this equation, which depends on the $N$ arbitrary solutions of the two-dimensional Laplace equation.

PACS numbers: 03.20. + i, 02.90. + p

## I. INTRODUCTION

The last decade has shown the exciting prospect of tackling the classical solutions for the Yang–Mills field theory for the different gauge groups. The well-known Belavin *et al.* instanton solution for the SU(2) gauge group[1] has been extended by Witten[2] to the spherically symmetrical instantons solutions for the same group. Next, Leznov and Saveliev[3] generalized the Witten construction to the arbitrary compact gauge group. Their construction, more precisely the self-dual equations for the SU($N$) gauge group, is reduced to two-dimensional Toda lattice with free end points.

On the other hand, the one-dimensional periodic Toda lattice has been extensively studied in the last decade[4] by many authors. It was shown that this system describes a completely integrable Hamiltonian system and can be solved by the inverse scattering transformation[5] or by Bäcklund transformations.[6] The one-dimensional Toda lattice with free end points was considered by Kostant[7] and by Olshanetsky and Perelomov.[8]

Moreover, there were proposed several different kinds of generalizations[9–11] of the Toda lattice. Here we will consider those proposed by Leznov and Saveliev, which we call the SU($N$) Toda lattice with the free end points in the two-dimensional space-time, [hereafter referred to as the SU($N$) Toda lattice]. We will use a slightly different terminology than that used in the Yang–Mills field theory. Our SU($N$) Toda lattice corresponds to the SU($N+1$) spherically symmetrical instanton solutions.

The two-dimensional periodic Toda lattice has been solved by Mikhajlov[12] by the inverse scattering transformation and by Fordy and Gibbon[13] by the Bäcklund transformation. For the SU($N$) Toda lattice, Leznov and Saveliev proposed two different methods for the solutions.[14–16] One of them uses the representation theory of the compact group. The second is pure algebraic and uses the differential calculus only. In both cases they obtained the closed formulas on the solutions of the SU($N$) Toda lattice as a functional of $N$ arbitrary solutions of the two-dimensional Laplace equations.

On the other hand the SU($N$) Toda lattice for $N=1$ reduces to the Liouville equation for which there is known a Bäcklund transformation which relates this equation to the two-dimensional Laplace equation. In this paper we generalize the Bäcklund transformation to arbitrary $N$. This transformation we will call the Liouville–Bäcklund transformation in order to distinguish it from the Bäcklund transformation for the periodic Toda lattice found by Fordy and Gibbon.[13] Our transformation joins $N$ arbitrary solutions of the two-dimensional Laplace equations with the SU($N$) Toda lattice. Moreover, this transformation contains one arbitrary constant. Integrating transformation, we obtain the solutions of the SU($N$) Toda lattice which can be reduced to those proposed by Leznov.[14] Therefore, we establish the correspondence between the Liouville–Bäcklund transformation method and with Leznov's method.

The paper is organized as follows. In the second section we describe a method of finding the Liouville–Bäcklund transformation for the Liouville equation which is different from that proposed by Lamb.[17] From this we find the Liouville–Bäcklund transformation first for the SU(2) case, which is described in the third section and then for arbitrary $N$ which is described in the fourth section. The last section contains concluding remarks.

## II. THE LIOUVILLE–BÄCKLUND TRANSFORMATION FOR THE LIOUVILLE EQUATION

In the last century, Liouville[17] found the solution of the nonlinear partial differential equation

$$h_{z\bar{z}} = \frac{\partial^2}{\partial z\, \partial \bar{z}} h = e^{2h}, \tag{1}$$

where $z = x + it$ and $\bar{z} = x - it$, depending on two arbitrary functions

$$e^{2h} = f_z\, g_{\bar{z}} (f+g)^{-2}, \tag{2}$$

where $f = f(z)$ and $g = (\bar{z})$ are arbitrary functions of their arguments. In order to check formula (2), let us assume that

$$h_z = Ae^h - F_z, \tag{3}$$

$$h_{\bar{z}} = Be^h - G_{\bar{z}}, \tag{4}$$

where $F = F(z)$ and $G = G(\bar{z})$ are arbitrary functions of their arguments and $A$ and $B$ are unknown functions which we want to find. We can determine these functions from the integrability conditions and from the assumption that $h$ satisfies Liouville equation. These two assumptions give us

$$A_{\bar{z}} - AG_{\bar{z}} = B_z - BF_z, \tag{5}$$

$$AB = 1 - (A_{\bar{z}} - AG_{\bar{z}})\, e^{-h}. \tag{6}$$

In order to find the Bäcklund transformation, let us assume

$$A_{\bar{z}} - AG_{\bar{z}} = 0. \tag{7}$$

[a)] Permanent address: Institute of Theoretical Physics, University of Wroclaw, ul. Cybulskiego 36, 50-205 Wroclaw, Poland.

  

Solving Eqs. (7) and (5), and introducing the solution to (3) and (4), we obtain

$$h_z = a\, e^{h+G-F} - F_z, \tag{8}$$

$$h_{\bar z} = (1/a)\, e^{h+F-G} - G_{\bar z}. \tag{9}$$

This is our Bäcklund transformation. Here $a$ is an arbitrary parameter different from zero. Now we can integrate Eqs. (8) and (9), and we obtain formula (2) in which

$$a \int_\infty^z e^{-2F}\, dz' = -f, \tag{10}$$

$$\frac{1}{a}\int_\infty^{\bar z} e^{-2G}\, d\bar z' = -g. \tag{11}$$

## III. THE LIOUVILLE-BÄCKLUND TRANSFORMATION FOR THE SU(2) TODA LATTICE

Let us consider the following generalization of the Liouville equation which we call the SU(2) Toda lattice:

$$h_{1z\bar z} = \exp(2h_1 - h_2), \tag{12}$$

$$h_{2z\bar z} = \exp(2h_2 - h_1). \tag{13}$$

Let us assume similarly as in the previous case that we have the following form for the derivative of $h_1$:

$$\partial_z(h_1 + \phi') = Ae^{h_1}, \tag{14}$$

$$\partial_{\bar z}(h_1 + \gamma') = Be^{h_1}, \tag{15}$$

where $\phi' = \phi'(z)$ and $\gamma' = \gamma'(\bar z)$ are arbitrary functions of their arguments and $A$ and $B$ are unknown functions which we determine from the integrability conditions and from the assumption that $h_1$ satisfies Eq. (12). The integrability condition with (12) gives us

$$X = \partial_{\bar z}A - A\gamma'_{\bar z} = \partial_z B - B\phi'_z, \tag{16}$$

$$AB = e^{-h_2} - Xe^{-h_1}. \tag{17}$$

On the other hand, the formula (17) can be computed directly from (12) and (14), (15). Indeed introducing $e^{-h_1} = H$, we obtain

$$e^{-h_2} = H_z H_{\bar z} - HH_{z\bar z}; \tag{18}$$

then, computing $H_z$, $H_{\bar z}$, and $H_{z\bar z}$ with the help of (14) and (15), we obtain (17). Now with the help of (17) or (18) and (14), (15), we can compute the derivatives $\partial_z h_2$ and $\partial_{\bar z} h_2$:

$$\partial_z(h_2 + \ln(X) + \phi') = -(\partial_z A\cdot B - AB\partial_z \ln(X))\, e^{h_2}, \tag{19}$$

$$\partial_{\bar z}(h_2 + \ln(X) + \gamma') = -(A\cdot\partial_{\bar z}B - AB\partial_{\bar z}\ln(X))\, e^{h_2}. \tag{20}$$

Let us assume that

$$\partial_{z\bar z} \ln X = 0. \tag{21}$$

This assumption is a purely heuristic assumption, which can be motivated, that we would like to consider the symmetric form of (19) and (20) to the formula (14) and (15). As we show this assumption does not contradict either the integrability of (19) and (20) or the assumption that $h_2$ satisfies (13). Indeed introducing $\exp(-h_2) = G$ we have

$$e^{-h_1} = G_z G_{\bar z} - GG_{z\bar z}. \tag{22}$$

Next differentiating Eqs. (19) and (20) with respect to $\bar z$ and $z$, respectively, and computing $G_z$, $G_{\bar z}$, and $G_{z\bar z}$ with the help of (19) and (20), we easily check the integrability of (19) and (20), and next we easily recognize that $h_2$ indeed satisfies Eq. (13).

Now we can easily solve Eq. (21) which gives us, with the help of (16),

$$A = \left\{\int^{\bar z} e^{\gamma^2-\gamma'}\, d\bar z' + \chi(z)\right\} e^{\phi^2+\gamma'}, \tag{23}$$

$$B = \left\{\int^{z} e^{\phi^2-\phi'}\, dz' + \chi(\bar z)\right\} e^{\gamma^2+\phi'}, \tag{24}$$

$$X = e^{\phi^2+\gamma^2} = F(z)\, K(\bar z), \tag{25}$$

where $\phi^2 = \phi^2(z)$ and $\gamma^2 = \gamma^2(\bar z)$ are the arbitrary functions of the arguments. Here we use the special notation on the function $X$ which we will use in the next sections. The functions $\chi(z)$ and $\chi(\bar z)$ are unknown functions which we determine in the following manner. Substituting the formula (18) into (22), we obtain

$$1 = -\det\begin{pmatrix} H, & H_z, & H_{zz} \\ H_{\bar z}, & H_{z\bar z}, & H_{zz\bar z} \\ H_{\bar z\bar z}, & H_{\bar z\bar z z}, & H_{zz\bar z\bar z}\end{pmatrix}. \tag{26}$$

Using Eqs. (14) and (15), Eq. (26) reduces to

$$-1 = \det\begin{pmatrix} 0, & -A, & -A_z \\ -B, & X, & X_z \\ -B_{\bar z}, & X_{\bar z}, & X_{z\bar z}\end{pmatrix}. \tag{27}$$

Introducing (23), (24), (25)–(27) after algebraic manipulations, we obtain

$$\partial_{\bar z}\chi(\bar z)\, \partial_z\chi(z) = \exp(-2\phi^2 - 2\gamma^2 - \phi' - \gamma'). \tag{28}$$

Equation (28) we solve by separation of variables, which gives us

$$\chi(z) = \mu\int^z e^{-\phi'-2\phi^2}\, dz', \tag{29}$$

$$\chi(\bar z) = \frac{1}{\mu}\int^{\bar z} e^{-\gamma'-2\gamma^2}\, d\bar z'. \tag{30}$$

Here $\mu$ is the arbitrary nonzero separation constant. By introducing (23) and (24) with (29) and (30) to (12), (13) and (19), (20) these equations become the Liouville–Bäcklund transformation. Carrying out the integration of this transformation, we obtain the general form of the solutions of the SU(2) Toda lattice. These are

$$\begin{aligned} e^{-h_1} = -e^{\gamma'+\phi'}\Bigg(&\int^z e^{\phi^2-\phi'}\, dz'\, \mu\int^{z'} e^{-\phi'-2\phi^2}\, dz'' \\ &+ \int^{\bar z} e^{\gamma^2-\gamma'}\, d\bar z'\, \frac{1}{\mu}\int^{\bar z'} e^{-\gamma'-2\gamma^2}\, d\bar z'' \\ &+ \int^{\bar z} e^{\gamma^2-\gamma'}\, d\bar z' \int^z e^{\phi^2-\phi'}\, d\bar z'\Bigg). \end{aligned} \tag{31}$$

$e^{-h_2}$ can be computed by formula (18).

In this way our solutions depends on the two arbitrary solutions of the two-dimensional Laplace's equations and on one arbitrary constant different from zero. These solutions can be reduced to those proposed by Leznov.[14]

## IV. THE LIOUVILLE-BÄCKLUND TRANSFORMATION FOR THE SU(N) TODA LATTICE

Let us consider a more complicated generalization of the Liouville equation, which we call the SU($N$) Toda lattice, proposed by Leznov and Saveliev,[3]

$$h_{1z\bar z} = \exp(2h_1 - h_2), \tag{32}$$

$$h_{2z\bar z} = \exp(-h_1 + 2h_2 - h_3), \tag{33}$$

$$h_{\alpha z\bar z} = \exp(-h_{\alpha-1} + 2h_\alpha - h_{\alpha+1}), \tag{34}$$

$$h_{Nz\bar z} = \exp(-h_{N-1} + 2h_N). \tag{35}$$

In order to see the connection of Eqs. (32)–(35) with the Toda lattice, let us write down the equation of motion for the two-dimensional Toda lattice in the following form[6,16]:

$$\frac{\partial^2}{\partial z\,\partial\bar z} g_\alpha = \sum_{\beta=1}^{N} K_{\alpha\beta} \exp g_\beta. \tag{36}$$

Here $K = \{K_{\alpha\beta}\}$ is the Cartan matrix for the SU$(N+1)$ group and has the following form:

$$K = \begin{pmatrix} 2, & -1, & 0, & 0, & \cdot \\ -1, & 2, & -1, & 0, & \cdot \\ 0, & -1, & 2, & -1, & \cdot \\ & & \cdots & & \end{pmatrix}. \tag{37}$$

Assuming that $g_0 = g_{N+1} = -\infty$ and transforming $g_\alpha$ to $g_\alpha = \Sigma_{\beta=1}^{N} K_{\alpha\beta}\, h_\beta$ Eqs. (36) reduce to Eqs. (32)–(35).

Now let us introduce the following notation, $e^{-h_1} = H$. Then, as one can easily find, we have

$$\exp(-h_2) = H_z H_{\bar z} - H H_{z\bar z} = -\det{}_2(H). \tag{38}$$

Using Eq. (33), we find

$$-\exp(-h_3) = \det\begin{pmatrix} H, & H_z, & H_{zz} \\ H_{\bar z}, & H_{\bar z z}, & H_{\bar z zz} \\ H_{\bar z\bar z}, & H_{\bar z\bar z z}, & H_{\bar z\bar z zz} \end{pmatrix} = \det{}_3(H), \tag{39}$$

and in the general case

$$\exp(-h_\alpha) = (-1)^{\alpha(\alpha-1)/2} \det{}_\alpha(H), \quad 1 \leqslant \alpha \leqslant N, \tag{40}$$

$$\exp(-h_{N+1}) = (-1)^{N(N+1)/2} \det{}_{N+1}(H) = 1. \tag{41}$$

Now let us assume, as in the previous sections, that

$$h_{1z} = A\, e^{h_1} - \phi^1_z, \tag{42}$$

$$h_{1\bar z} = B\, e^{h_1} - \gamma^1_{\bar z}, \tag{43}$$

where $\phi^1 = \phi^1(z)$ and $\gamma^1 = \gamma^1(\bar z)$ are arbitrary functions of their arguments and $A$ and $B$ are unknown functions which we want to determine.

Due to the formulas (38)–(41) and the assumptions (42) and (43), we can write down $\exp(-h_\alpha)$ as a functional of $H$, $A$, $B$, $\phi'$, $\gamma'$, namely, we have

$$\exp(-h_2) = HX - \det\begin{pmatrix} 0, & -A \\ -B, & -A_{\bar z} \end{pmatrix}, \tag{44}$$

$$\exp(-h_3) = -H\det\begin{pmatrix} X, & X_z \\ X_{\bar z}, & X_{z\bar z} \end{pmatrix} - \det\begin{pmatrix} 0, & -A, & -A_z \\ -B, & X, & X_z \\ -B_{\bar z}, & X_{\bar z}, & X_{\bar z z} \end{pmatrix} = -H\det{}_2(X) - \det{}_3(A,B,X). \tag{45}$$

For arbitrary $\alpha$, $1 < \alpha \leqslant N$, we have

$$\exp(-h_\alpha) = (-1)^{\alpha(\alpha-1)/2} H \det{}_{\alpha-1}(X) - \det{}_\alpha(A,B,X), \tag{46}$$

where we use the following notation:

$$X = A_{\bar z} - \gamma'_{\bar z} A. \tag{47}$$

We can expand $\exp(-h_2)$ in the slightly different form also using the following formula:

$$\exp(-h_2) = H\det\begin{pmatrix} 1, & 0 \\ -B, & -B_z - B\phi'_z \end{pmatrix} - \det\begin{pmatrix} 0, & -A \\ -B, & -B_z \end{pmatrix}. \tag{48}$$

Here we use Eq. (43) instead of (42). Then comparison of (44) with (48) gives us

$$X = B_z - B\phi'_z. \tag{49}$$

Equations (44) and (47) guarantee us the integrability of (42) and (43). To obtain the explicit form for the derivatives of the $h_\alpha$, $\alpha > 1$, let us differentiate (46) with respect to $z$ and $\bar z$, respectively, and use (42), (43), and (46) again, obtaining

$$\begin{aligned}(h_\alpha + \ln\det{}_{\alpha-1}(X) + \phi')_z &= (-1)^{\alpha(\alpha-1)/2} \times A \det{}_{\alpha-1}(X)\, e^{h_\alpha} \\ &\quad - (\phi' + \ln\det{}_{\alpha-1}(x))_z \det{}_\alpha(A,B,X) \\ &\quad \times e^{h_\alpha} + \partial_z \det(A,B,X)\, e^{h_\alpha},\end{aligned} \tag{50}$$

$$\begin{aligned}(h_\alpha + \ln\det{}_{\alpha-1}(X) + \gamma')_{\bar z} &= (-1)^{\alpha(\alpha-1)/2} \times B \det{}_{\alpha-1}(X)\, e^{h_\alpha} \\ &\quad - (\gamma' + \ln\det{}_{\alpha-1}(X))_{\bar z} \det{}_\alpha(A,B,X)\, e^{h_\alpha} \\ &\quad + \partial_{\bar z} \det(A,B,X)\, e^{h_\alpha}.\end{aligned} \tag{51}$$

It will be very useful for us to introduce the special notation for the derivative of $h_N$

$$(h_N + \ln\det{}_{N-1}(X) + \phi')_z = C_N\, e^{h_N}, \tag{52}$$

$$(h_N + \ln\det{}_{N-1}(X) + \gamma')_{\bar z} = D_N\, e^{h_N}, \tag{53}$$

where $C_N$ and $D_N$ can be computed from (50) and (51), respectively.

We are now prepared to find the equation from which we determine the functions $A$ and $B$. First, as one can easily notice, it is possible to define $\exp(-h_\alpha)$ successively as a functional of $\exp(-h_N)$ also, in the reverse order to (38)–(41). Indeed, introducing $\exp(-h_N) = G$, we obtain

$$\exp(-h_{N-1}) = G_z G_{\bar z} - G G_{z\bar z} = -\det{}_2(G), \tag{54}$$

$$\exp(-h_\alpha) = (-1)^{\alpha(\alpha-1)/2} \det{}_\alpha(G). \tag{55}$$

As in Sec. III, we would like to have the derivative of $h_N$ a symmetrical form to the derivative of $h_1$. Therefore, we assume it and that it can be denoted by

$$(h_N + \phi + \phi')_z = C'_N\, e^{h_N}, \tag{56}$$

$$(h_N + \gamma + \phi')_{\bar z} = D'_N\, e^{h_N}, \tag{57}$$

where $\phi = \phi(z)$ and $\gamma = \gamma(\bar z)$ are the arbitrary functions of their arguments. Moreover, we assume that

$$\phi_z = \partial_z \ln\det{}_{N-1}(X) \tag{58}$$

$$\gamma_{\bar z} = \partial_{\bar z} \ln\det{}_{N+1}(X) \tag{59}$$

$$C'_N = C_N, \quad D'_N = D_N, \tag{60}$$

$$\phi = \sum_{i=2}^{N} \phi^i(z), \quad \gamma = \sum_{i=2}^{N} \gamma^i(\bar z). \tag{61}$$

We can assume that $\phi$ and $\gamma$ have the form (61) because $\phi$ and $\gamma$ are arbitrary functions. As we show, the assumptions (58)–(61) do not contradict the integrability of (56) and (57). Indeed differentiating (56) with respect to $\bar z$ and using (57), (59), and (54), one can check that $h_N$ satisfies (35). Now we can do the same with Eq. (57) and obtain that $h_N$ satisfies (35) again. Therefore, we prove the integrability of (52) and (53). Because $h_{N-\alpha}$ is the function of $h_N$ or $h_1$, we immediately conclude that the integrability of (50) and (51) is the direct consequence of the integrability of $h_N$ and $h_1$. Equations (50) and (51) define for us the Liouville–Bäcklund transformation for the arbitrary $N$ in the SU($N$) Toda lattice. To obtain the explicit formulas on this transformation, we should find the functions $A$ and $B$. Equations (59), (60), (47), (49), and (41) are our basic equations from which we find these functions.

Preparing the first integrations of the (58) and (59), we find

$$\det{}_{N-1}(X) = \exp\Big(\sum_{i=2}^{N} \phi^i + \sum_{i=2}^{N} \gamma^i\Big). \tag{62}$$

In this way we obtain the similar but not identical equation [Eq. (62)] to that found by Leznov.[14] We solve it in a similar manner to Leznov. Namely, we assume that

$$X = \sum_{\alpha=1}^{N-1} F^\alpha(z)\cdot K^\alpha(\bar z); \tag{63}$$

then (62) becomes

$$\det{}_{N-1}(F)\cdot\det{}_{N-1}(K) = e^{\phi+\gamma}, \tag{64}$$

where

$$F_{ij} = F^i_{\underbrace{z,z,\ldots,z}_{j-1}}, \quad K_{ij} = K^j_{\underbrace{\bar z,\bar z,\ldots,\bar z}_{i-1}}.$$

Let us now assume by induction that the functions $F_{N-2}$ and $K_{N-2}$, $1 \leqslant \alpha \leqslant N-2$, satisfy Eq. (64) for the SU($N-1$) Toda lattice. The first step in this construction corresponds to the SU(2) Toda lattice considered in the previous section. For this first step we have

$$F^1_1(z) = e^{\phi^2}, \quad K^1_1(\bar z) = e^{\gamma^2}. \tag{65}$$

Due to (65) we immediately obtain one particular solution for the functions $F^\alpha$ and $K^\alpha$ in the SU(3) case.

$$F^1 = (-1)\, e^{\phi^3/2} \int^z e^{\phi^2}\, dz',$$

$$K^1 = (-1)\, e^{\gamma^3/2} \int^{\bar z} e^{\gamma^2}\, d\bar z, \tag{66}$$

$$F^2 = e^{\phi^3/2}, \quad K^2 = e^{\gamma^3/2}. \tag{67}$$

Therefore, the continuations of this procedure give us that, for arbitrary $N$ in the SU($N$) Toda lattice, we have

$$F^\alpha = (-1)^\alpha e^{\phi^N/(N-1)} \int^z e^{\phi^{N-1}/(N-2)}\, dz \times \int^z \cdots \int^{z_\alpha} e^{\phi^{\alpha+1}/\alpha}\, dz_{\alpha+1}, \tag{68}$$

$$F^{N-1} = (-1)^{N-1} e^{\phi^N/(N-1)},$$
$$K^{N-1} = (-1)^{N-1} e^{\gamma^N/(N-1)}, \tag{69}$$

$$K^\alpha = (-1)^\alpha e^{\gamma^N/(N-1)} \int^{\bar z} e^{\gamma^{N-1}/(N-2)}\, d\bar z \times \int^z \cdots \int^{\bar z_\alpha} e^{\gamma^{\alpha+1}/\alpha}\, d\bar z_{\alpha+1}. \tag{70}$$

Substituting these formulas into (63) and next substituting into (47) and (49), we obtain

$$A = \Big(\sum_{\alpha}^{N-1} F^\alpha \int^{\bar z} e^{-\gamma^1} K^\alpha\, d\bar s + \chi(z)\Big) e^{\gamma^1}, \tag{71}$$

$$B = \Big(\sum_{\alpha}^{N-1} \int^z F^\alpha e^{-\phi^1}\, ds\cdot K^\alpha + \chi(\bar z)\Big) e^{\phi^1}. \tag{72}$$

Here the functions $\chi(z)$ and $\chi(\bar z)$ are unknown functions which play the role of the constants of the integrations. We determine them in the following manner. Preparing the integrations of our Bäcklund transformation for $h_1$, we obtain

$$e^{-h_1} = \sum_{\alpha=1}^{N+1} \mathscr{F}^\alpha \mathscr{K}^\alpha, \tag{73}$$

where

$$\mathscr{F}^\alpha = e^{\phi^1} \int^z e^{-\phi^1} F^\alpha\, ds, \tag{74}$$

$$\mathscr{K}^\alpha = e^{\gamma^1} \int^{\bar z} e^{-\gamma^1} K^\alpha\, d\bar s, \tag{75}$$

for $1 \leqslant \alpha \leqslant N-1$ and

$$\mathscr{F}^{N+1} = e^{\phi^1}, \quad \mathscr{K}^N = e^{\gamma^1}, \tag{76}$$

$$\mathscr{F}^N = e^{\phi^1} \int^z e^{-\phi^1} \chi(z)\, dz,$$
$$\mathscr{K}^{N+1} = e^{\gamma^1} \int^{\bar z} e^{-\gamma^1} \chi(\bar z)\, d\bar z. \tag{77}$$

Now we determine the functions $\chi(z)$ and $\chi(\bar z)$ in such a way to satisfy the condition (41). Substituting (73) into (41), we easily recognize that this formula reduces to

$$(-1)^{N(N+1)/2} \det{}_{N+1} \mathscr{F} \det{}_{N+1} \mathscr{K} = 1, \tag{78}$$

where

$$\mathscr{F}_{ij} = \mathscr{F}^i_{\underbrace{z,z,\ldots,z}_{j-1}}, \quad \mathscr{K}_{ij} = \mathscr{K}^j_{\underbrace{\bar z,\bar z,\ldots,\bar z}_{i-1}}.$$

By introducing the explicit form of $\mathscr{F}^\alpha$ and $\mathscr{K}^\alpha$ to (78), this formula reduces to

$$(-1)^{N(N+1)/2} e^{\phi^1+\gamma^1} \det{}_N F\cdot\det{}_N K = 1, \tag{79}$$

where $F_{ij}$ and $K_{ij}$ for $1 < i,j < N-1$ are the same functions as in Eq. (64) and

$$F_{Ni} = \chi(z)_{\underbrace{z,z,\ldots,z}_{i-1}} \tag{80}$$

$$K_{iN} = \chi(\bar z)_{\underbrace{\bar z,\bar z,\ldots,\bar z}_{i-1}}. \tag{81}$$

Equation (79) is similar to Eq. (64), and, using this equation, we obtain

$$\chi(z) = e^{\phi^N/(N-1)} \int^z \cdots \int^{z'} e^{\phi^2}\, dz'' \int^{z''} e^{\phi_0}\, dz''', \tag{82}$$

$$\chi(\bar z) = (-1)^{N+1} e^{\gamma^N/(N-1)} \int^{\bar z} \cdots \int^{\bar z'} e^{\gamma^2}\, d\bar z'' \int^{\bar z''} e^{\gamma_0}\, d\bar z''', \tag{83}$$

where

$$-\phi_0 = \phi^1 + \sum_{i=2}^{N} \frac{i\phi^i}{(i-1)} + \lambda, \tag{84}$$

$$-\gamma_0 = \gamma^1 + \sum_{i=2}^{N} \frac{i\gamma^i}{(i-1)} - \lambda, \tag{85}$$

where $\lambda$ is an arbitrary constant. This constant can be absorbed by the redefinition of $\phi^2$ and $\gamma^2$. Then this constant will not appear in the solutions of the SU($N$) Toda lattice, and therefore these solutions are reduced to those found by Leznov.

## V. CONCLUDING REMARKS

Here we have found the Liouville–Bäcklund transformation for the SU($N$) Toda lattice. This transformation relates the $N$ arbitrary solutions of the two-dimensional Laplace equation with our equation. Moreover, let us notice that this transformation is invariant under the Weyl group. Indeed notice that the arbitrary permutation of $\mathscr{F}^\alpha$ together with arbitrary permutation of $\mathscr{K}^\alpha$ in (73) is also the solution of Eq. (78) and hence is the solution of our equation. But this invariance, as was pointed by Leznov, corresponds to the invariance under Weyl group in the SU($N+1$) gauge theory.

Finally let us notice that it will be very interesting to extend this transformation to an arbitrary compact gauge group for the self-dual equations. In this case we have slightly different Toda lattice in the two-dimensional space-time. For the classical compact gauge group Leznov solved this equations by the same method as for the SU($N$) case,[14] and hence probably the Liouville–Bäcklund method can be extended too.

## ACKNOWLEDGMENTS

The author wishes to thank Dr. Ling-Lie Chau and the High Energy Theory Group for the warm hospitality and discussions during his stay at Brookhaven National Laboratory and for the financial support. This work has been authored under Contract Nos. DE-AC02-76CH00016 and DE-AC02-83ER40033.B000 with the U.S. Department of Energy, and the National Science Foundation Grant PHY 8116102.

[1]A. Belavin, A. Polyakov, A. Schwartz, and Y. Tupkin, Phys. Lett. B **59**, 85 (1975).
[2]E. Witten, Phys. Rev. Lett. **38**, 121 (1976).
[3]A. Leznov and M. Saveliev, Lett. Math. Phys. **3**, 489 (1979).
[4]M. Toda, Phys. Rep. C **18**, 1 (1975).
[5]V. Zakharov, S. Manakov, P. Novikov, and A. Pytaevsky, *The Theory of Solitons* (Nauka, Moscow, 1980) (in Russian).
[6]M. Wadati and M. Toda, J. Phys. Soc. Jpn. **39**, 1196 (1975).
[7]B. Kostant, Adv. Math. **34**, 195 (1979).
[8]M. Olshanetsky and A. Perelomov, Teor. Math. Phys. **45**, 3 (1980) (in Russian).
[9]Z. Popowicz, Phys. Lett. A **81**, 235 (1981).
[10]D. Levi, L. Pinlloni, and P. Santini, J. Phys. A **14**, 1567 (1981).
[11]Z. Popowicz, Z. Phys. C **19**, 79 (1983).
[12]A. Mihajlov, Pis'ma Zh. Eks. Teor. Fiz. **30**, 443 (1979).
[13]A. Fordy and J. Gibbon, Commun. Math. Phys. **77**, 21 (1980).
[14]A. Leznov, Teor. Math. Phys. **42**, 343 (1980) (in Russian).
[15]A. Leznov and M. Saveliev, Phys. Elem. Part. At. Nuclei **11**, **40** (1980) (in Russian).
[16]A. Leznov and M. Saveliev, Phys. Elem. Part. At. Nuclei **12**, 125 (1981).
[17]G. Lamb, Lect. Not. Math. **515**, 69 (1976).

# Dynamical invariants for two-dimensional time-dependent classical systems

S. C. Mishra, R. S. Kaushal,[a] and K. C. Tripathy
*Department of Physics and Astrophysics, University of Delhi, Delhi-110 007, India*

(Received 15 June 1983; accepted for publication 18 November 1983)

General equations are formulated to determine all potentials for two-dimensional systems of the type $L = \frac{1}{2}(p_1^2 + p_2^2) - V(q_1,q_2,t)$, which admits invariants of the form $I = a_0 + a_i\xi_i + \frac{1}{2}a_{ij}\xi_i\xi_j$, $i,j = 1,2$, where $\xi_1 = \dot z = \dot q_1 + i\dot q_2$, $\xi_2 = \dot{\bar z} = \dot q_1 - i\dot q_2$, $a_0$, $a_i$, $a_{ij}$ are arbitrary functions of $t$, $z = q_1 + iq_2$, and $\bar z = q_1 - iq_2$. Simplifying restrictions reduce the general equation to a tractable form. The resulting equations are solved for a special class of time-separable potentials and derive (i) the van der Waals-type long-range potential, $V(r,t) = \beta(t)(b/r^4 + d)$ and (ii) the quark-confining logarithmic potential, $V(r,t) = \beta(t)\lambda(\ln r + b_1/r^4 + d_1)$. Invariants $I$ for the resulting dynamical systems are found. Some observations on the present method in the context of Katzin and Levine and of Lewis and Leach analyses have also been made.

PACS numbers: 03.20. + i, 03.40. − t, 11.30. − j

## I. INTRODUCTION

Recently, considerable activities in constructing exact invariants for time-dependent classical dynamical systems described by the Hamiltonian $H = \frac{1}{2}p^2 + V(q,t)$ or the Lagrangian $L = \frac{1}{2}p^2 - V(q,t)$ have been initiated.[1–6] Such studies have a lot of bearing in plasma physics, time-dependent Kepler and harmonic oscillator motions,[4–6] $\alpha$-decay, time-dependent gravitational constants, time-varying mass for accelerating dynamical systems, and time-dependent magnetic monopole problems.[7] So far, the analysis is mainly directed towards one-dimensional dynamical systems.[4] Katzin and Levine have, however, discussed this problem for the restricted class of Kepler, harmonic oscillator, and their linearly combined potentials in two dimensions.[5,6] Following the recipe of Ref. 5, we reexamine the classical Lagrangian system,

$$L = \tfrac{1}{2}|\dot z|^2 - V(z,\bar z,t), \quad z = q_1 + iq_2, \quad \dot z = p_1 + ip_2$$

and restrict ourselves to the determination of the constants of the motion of the form

$$I = a_0 + a_i\xi_i + \tfrac{1}{2}a_{ij}\xi_i\xi_j, \quad \xi_1 = \dot z, \quad \xi_2 = \dot{\bar z},$$

where the coefficients $a_0$, $a_i$, $a_{ij}$ explicitly depend on time $t$, $z$, and $\bar z$ and $a_{ij} = a_{ji}$. Our material is arranged as follows.

In Sec. II, we consider the Lagrangian, $L = \frac{1}{2}|\dot z|^2 - V(z,\bar z,t)$, and, requiring that $dI/dt = 0$ and using the *ansatz* (2.35), we obtain a second-order differential equation for the potential (2.36). The potentials satisfying such equations are derived, and the corresponding invariants are constructed. In Sec. III, we restrict our analysis to the potential of the form $V(z,\bar z,t) = V(|z|,t) = \beta(t)v(|z|)$ and derive two important class of potentials, namely, case (1), $V(|z|,t) = \beta(t)(b/r^4 + d)$, and case (2), $V(|z|,t) = \beta(t)\lambda(\ln r + b_1/r^4 + d_1)$. The corresponding invariants for these two cases are constructed. In Sec. IV, we rewrite the invariant $I$ in the form

$$I = \sum_{m,n=0}^{\infty} f_{mn}(z,\bar z,t)\xi_1^m\xi_2^n, \quad \xi_1 = \dot z, \quad \xi_2 = \dot{\bar z},$$

and the corresponding Hamiltonian $H = \frac{1}{2}\xi_1\xi_2 + V(z,\bar z,t)$. On demanding $dI/dt = \partial I/\partial t + [I,H]_t = 0$, we obtain a recursion relation for the coefficients $f_{mn}$. On restriction of $m,n$, i.e., $0 \leqslant m + n \leqslant 2$, and properly identifying $f_{mn}$ with $a_0$, $a_i$, $\frac{1}{2}a_{ij}$ of Sec. II, we establish the correspondence with the Lewis and Leach approach[4] and our analysis. In Sec. V, we examine the potential $V(z,\bar z,t) = \frac{1}{2}\beta(t)|z|^2$ and from the potential equation fix $\sigma_1$ and $\sigma_2$. On substituting $\sigma_1$, $\sigma_2$ and by suitably fixing other parameters, $a_0$, $a_i$, $a_{ij}$ can be determined which in turn yield the invariant $I$.

We summarize our discussions in Sec. VI.

## II. CONSTRUCTION OF THE POTENTIALS AND CORRESPONDING INVARIANTS

### A. The method

We consider a dynamical system described by the Lagrangian

$$L = \tfrac{1}{2}|\dot z|^2 - V(z,\bar z,t) \tag{2.1}$$

with the concomitant equations of motion,

$$\ddot z = -2\frac{\partial V}{\partial \bar z}, \quad \ddot{\bar z} = -2\frac{\partial V}{\partial z}. \tag{2.2}$$

Let us consider the constants of the motion of the form

$$I = a_0 + a_i\xi_i + \tfrac{1}{2}a_{ij}\xi_i\xi_j, \quad i,j = 1,2. \tag{2.3}$$

The coefficients $a_0$, $a_i$, $a_{ij}$ explicitly depend on $t$, $z$, and $\bar z$. Using $dI/dt = 0$, we find from (2.3),

$$(\dot a_0 + a_i\dot\xi_i) + (a_{0,i} + \dot a_i + a_{ij}\dot\xi_j)\xi_i + (a_{i,j} + \tfrac{1}{2}\dot a_{ij})\xi_i\xi_j + \tfrac{1}{2}a_{ij,k}\xi_i\xi_j\xi_k = 0. \tag{2.4}$$

[a] Department of Physics, Ramjas College, University of Delhi, Delhi-110 007, India.

 0022-2488/84/072217-05$02.50  

Taking into account the proper symmetrization of the coefficiants $a_0$, $a_i$, $a_{ij}$ we obtain from (2.4)

$$a_{ij,k} + a_{jk,i} + a_{ki,j} = 0, \tag{2.5}$$

$$a_{i,j} + a_{j,i} = -\dot{a}_{ij}, \tag{2.6}$$

$$a_{0,i} = -\dot{a}_i - a_{ij}\dot{\xi}_j, \tag{2.7}$$

and

$$\dot{a}_0 = -a_i\dot{\xi}_i. \tag{2.8}$$

Since $a_{12} = a_{21}$, Eq. (2.5) yields

$$\frac{\partial a_{11}}{\partial z} = 0, \tag{2.9}$$

$$\frac{\partial a_{22}}{\partial \bar{z}} = 0, \tag{2.10}$$

$$2\frac{\partial a_{12}}{\partial z} + \frac{\partial a_{11}}{\partial \bar{z}} = 0, \tag{2.11}$$

and

$$2\frac{\partial a_{12}}{\partial \bar{z}} + \frac{\partial a_{22}}{\partial z} = 0, \tag{2.12}$$

whereas Eqs. (2.6)–(2.8) and (2.2) yield

$$2\frac{\partial a_1}{\partial z} = -\frac{\partial a_{11}}{\partial t}, \tag{2.13}$$

$$2\frac{\partial a_2}{\partial \bar{z}} = -\frac{\partial a_{22}}{\partial t}, \tag{2.14}$$

$$\frac{\partial a_1}{\partial \bar{z}} + \frac{\partial a_2}{\partial z} = -\frac{\partial a_{12}}{\partial t}, \tag{2.15}$$

$$\frac{\partial a_0}{\partial z} = -\frac{\partial a_1}{\partial t} + 2a_{11}\frac{\partial V}{\partial \bar{z}} + 2a_{12}\frac{\partial V}{\partial z}, \tag{2.16}$$

$$\frac{\partial a_0}{\partial \bar{z}} = -\frac{\partial a_2}{\partial t} + 2a_{12}\frac{\partial V}{\partial \bar{z}} + 2a_{22}\frac{\partial V}{\partial z}, \tag{2.17}$$

$$\frac{\partial a_0}{\partial t} = 2a_1\frac{\partial V}{\partial \bar{z}} + 2a_2\frac{\partial V}{\partial z}. \tag{2.18}$$

Now, we solve Eqs. (2.9)–(2.18) for determining $a_0$, $a_i$, and $a_{ij}$.

### B. Determination of $a_{ij}$

From Eqs. (2.9) and (2.10), $a_{11} = a_{11}(\bar{z},t)$ and $a_{22} = a_{22}(z,t)$. Since $\partial^2 a_{12}/\partial z\partial\bar{z} = \partial^2 a_{12}/\partial\bar{z}\partial z$, Eqs. (2.11) and (2.12) yield

$$\frac{\partial^2 a_{11}}{\partial \bar{z}^2} = \frac{\partial^2 a_{22}}{\partial z^2} = 2\sigma_0(t) \quad \text{(say)}. \tag{2.19}$$

Solving for $a_{11}$, $a_{22}$, we have

$$a_{11} = \sigma_0(t)\bar{z}^2 + \sigma_2(t)\bar{z} + \sigma_3(t) \tag{2.20}$$

and

$$a_{22} = \sigma_0(t)z^2 + \sigma_1(t)z + \sigma_4(t). \tag{2.21}$$

Substituting for $a_{11}$, $a_{22}$ in (2.11) and (2.12), we obtain

$$a_{21}(z,\bar{z},t) = a_{12}(z,\bar{z},t) = -\sigma_0(t)\bar{z}z - \tfrac{1}{2}\sigma_2(t)z - \tfrac{1}{2}\sigma_1(t)\bar{z} + \tfrac{1}{4}\mu(t), \tag{2.22}$$

$\mu(t)$ being the integration constant.

### C. Determination of $a_i$

Substituting for $a_{11}$, $a_{22}$ from (2.20) and (2.21) in (2.13) and (2.14), we get

$$a_1 = -\tfrac{1}{2}[\dot{\sigma}_0(t)\bar{z}^2 z + \dot{\sigma}_2(t)\bar{z}z + \dot{\sigma}_3(t)z] + \tfrac{1}{2}\tau_1(\bar{z},t) \tag{2.23}$$

and

$$a_2 = -\tfrac{1}{2}[\dot{\sigma}_0(t)z^2\bar{z} + \dot{\sigma}_1(t)z\bar{z} + \dot{\sigma}_4(t)\bar{z}] + \tfrac{1}{2}\tau_2(z,t); \tag{2.24}$$

using (2.23), (2.24), and (2.22) in (2.15), we have

$$3\dot{\sigma}_0(t)z\bar{z} + \dot{\sigma}_2(t)z + \dot{\sigma}_1(t)\bar{z} - \tfrac{1}{4}\dot{\mu}(t) - \frac{1}{2}\frac{\partial\tau_1}{\partial\bar{z}} - \frac{1}{2}\frac{\partial\tau_2}{\partial z} = 0. \tag{2.25}$$

Differentiating (2.25) w.r.t. $z$, then $\bar{z}$, we get

$\dot{\sigma}_0(t) = 0$, i.e.,

$$\sigma_0(t) = c_1 \quad \text{(some constant independent of time)}. \tag{2.26}$$

Substituting for $\sigma_0$ in (2.25) yields

$$\dot{\sigma}_2 z + \dot{\sigma}_1\bar{z} - \frac{1}{2}\dot{\mu}(t) - \frac{1}{2}\frac{\partial\tau_1}{\partial\bar{z}} - \frac{1}{2}\frac{\partial\tau_2}{\partial z} = 0. \tag{2.27}$$

Differentiating (2.27) w.r.t. $z$, we have

$$\dot{\sigma}_2 - \frac{1}{2}\frac{\partial^2\tau_2}{\partial z^2} = 0. \tag{2.28}$$

Solving for $\tau_2$, we obtain, from (2.28),

$$\tau_2 = \tau_2(z,t) = \dot{\sigma}_2(t)z^2 + \sigma_5(t)z + \sigma_6(t); \tag{2.29}$$

similarly, differentiating (2.27) w.r.t. $\bar{z}$, and solving for $\tau_1$, we find

$$\tau_1 = \tau_1(\bar{z},t) = \dot{\sigma}_1(t)\bar{z}^2 - \dot{\mu}(t)\bar{z} - \sigma_5(t)\bar{z} + \sigma_7(t). \tag{2.30}$$

Substituting for $\tau_1$, $\tau_2$ from (2.29) and (2.30) and $\sigma_0 = c_1$ from (2.26) in (2.23) and (2.24), we obtain

$$a_1 = -\tfrac{1}{2}[\dot{\sigma}_2(t)\bar{z} + \dot{\sigma}_3(t)]z + \tfrac{1}{2}\dot{\sigma}_1(t)\bar{z}^2 - \tfrac{1}{2}[\dot{\mu}(t) + \sigma_5(t)]\bar{z} + \tfrac{1}{2}\sigma_7(t) \tag{2.31}$$

and

$$a_2 = -\tfrac{1}{2}[\dot{\sigma}_1(t)z + \dot{\sigma}_4(t)]\bar{z} + \tfrac{1}{2}\dot{\sigma}_2(t)z^2 + \tfrac{1}{2}[\sigma_5(t)]z + \tfrac{1}{2}\sigma_6(t). \tag{2.32}$$

### D. Determination of $a_0$

Differentiating (2.16) and (2.17) w.r.t. $\bar{z}$ and $z$, respectively, and using

$$\frac{\partial^2 V}{\partial z\partial\bar{z}} = \frac{\partial^2 V}{\partial\bar{z}\partial z},$$

we obtain

$$-\frac{\partial}{\partial\bar{z}}\left(\frac{\partial a_1}{\partial t}\right) + 2\frac{\partial a_{11}}{\partial\bar{z}}\cdot\frac{\partial V}{\partial\bar{z}} + 2a_{11}\frac{\partial^2 V}{\partial\bar{z}^2} + 2\frac{\partial a_{12}}{\partial\bar{z}}\cdot\frac{\partial V}{\partial z} = -\frac{\partial}{\partial z}\left(\frac{\partial a_2}{\partial t}\right) + 2\frac{\partial a_{22}}{\partial z}\cdot\frac{\partial V}{\partial z} + 2a_{22}\frac{\partial^2 V}{\partial z^2} + 2\frac{\partial a_{12}}{\partial z}\cdot\frac{\partial V}{\partial\bar{z}}. \tag{2.33}$$

Substituting for $a_1$, $a_2$, $a_{11}$, $a_{22}$, $a_{12}$ from (2.31), (2.32), (2.20), (2.21), and (2.22), we have from (2.33)

$$2\{c_1z^2+\sigma_1(t)z+\sigma_4(t)\}\frac{\partial^2 V}{\partial z^2}+3\{2c_1z+\sigma_1(t)\}\frac{\partial V}{\partial z}$$
$$+\{-\tfrac{3}{2}\ddot\sigma_2(t)z-\dot\sigma_5(t)-\tfrac{1}{2}\dot\mu(t)\}$$
$$=2\{c_1\bar z^2+\sigma_2(t)\bar z+\sigma_3(t)\}\frac{\partial^2 V}{\partial \bar z^2}$$
$$+3\{2c_1\bar z+\sigma_2(t)\}\frac{\partial V}{\partial \bar z}$$
$$+\{-\tfrac{3}{2}\ddot\sigma_1(t)\bar z\}. \tag{2.34}$$

Let us make the *ansatz*

$$\dot\mu(t)=0,\quad \dot\sigma_5(t)=0,\quad \sigma_1=\bar\sigma_2=\sigma_2,$$

and

$$\sigma_3=\bar\sigma_4=\sigma_4. \tag{2.35}$$

Then,

$$\sigma_5=c_2\quad \text{(say)}.$$

Using (2.35) in (2.34), we find

$$A\frac{\partial^2 V}{\partial z^2}+B\frac{\partial V}{\partial z}+C=\bar A\frac{\partial^2 V}{\partial \bar z^2}+\bar B\frac{\partial V}{\partial \bar z}+\bar C$$
$$=\varphi(t)\quad \text{(say)}, \tag{2.36}$$

where

$$A=2\{c_1z^2+\sigma_1(t)z+\sigma_3(t)\},$$
$$B=3\{2c_1z+\sigma_1(t)\}, \tag{2.37}$$
$$C=-\tfrac{3}{2}\ddot\sigma_2(t)z.$$

Equations (2.36) are called "potential equations," and solutions of (2.36) give a class of potentials. Before we consider some special cases for solving (2.36), certain remarks are in order. Katzin and Levine,[5] in order to solve the time-dependent Kepler problem, had assumed $\mu=0$, $\ddot\sigma_1=\ddot\sigma_2=0$, $\sigma_3=\sigma_4$, $\sigma_5=\text{const}$, $\sigma_6=\sigma_7=0$, $\sigma_0=\text{const}$. In our case, we resorted to the *ansatz* (2.35), so that we can reduce (2.34) into a pair of conjugate equations for the potential equation (2.36). Secondly, our Eq. (2.34) in its general form when supplemented with Eqs. (2.16)–(2.18) provides an explicit form for the invariants for the time-dependent Kepler,[8] harmonic oscillator,[9] and their linearly combined potentials.[10]

Solving for the potential $V$ from (2.36), we fix the coefficient $a_0$, which in turn together with $a_i$, $a_{ij}$ determine the invariant $I(z,\bar z,t)$.

## III. SOME SPECIAL CASES

Here, we consider the potential

$$V(z,\bar z,t)=V(|z|,t)\equiv\beta(t)v(|z|)\quad \text{(say)}. \tag{3.1}$$

Thus,

$$\frac{\partial V}{\partial z}=\frac{\beta(t)}{2z}|z|\frac{dv}{d|z|},$$

$$\frac{\partial^2 V}{\partial z^2}=\frac{\beta(t)|z|}{4z^2}\left\{-\frac{dv}{d|z|}+|z|\frac{d^2v}{d|z|^2}\right\}. \tag{3.2}$$

Substituting (3.2) in the potential equations (2.36),

$$A\frac{\partial^2 V}{\partial z^2}+B\frac{\partial V}{\partial z}+C=\varphi(t),$$
$$\bar A\frac{\partial^2 V}{\partial \bar z^2}+\bar B\frac{\partial V}{\partial \bar z}+\bar C=\varphi(t), \tag{2.36'}$$

we have

$$\left\{c_1+\frac{\sigma_1}{z}+\frac{\sigma_3}{z^2}\right\}\frac{\beta}{2}|z|\cdot\left\{-\frac{dv}{d|z|}+|z|\frac{d^2v}{d|z|^2}\right\}$$
$$+3\left\{2c_1+\frac{\sigma_1}{z}\right\}\frac{\beta}{2}|z|\frac{dv}{d|z|}-\frac{3}{2}\ddot\sigma_1 z=\varphi(t) \tag{3.3a}$$

and

$$\left\{c_1+\frac{\sigma_1}{\bar z}+\frac{\sigma_3}{\bar z^2}\right\}\frac{\beta}{2}|z|\left\{-\frac{dv}{d|z|}+|z|\frac{d^2v}{d|z|^2}\right\}$$
$$+3\left\{2c_1+\frac{\sigma_1}{\bar z}\right\}\frac{\beta}{2}|z|\frac{dv}{d|z|}-\frac{3}{2}\ddot\sigma_1\bar z=\varphi(t). \tag{3.3b}$$

In order that (3.3a) and (3.3b) be simultaneously satisfied by $v(|z|)$, we must have

$$\sigma_1=\sigma_3=0\quad\text{and}\quad\ddot\sigma_1=0. \tag{3.4}$$

Thus, (3.3) reduces to

$$|z|^2\frac{d^2v}{d|z|^2}+5|z|\frac{dv}{d|z|}=\lambda,$$

where

$$\lambda=\frac{2\varphi(t)}{\beta(t)}. \tag{3.5}$$

Note $\lambda$ is a constant independent of time.

We consider the following two interesting cases.

*Case (a)*: $\lambda=0$: Eq. (3.5) reduces to the form ($|z|=r$)

$$r^2\frac{d^2v}{dr^2}+5r\frac{dv}{dr}=0.$$

Thus, we have the nontrivial solution for $v$:

$$v=v(r)=(b/r^4)+d, \tag{3.6}$$

where $b$, $d$ are some arbitrary constants. (3.6) is the well-known van der Waals-type potential. Now, using the *ansatz* (2.35) and (3.4) in the expressions for $a_0$, $a_1$, $a_2$, $a_{11}$, $a_{12}$, and $a_{22}$, we obtain

$$a_1=-\tfrac{1}{2}c_2\bar z,\quad a_2=\tfrac{1}{2}c_2 z,$$
$$a_{11}=c_1\bar z^2,\quad a_{12}=a_{21}=-c_1z\bar z,\quad a_{22}=c_1z^2, \tag{3.7}$$
$$a_0=-2c_2bB(t)/r^4,$$

where $B(t)=\int\beta(t')\,dt'$ and $\sigma_6=\sigma_7=0$ is assumed.

Finally, the invariant (2.3) can be written in the form

$$I=\alpha(t)/r^4+\alpha_1(t)L+\alpha_2L^2, \tag{3.8}$$

where

$$\alpha(t)=-2c_2bB(t),\quad \alpha_1=-2ic_2,\quad \alpha_2=-8c_1, \tag{3.9}$$
$$L=q_1p_2-q_2p_1=(1/4i)(\xi_1\bar z-\xi_2 z).$$

*Case (b)*: $\lambda=\lambda_0\neq0$: Eq. (3.5) yields

$$r^2\frac{d^2v}{dr^2}+5r\frac{dv}{dr}=\lambda_0. \tag{3.10}$$

Solving for $v$ in (3.10), we obtain

$$v(r)=\tfrac{1}{4}\lambda_0(\ln r+b_1/r^4+d_1). \tag{3.11}$$

The invariant for this case turns out to be

$$I = \delta_0(t) + \delta_1(t)/r^4 + \delta_2 L + \delta_3 L^2, \tag{3.12}$$

where

$$\delta_0(t) = \tfrac{1}{2}c_2 B(t), \quad \delta_1(t) = -2c_2 B(t) b_1,$$
$$\delta_2 = -2ic_2, \quad \delta_3 = -8c_1.$$

## IV. CORRESPONDENCE WITH LEWIS AND LEACH METHOD[4]

In this section, we extend the method of Lewis and Leach[4] to two dimensions and show that up to quadratic terms in momentum, the recursion formula method yields the same set of equations for the coefficients and determines correspondingly the same invariant $I$.

Consider the Hamiltonian of the classical dynamical system (2.1),

$$H = \tfrac{1}{2}(p_1^2 + p_2^2) + V(q_1,q_2,t)$$
$$\equiv \tfrac{1}{2}\xi_1\xi_2 + V(z,\bar{z},t). \tag{4.1}$$

Let the invariant $I$ for (4.1) be expressed as a double power series in $\xi_1$, $\xi_2$, i.e.,

$$I = \sum_{m,n=0}^{\infty} f_{mn}(z,\bar{z},t)\xi_1^m\xi_2^n. \tag{4.2}$$

Using the equation for the invariant $I$,

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + \sum_i \left(\frac{\partial I}{\partial q_i}\cdot\frac{\partial H}{\partial p_i} - \frac{\partial I}{\partial p_i}\frac{\partial H}{\partial q_i}\right)$$
$$= \frac{\partial I}{\partial t} + 2\left(\frac{\partial I}{\partial z}\frac{\partial H}{\partial \xi_2} + \frac{\partial I}{\partial \bar{z}}\frac{\partial H}{\partial \xi_1} - \frac{\partial I}{\partial \xi_1}\frac{\partial H}{\partial \bar{z}} - \frac{\partial I}{\partial \xi_2}\cdot\frac{\partial H}{\partial z}\right)$$
$$= 0,$$

and demanding the coefficients of $\xi_1^m\xi_2^n$ to vanish, we obtain the following recursion relation for $f_{mn}$:

$$\dot{f}_{mn} + \frac{\partial f_{m-1,n}}{\partial z} + \frac{\partial f_{m,n-1}}{\partial \bar{z}} - 2(m+1)f_{m+1,n}\frac{\partial V}{\partial \bar{z}}$$
$$- 2(n+1)f_{m,n+1}\frac{dV}{dz} = 0. \tag{4.3}$$

If we restrict our analysis to the case $0 \leqslant m + n \leqslant 2$, then

$$I = f_{00} + f_{01}\xi_2 + f_{10}\xi_1 + f_{11}\xi_1\xi_2 + f_{02}\xi_2^2 + f_{20}\xi_1^2. \tag{4.4}$$

Equation (4.3) then yields

$$\dot{f}_{00} - 2f_{10}\frac{\partial V}{\partial \bar{z}} - 2f_{01}\frac{\partial V}{\partial z} = 0,$$
$$\dot{f}_{01} + \frac{\partial f_{00}}{\partial \bar{z}} - 2f_{11}\frac{\partial V}{\partial \bar{z}} - 4f_{02}\frac{\partial V}{\partial z} = 0,$$
$$\dot{f}_{10} + \frac{\partial f_{00}}{\partial z} - 2f_{11}\frac{\partial V}{\partial z} - 4f_{20}\frac{\partial V}{\partial \bar{z}} = 0,$$
$$\dot{f}_{02} + \frac{\partial f_{01}}{\partial \bar{z}} = 0, \quad \dot{f}_{20} + \frac{\partial f_{10}}{\partial z} = 0, \tag{4.5}$$
$$\dot{f}_{11} + \frac{\partial f_{01}}{\partial z} + \frac{\partial f_{10}}{\partial \bar{z}} = 0,$$
$$\frac{\partial f_{20}}{\partial z} = 0, \quad \frac{\partial f_{02}}{\partial \bar{z}} = 0,$$
$$\frac{\partial f_{11}}{\partial z} + \frac{\partial f_{20}}{\partial \bar{z}} = 0, \quad \frac{\partial f_{02}}{\partial z} + \frac{\partial f_{11}}{\partial \bar{z}} = 0.$$

We note that (4.5) coincides with Eqs. (2.9)–(2.18). In the method of Lewis and Leach, the symmetry is built in and this gives a general method of constructing invariants involving higher powers of momenta.

## V. TIME-DEPENDENT, UNCOUPLED HARMONIC OSCILLATOR (TWO-DIMENSIONAL) MOTION

Let us consider $V(z,\bar{z},t) = \beta(t)v(|z|) \equiv \tfrac{1}{2}\beta(t)|z|^2$. Then,

$$\frac{\partial V}{\partial z} = \frac{1}{2}\frac{\beta(t)}{z}|z|^2, \quad \frac{\partial V}{\partial \bar{z}} = \frac{1}{2}\frac{\beta(t)}{\bar{z}}|z|^2,$$
$$\frac{\partial^2 V}{\partial z^2} = \frac{\partial^2 V}{\partial \bar{z}^2} = 0. \tag{5.1}$$

Thus, Eq. (2.34) reduces to

$$\tfrac{3}{2}\{2c_1 z + \sigma_1(t)\}[\beta(t)/z]|z|^2 - \tfrac{3}{2}\ddot{\sigma}_2(t)z - \dot{\sigma}_5(t) - \tfrac{1}{2}\ddot{\mu}(t)$$
$$= \tfrac{3}{2}\{2c_1\bar{z} + \sigma_2(t)\}[\beta(t)/\bar{z}]|z|^2 - \tfrac{3}{2}\ddot{\sigma}_1(t)\bar{z}. \tag{5.2}$$

Using the *ansatz* $\sigma_5 = c_2$, $\dot{\sigma}_5 = 0$, $\ddot{\mu} = 0$, Eq. (5.2) reduces to

$$\frac{\beta(t)}{z}\sigma_1(t) + \frac{\ddot{\sigma}_1(t)}{z} = \frac{\beta(t)}{\bar{z}}\sigma_2(t) + \frac{\ddot{\sigma}_2(t)}{\bar{z}} = k_1 \quad \text{(say)}. \tag{5.3}$$

For $k_1 = 0$,

$$\ddot{\sigma}_1(t) + \beta(t)\sigma_1(t) = 0, \tag{5.4a}$$
$$\ddot{\sigma}_2(t) + \beta(t)\sigma_2(t) = 0. \tag{5.4b}$$

The solutions of (5.4a) or (5.4b) are given by[4]

$$\left[\sigma = \rho\sin T, \quad \rho\cos T, \quad T = \int \rho^{-2}(t')\,dt'\right], \tag{5.5}$$

where $\rho$ satisfies the auxiliary equation $\ddot{\rho} + \beta(t)\rho = \rho^{-3}$. Substituting for $\sigma_1$, $\sigma_2$ in the expressions for $a_{11}$, $a_{12}$, $a_{22}$, $a_1$, $a_2$, and $a_0$, the invariants can be found out.

## VI. CONCLUSIONS

Our analysis has the following features:

(i) It establishes the correspondence with the Katzin–Levine and Lewis–Leach methods when $I$ has terms up to quadratic in momenta.

(ii) Writing $I = \sum_{m,n=0}^{\infty} f_{mn}\xi_1^m\xi_2^n$, $\xi_1 = \dot{z}$, $\xi_2 = \dot{\bar{z}}$, we have extended, in fact, Lewis–Leach analysis to double series expansion in $\xi_1$, $\xi_2$. The suitable convergence of the series is assumed. Our prescription, in principle, can be used to determine analytic potentials and the corresponding invariants.

(iii) By restricting $0 \leqslant m + n \leqslant 2$, i.e., considering $I$ in the form: $I = a_0 + a_i\xi_i + \tfrac{1}{2}a_{ij}\xi_i\xi_j$, and using the *ansatz* (2.35), we have derived two interesting types of potentials, namely, (1) the van der Waals-type long-range potential and (2) the quark-confining logarithmic potential, which are both time-dependent. The later potential can have a lot of applications in string models of quark confinement,[11] particularly when the coupling coefficient becomes time-dependent.


[1]J. R. Ray and J. L. Reid, J. Math. Phys. **20**, 2054 (1979); **23**, 1042 (1982); Phys. Lett. A **71**, 317 (1979).

[2]H. J. Korsch, Phys. Lett. A **74**, 294 (1979); R. S. Kaushal and H. J. Korsch, J. Math. Phys. **22**, 1904 (1981).

[3]M. Lutzky, J. Phys. A **11**, 249 (1978); Phys. Lett. A **68**, 3 (1978); P. G. L. Leach, J. Math. Phys. **22**, 465 (1981); N. J. Gunther and P. G. L. Leach, J. Math. Phys. **18**, 572 (1977).

[4]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 237 (1982).

[5]G. H. Katzin and J. L. Levine, J. Math. Phys. **23**, 552 (1982).

[6]G. H. Katzin and J. L. Levine, J. Math. Phys. **18**, 1267 (1977); **24**, 1761 (1983).

[7]H. Arodź, Phys. Rev. D **27**, 1903 (1983).

[8]Using $V(z,\bar{z},t) = V(|z|,t) = -\beta(t)/r = -\beta(t)/(z\bar{z})^{1/2}$ Coulomb case, we obtain from Eq. (2.34) $\ddot{\sigma}_1 = 0, \ddot{\sigma}_2 = 0, \dot{\sigma}_5 + \frac{1}{2}\ddot{\mu} = 0, \sigma_3 = \sigma_4 = 0$. Differentiating partially Eq. (2.16) w.r.t. $t$ and Eq. (2.18) w.r.t. $z$ and using $\partial^2 a_0/\partial t\cdot\partial z = \partial^2 a_0/\partial z\cdot\partial t$, we finally obtain $\sigma_6 = 0$, $\sigma_7 = 0$, $\beta = (at + b)^{-1}$, and $\mu = (a't + b')^2$. We have used here $\sigma_1 = \sigma_2 = at + b$.

[9]For time-dependent harmonic oscillator motion, see Sec. V.

[10]Substituting $V(r,t) = -\frac{1}{2}(\ddot{U}/U)r^2 - (\mu_0/U)\cdot(1/r)$ [see G. Katzin and J. Levine, J. Math. Phys. **24**, 1761 (1983)] or $V(z,\bar{z},t) = -\frac{1}{2}(\ddot{U}/U)z\bar{z} - (\mu_0/U)(z\bar{z})^{-1/2}$ in Eq. (2.34), and, using Eqs. (2.16)–(2.18), we find $\sigma_1 = k_1 U(t)$, $\sigma_2 = k_2 U(t), \sigma_3 = \sigma_4 = 0, \sigma_5 = -U\dot{U}/\mu^2 + k_3, \sigma_6 = \sigma_7 = 0, \mu = U^2/\mu_0^2$ and $k_1, k_2, k_3$ being some arbitrary constants. On substituting these values for $a_i$, $a_{ij}$, the invariant $I$ can be obtained.

[11]See, for example, H. J. W. Müller-Kirsten and S. K. Bose, J. Math. Phys. **20**, 1878 (1979) for the discussion on logarithmic potential and H. H. Aly, H. J. W. Müller-Kirsten and N. Vahedi-Faridi, J. Math. Phys. **16**, 961 (1975) for $1/r^4$ potential.

# General prolongations and (x, t)-depending pseudopotentials for the KdV equation

Pierre Molino
*Département de Mathématiques, Université des Sciences et Techniques du Languedoc, Place E. Bataillon, 34060 Montpellier, France*

(Received 4 January 1984; accepted for publication 20 January 1984)

Given an exterior differential system on a manifold $M$, we study general prolongations of the system on a locally trivial fiber bundle $(\widetilde{M}, \tilde{\pi}, M)$ by a Cartan–Ehresmann connection. We characterize such prolongations for the system associated with the KdV equation without any assumption of "$(x, t)$ independence." The partial Lie algebra discovered by Wahlquist–Estabrook [J. Math. Phys. **16**, 1 (1975)] appears by this way as an intrinsic tool. Simple analytic pseudopotentials are classified up to diffeomorphism.

PACS numbers: 03.40.Kf

## I. INTRODUCTION

Differentiability is assumed to be $C^\infty$.

Following Wahlquist–Estabrook,[1] we consider $M = \mathbb{R}^5$ with coordinates $(x, t, u, z, p)$ and the projection $\pi: M \to \mathbb{R}^2$ defined by $\pi(x, t, u, z, p) = (x, t)$.

On $M$, we consider the following *exterior differential system* (EDS):

$$\alpha \equiv du \wedge dt - z\, dx \wedge dt = 0,$$
$$\beta \equiv dz \wedge dt - p\, dx \wedge dt = 0, \qquad (1)$$
$$\gamma \equiv -du \wedge dx + dp \wedge dt + 12uz\, dx \wedge dt = 0.$$

A submanifold $S$ of $M$ is an *integral manifold* of (1) in the sense of Cartan[2] iff the induced forms $\alpha_S, \beta_S, \gamma_S$ vanish. We denote by $\mathscr{I}$ the ideal of differential forms on $M$ which is generated by $\alpha, \beta, \gamma$. From the point of view of integral manifolds, the EDS (1) is completely determined by the associated ideal $\mathscr{I}$. Moreover, we will observe that $\mathscr{I}$ is closed, that is to say, $d\mathscr{I} \subset \mathscr{I}$.

Let $s: \mathbb{R}^2 \to M$ be a section of $\pi$. We denote $s^*u(x,t) = u(x,t); s^*z(x,t) = z(x,t); s^*p(x,t) = p(x,t)$. Then the image $S = s(\mathbb{R}^2)$ is an integral manifold of (1) iff one has

$$z(x, t) = u_x(x, t) \quad \text{and} \quad p(x, t) = u_{xx}(x, t),$$

where $u(x, t)$ is a solution of the KdV equation

$$u_t + u_{xxx} + 12u\, u_x = 0. \qquad (2)$$

## II. PROLONGATIONS BY CARTAN–EHRESMANN CONNECTIONS

Let us consider a locally trivial fibration $\tilde{\pi}: \widetilde{M} \to M$, with $F$ as typical fiber. A *Cartan–Ehresmann connection* on $(\widetilde{M}, \tilde{\pi}, M)$ is a field $H$ of horizontal contact elements on $\widetilde{M}$ which is supplementary of the field $V$ of the $\tilde{\pi}$-vertical contact elements. Moreover, one assumes that $H$ is complete, that is to say, every complete vector field $X$ on $M$ has a complete horizontal lift $\widetilde{X}$ on $\widetilde{M}$.

Let $\widetilde{\mathscr{H}}^*$ be the set of 1-forms on $\widetilde{M}$ which vanish on the field $H$. The ideal $\widetilde{\mathscr{I}}$ of differential forms on $\widetilde{M}$, which is generated by $\tilde{\pi}^*\mathscr{I} \cup \widetilde{\mathscr{H}}^*$, determines on $\widetilde{M}$ an EDS.

If, moreover, the ideal $\widetilde{\mathscr{I}}$ is closed, that is to say $d\widetilde{\mathscr{I}} \subset \widetilde{\mathscr{I}}$, then we will say that *the connection H is adapted to (1)*. In this case, the EDS on $\widetilde{M}$ defined by $\widetilde{\mathscr{I}}$ will be referred to as *the prolongation of (1) on $(\widetilde{M}, \tilde{\pi}, M)$ by the Cartan–Ehresmann connection H*. (See Ref. 5.)

The geometrical interpretation is the following one: if the connection $H$ is adapted to (1), then each integral manifold $S$ of (1) admits horizontal coverings in $\widetilde{M}$ which are integral manifolds of the prolonged system defined by $\widetilde{\mathscr{I}}$.

*Example*: Let us consider the case where $\widetilde{M}$ is the trivial bundle $M \times \mathbb{R}^q$ with global coordinates $(x, t, u, z, p, y^1, ..., y^q)$. Moreover we assume that the connection $H$ is defined by the system.

$$\omega^i \equiv dy^i - A^i\, dx - B^i\, dt = 0, \quad i = 1, ..., q, \qquad (3)$$

where $A^i$, $B^i$ are functions of $(x, t, u, z, p, y^1, ..., y^q)$.

Then, if $H$ is adapted to (1), it defines a *multiple pseudopotential* in the sense of Ref. 1. Wahlquist–Estabrook studied such particular prolongations, assuming moreover that $A^i$, $B^i$ do not have explicit $(x, t)$ dependence.

Our purpose is to study general prolongations of (1) by Cartan–Ehresmann (CE) connections without any particular assumption.

Besides, it is interesting to observe that the $(x, t)$-independence assumption has not an intrinsic signification: it is essentially related to the choice of a particular trivialization $\widetilde{M} \simeq M \times \mathbb{R}^q$.

## III. FOLIATED STRUCTURE AND ADAPTED COORDINATES IN $\widetilde{M}$

From now on, $(\widetilde{M}, \tilde{\pi}, M)$ is a locally trivial fiber bundle with $F$ as typical fiber; $H$ is a CE connection on $(\widetilde{M}, \tilde{\pi}, M)$ which is assumed to be adapted to (1).

Let us observe first that the submanifolds in $M$ defined by

$$x = \text{const}, \quad t = \text{const}$$

are integral manifolds of (1). Hence, the CE connection induced by $H$ over such a submanifold is integrable and defines a horizontal foliation.

By this way, we define on $\widetilde{M}$ an *H-horizontal foliation* $\widetilde{\mathscr{F}}$, the leaves of which are sections of $\widetilde{M}$ over the submanifolds defined by $x = \text{const}$, $t = \text{const}$.

Now, let us consider local coordinates $(x, t, u, z, p, y^1, ..., y^q)$ in $\widetilde{M}$ such that the foliation $\widetilde{\mathscr{F}}$ is locally defined by

$$dx = 0, \quad dt = 0, \quad dy^1 = 0, ..., dy^q = 0.$$

We will say that such local coordinates are *adapted to the foliation $\widetilde{\mathscr{F}}$*.

 0022-2488/84/072222-04$02.50  

With respect to such $\mathscr{F}$-adapted coordinates, the connection $H$ is defined by equations like (3):

$$dy^i = A^i\,dx + B^i\,dt, \quad i = 1, ..., q. \tag{4}$$

Moreover, if we denote by $\widetilde{A}, \widetilde{B}, \widetilde{U}, \widetilde{Z}, \widetilde{P}$ the $H$-horizontal vector fields on $\widetilde{M}$ whose respective projections in $M$ are $\partial/\partial x, \partial/\partial t, \partial/\partial u, \partial/\partial z, \partial/\partial p$, then, in $\mathscr{F}$-adapted local coordinates, we have

$$\widetilde{A} = \frac{\partial}{\partial x} + \sum_{i=1}^{q} A^i \frac{\partial}{\partial y^i}, \quad \widetilde{B} = \frac{\partial}{\partial t} + \sum_{i=1}^{q} B^i \frac{\partial}{\partial y^i},$$
$$\widetilde{U} = \frac{\partial}{\partial u}, \quad \widetilde{Z} = \frac{\partial}{\partial z}, \quad \widetilde{P} = \frac{\partial}{\partial p}. \tag{5}$$

Of course, these vector fields define completely the connection $H$.

## IV. THE CLOSURE CONDITION $d\mathscr{F} \subset \mathscr{F}$

If, in local $\mathscr{F}$-adapted coordinates, $H$ is defined by (4), then the closure condition $d\mathscr{F} \subset \mathscr{F}$ gives

$$A^i_z = A^i_p = 0, \quad B^i_p = -A^i_u,$$
$$\sum_{j=1}^{q} (A^j B^i_{y^j} - B^j A^i_{y^j}) + B^i_x - A^i_t + zB^i_u + pB^i_z + 12uzA^i_u = 0 \tag{6}$$
$$i = 1, ..., q.$$

In order to simplify conditions (6), we define

$$[\widetilde{U}, \widetilde{A}] = \widetilde{A}_u, \quad [\widetilde{Z}, \widetilde{A}] = \widetilde{A}_z, \quad [\widetilde{P}, \widetilde{A}] = \widetilde{A}_p,$$
$$[\widetilde{U}, \widetilde{B}] = \widetilde{B}_u, \quad [\widetilde{Z}, \widetilde{B}] = \widetilde{B}_z, \quad [\widetilde{P}, \widetilde{B}] = \widetilde{B}_p.$$

Then (6) becomes

$$\widetilde{A}_z = \widetilde{A}_p = 0, \quad \widetilde{A}_u = -\widetilde{B}_p,$$
$$[\widetilde{A}, \widetilde{B}] + z\widetilde{B}_u + p\widetilde{B}_z + 12uz\widetilde{A}_u = 0. \tag{7}$$

By a calculation which is essentially the same as in WE,[1] we obtain

$$\widetilde{A} = 2\widetilde{X}_1 + 2u\widetilde{X}_2 + 3u^2\widetilde{X}_3,$$
$$\widetilde{B} = (-2p - 12u^2)\widetilde{X}_2 + (-6up + 3z^2 - 24u^3)\widetilde{X}_3 - 4z\widetilde{X}_7 + 4u^2\widetilde{X}_6 + 8u\widetilde{X}_5 + 8\widetilde{X}_4, \tag{8}$$

where coefficients in $\widetilde{A}$ are introduced in accordance with notations of WE, and where vector fields $\widetilde{X}_1, ..., \widetilde{X}_7$ have to satisfy the following conditions:

$\widetilde{X}_2, \widetilde{X}_3, \widetilde{X}_7, \widetilde{X}_5, \widetilde{X}_6$ are $\tilde{\pi}$-vertical; (9a)

$\widetilde{X}_1, \widetilde{X}_4$ are $\tilde{\pi}$-projectable, respectively, on

$$\frac{1}{2}\frac{\partial}{\partial x} \text{ and } \frac{1}{8}\frac{\partial}{\partial t}; \tag{9b}$$

$\widetilde{X}_1, ..., \widetilde{X}_7$ commute with $\widetilde{U}, \widetilde{Z}, \widetilde{P}$; (9c)

$$[\widetilde{X}_1, \widetilde{X}_3] = [\widetilde{X}_2, \widetilde{X}_3] = [\widetilde{X}_2, \widetilde{X}_6] = [\widetilde{X}_1, \widetilde{X}_4] = 0,$$
$$[\widetilde{X}_1, \widetilde{X}_2] = -\widetilde{X}_7, \quad [\widetilde{X}_1, \widetilde{X}_7] = \widetilde{X}_5, \quad [\widetilde{X}_2, \widetilde{X}_7] = \widetilde{X}_6, \tag{9d}$$
$$[\widetilde{X}_1, \widetilde{X}_5] + [\widetilde{X}_2, \widetilde{X}_4] = \widetilde{X}_7 + [\widetilde{X}_3, \widetilde{X}_4] + [\widetilde{X}_1, \widetilde{X}_6] = 0.$$

We observe that the vector fields $\widetilde{X}_2, \widetilde{X}_3, \widetilde{X}_5, \widetilde{X}_6, \widetilde{X}_7$ are precisely the vector fields $X_2, X_3, X_5, X_6, X_7$ introduced in WE, while $X_1, X_4$ have horizontal components, the introduction of which allows us to avoid the (unintrinsic) assumption of $(x, t)$ independence.

## V. GENERAL PROLONGATIONS OF (1) AND GEOMETRIC REALIZATIONS OF THE WE PARTIAL LIE ALGEBRA

Let us denote by $L$ a seven-dimensional $\mathbb{R}$-vector space with basis $\{\xi_1, ..., \xi_7\}$ and partial Lie algebra structure defined by

$$[\xi_1, \xi_3] = [\xi_2, \xi_3] = [\xi_2, \xi_6] = [\xi_1, \xi_4] = 0,$$
$$[\xi_1, \xi_2] = -\xi_7, \quad [\xi_1, \xi_7] = \xi_5, \quad [\xi_2, \xi_7] = \xi_6, \tag{10}$$
$$[\xi_1, \xi_5] + [\xi_2, \xi_4] = \xi_7 + [\xi_3, \xi_4] + [\xi_1, \xi_6] = 0.$$

We will say that $L$ is *the* WE *partial Lie algebra*. We denote by $A$ the subspace generated by $\{\xi_1, \xi_4\}$ and by $B$ the subspace generated by $\{\xi_2, \xi_3, \xi_5, \xi_6, \xi_7\}$. The subspace $A$ has the structure of an abelian Lie algebra.

If $M_0$ is a $C^\infty$ manifold, a pair $(\mathscr{L}, \varphi)$ is a *geometrical realization of* $L$ *in* $M_0$ if the following hold.

(i) $\mathscr{L} = \mathscr{A} \oplus \mathscr{B}$ is a transitive Lie algebra of vector fields in $M_0$, where $\mathscr{A}$, $\mathscr{B}$ are subalgebras whose values at each point of $M_0$ define supplementary contact elements, with $[\mathscr{A}, \mathscr{B}] \subset \mathscr{B}$.

(ii) $\varphi: L \to \mathscr{L}$ is a $\mathbb{R}$-linear homomorphism compatible with (partial) Lie algebra structures, and such that

$$\ker \varphi \cap A = \{0\}, \quad \varphi(A) = \mathscr{A}, \quad \varphi(B) \subset \mathscr{B}.$$

(iii) Vector fields in $\mathscr{A}$ are complete and linearly independent.

Now, returning to the situation in Sec. III, let us denote by $(\widetilde{M}_0, \tilde{\pi}_0, \mathbb{R}^2)$ the locally trivial fiber bundle on $\mathbb{R}^2$ induced from $(\widetilde{M}, \tilde{\pi}, M)$ by the section $s_0$: $\mathbb{R}^2 \to M$ defined by

$$s_0(x, t) = (x, t, 0, 0, 0).$$

The projection of $\widetilde{M}$ onto $\widetilde{M}_0$ along the leaves of $\widetilde{\mathscr{F}}$ allows us to identify

$$\widetilde{M} = \widetilde{M}_0 \times \mathbb{R}^3, \tag{11}$$

where coordinates in $\mathbb{R}^3$ are $(u, z, p)$.

Moreover, $\widetilde{X}_1, ..., \widetilde{X}_7$ induce vector fields $\widetilde{X}_{10}, ..., \widetilde{X}_{70}$ on $\widetilde{M}_0$ whose knowledge completely determines $\widetilde{X}_1, ..., \widetilde{X}_7$, thus $H$.

If $\widetilde{A}$ is the Lie algebra of vector fields in $\widetilde{M}_0$ generated by $\widetilde{X}_{10}, \widetilde{X}_{40}$, $\widetilde{\mathscr{B}}$ the Lie algebra of $\tilde{\pi}_0$-vertical vector fields, $\widetilde{\mathscr{L}} = \widetilde{\mathscr{A}} \oplus \widetilde{\mathscr{B}}$, $\tilde{\varphi}$ the $\mathbb{R}$-linear homomorphism $L \to \widetilde{\mathscr{L}}$ determined by

$$\tilde{\varphi}(\xi_i) = \widetilde{X}_{i0}, \quad i = 1, ..., 7,$$

then $(\widetilde{\mathscr{L}}, \tilde{\varphi})$ is a geometrical realization of $L$ and we obtain the following theorem.

**Theorem I:** Each prolongation of (1) by a Cartan–Ehresmann connection determines a geometrical realization of the WE partial Lie algebra $L$. Conversely, every geometrical realization of $L$ corresponds to such a prolongation.

In order to prove the second part of this result, let us consider a geometrical realization $(\mathscr{L}, \varphi)$ of $L$ on a manifold $M_0$. If $\mathscr{L} = \mathscr{A} \oplus \mathscr{B}$, orbits of the subalgebra $\mathscr{B}$ define a codimension 2 foliation $\mathscr{F}(\mathscr{B})$ on $M_0$. Moreover condition $[\mathscr{A}, \mathscr{B}] \subset \mathscr{B}$ implies that $\mathscr{A}$ is a Lie algebra of commuting foliate vector fields. Hence $\mathscr{F}(\mathscr{B})$ is a $\mathbb{R}^2$-Lie foliation in the sense of Fedida.[3] From Ref. 3 one knows that the pullback $\widetilde{\mathscr{F}}(\mathscr{B})$ of $\mathscr{F}(\mathscr{B})$ on a covering manifold $\widetilde{M}_0$ of $M_0$ is a simple foliation which (in accordance with completeness of foliate vector fields in $\mathscr{A}$) corresponds to a locally trivial fibration

$\tilde{\pi}_0$: $\widetilde{M}_0 \to \mathbb{R}^2$ such that $\varphi(\xi_1)$ and $\varphi(\xi_4)$ define $\tilde{\pi}_0$-projectable vector fields on $\widetilde{M}_0$ whose respective projections are $\frac{1}{2}\partial/\partial x$ and $\frac{1}{8}\partial/\partial t$.

From $(\mathscr{L}, \varphi)$ we obtain a covering geometrical realization $(\widetilde{\mathscr{L}}, \tilde{\varphi})$ of $L$ in $\widetilde{M}_0$. Now, if $\widetilde{M} = \widetilde{M}_0 \times \mathbb{R}^3$, where $\mathbb{R}^3$ has $(u, z, p)$ as natural coordinates, we denote by $\tilde{\pi}$: $\widetilde{M} \to \mathbb{R}^5$ the projection

$$\tilde{\pi} = \tilde{\pi}_0 \times \mathbf{1}_{\mathbb{R}^3}.$$

If $\tilde{m} = (\tilde{m}_0, m) \in \widetilde{M}_0 \times \mathbb{R}^3$, one has a natural identification

$$T_{\tilde{m}}\widetilde{M} = T_{\tilde{m}_0}\widetilde{M}_0 \oplus T_m \mathbb{R}^3, \tag{12}$$

and we will define vector fields $\widetilde{U}, \widetilde{Z}, \widetilde{P}, \widetilde{X}_1, ..., \widetilde{X}_7$ in $\widetilde{M}$ by

$$\widetilde{U}_{\tilde{m}} = 0 + \left.\frac{\partial}{\partial u}\right|_m, \quad \widetilde{Z} = 0 + \left.\frac{\partial}{\partial z}\right|_m, \quad \widetilde{P} = 0 + \left.\frac{\partial}{\partial p}\right|_m,$$

$$\widetilde{X}_{i\tilde{m}} = \tilde{\varphi}(\xi_i)_{\tilde{m}_0} + 0, \quad i = 1, ..., 7.$$

Equations (5) and (8) define vector fields $\widetilde{A}, \widetilde{B}$ in $\widetilde{M}$. If $H$ is the CE connection on $(\widetilde{M}, \tilde{\pi}, \mathbb{R}^5)$ which admits $\widetilde{A}, \widetilde{B}, \widetilde{U}, \widetilde{Z}, \widetilde{P}$ as horizontal vector fields, then $H$ is adapted to (1). Q.E.D.

*Remark 1:* $\{\widetilde{X}_1, \widetilde{X}_4, \widetilde{U}, \widetilde{Z}, \widetilde{P}\}$ define on $(M, \pi, M)$ on integrable connection. Thus they determine on $(\widetilde{M}, \tilde{\pi}, M)$ a *global foliate trivialization*.

*Remark 2:* Results of Wahlquist–Estabrook in Ref. 1 correspond to the following particular case: Let $X_1, ..., X_7$ be vector fields on a manifold $F$ such that

$$[X_1, X_3] = [X_2, X_3] = [X_2, X_6] = [X_1, X_4] = 0,$$

$$[X_1, X_2] = -X_7, \quad [X_1, X_7] = X_5, \quad [X_2, X_7] = X_6,$$

$$[X_1, X_5] + [X_2, X_4] = X_7 + [X_3, X_4] + [X_1, X_6] = 0.$$

Now, let us consider $M_0 = F \times \mathbb{R}^2$ with the natural identification

$$T_{(f,x,t)}M_0 = T_f F \oplus T_{(x,t)}\mathbb{R}^2.$$

We will define vector fields $\varphi(\xi_i)$, $i = 1, ..., 7$ on $M_0$ by

$$\varphi(\xi_1)_{(f,x,t)} = X_{1f} + \frac{1}{2}\left.\frac{\partial}{\partial x}\right|_{(x,t)},$$

$$\varphi(\xi_4)_{(f,x,t)} = X_{4f} + \frac{1}{8}\left.\frac{\partial}{\partial t}\right|_{(x,t)},$$

$$\varphi(\xi_i)_{(f,x,t)} = X_{if} + 0, \quad i = 2,3,5,6,7.$$

If $\pi_0$: $M_0 \to \mathbb{R}^2$ is the second projection and $\mathscr{B}$ is the Lie algebra of $\pi_0$-vertical vector fields, we will denote by $\mathscr{A}$ the abelian Lie algebra generated by $\{\varphi(\xi_1), \varphi(\xi_4)\}$ and by $\mathscr{L}$ the Lie algebra $\mathscr{A} \oplus \mathscr{B}$. Then $(\mathscr{L}, \varphi)$ is a geometrical realization of $L$ in $M_0$. Moreover, we have

$$\left(\frac{\partial}{\partial x}, \varphi(\xi_i)\right) \equiv \left(\frac{\partial}{\partial t}, \varphi(\xi_i)\right) \equiv 0, \quad i = 1, ..., 7.$$

This fact corresponds to the assumption of "$(x, t)$ independence" of the prolongation with respect to the trivialization $M_0 = F \times \mathbb{R}^2$.

From an intrinsic point of view, the existence of such a trivialization is equivalent to the existence of two-dimensional Lie algebra $\mathscr{A}'$ of vector fields which commutes with $\mathscr{L}$ and such that values of $\mathscr{A}'$ and $\mathscr{B}$ at every point define supplementary contact elements.

*Remark 3:* Let us give an example of $(x, t)$-depending prolongation: $\widetilde{M} = M \times \mathbb{R}$ with global coordinates $(x, t, u, z, p, y)$ and we consider the vector fields

$$\widetilde{U} \equiv \frac{\partial}{\partial u}, \quad \widetilde{Z} \equiv \frac{\partial}{\partial z}, \quad \widetilde{P} \equiv \frac{\partial}{\partial p},$$
$$\widetilde{X}_1 \equiv \frac{1}{2}\frac{\partial}{\partial x}, \quad \widetilde{X}_4 \equiv \frac{1}{8}\frac{\partial}{\partial t}, \tag{13}$$
$$\widetilde{X}_2 \equiv -2x\frac{\partial}{\partial y}, \quad \widetilde{X}_3 \equiv 8t\frac{\partial}{\partial y}, \quad \widetilde{X}_7 \equiv \frac{\partial}{\partial y}, \quad \widetilde{X}_5 \equiv \widetilde{X}_6 \equiv 0.$$

By this way, we obtain the $(x, t)$-depending potential

$$dy = (-xu + 6u^2 t)dx + (px + 6u^2 x - 12upt + 6z^2 t - 48u^3 t - z)dt. \tag{14}$$

## VI. CLASSIFICATION OF SIMPLE ANALYTIC PSEUDOPOTENTIALS

In this section, differentiability will be assumed to be real analytic. We study the case $F = \mathbb{R}$ (simple pseudopotentials).

In this case, by analycity, vector fields $\widetilde{X}_1, ..., \widetilde{X}_7$ on $\widetilde{M}$ are real analytic. Classification will be done by the following arguments (see details in Ref. 4)

(a) If $\widetilde{X}_3 \not\equiv 0$, let $\widetilde{\Omega} = \{\tilde{m} \in \widetilde{M} / \widetilde{X}_{3\tilde{m}} \neq 0\}$, where $\widetilde{\Omega}$ is an open dense set in $\widetilde{M}$.

If $\tilde{m} \in \widetilde{M}$, there exist, in a neighborhood of $\tilde{m}$, $\widetilde{\mathscr{F}}$-adapted local coordinates $(x, t, u, z, p, y)$ such that $\widetilde{X}_3 \equiv \partial/\partial y$. Equation (9d) gives

$$\widetilde{X}_2 \equiv [\alpha(t)x + \beta(t)]\frac{\partial}{\partial y}, \quad \text{with } \begin{cases} 4\alpha^2 - \alpha_t = 0, \\ 4\alpha\beta - \beta_t = 0. \end{cases} \tag{15}$$

Thus, either $\alpha = 0, \beta = \lambda$ or $\alpha = -1/4(t - t_0), \beta = x_0/4(t - t_0)$.

*If* $\alpha = 0, \beta = \lambda$, we obtain relations

$$\widetilde{X}_2 \equiv \lambda\widetilde{X}_3, \quad \widetilde{X}_5 \equiv \widetilde{X}_6 \equiv \widetilde{X}_7 \equiv 0,$$

which are true in $\widetilde{\Omega}$, thus, by analyticity, in $\widetilde{M}$. Now, using the integrable connection whose horizontal elements are generated by $\{\widetilde{X}_1, \widetilde{X}_4, \widetilde{U}, \widetilde{Z}, \widetilde{P}\}$, we obtain global $\widetilde{\mathscr{F}}$-adapted coordinates such that $H$ is defined by

$$dy = \varphi(y)[(2\lambda u + 3u^2)dx + (-2\lambda p - 12u^2\lambda - 6up + 3z^2 - 24u^3)dt], \tag{16}$$

where $\lambda \in \mathbb{R}$ and $\varphi$ is an arbitrary analytic function.

If $\alpha = -1/4(t - t_0), \beta = x_0/4(t - t_0)$, we obtain relations

$$4(t - t_0)\widetilde{X}_2 + (x - x_0)\widetilde{X}_3 \equiv 0, \ \widetilde{X}_5 \equiv \widetilde{X}_6 \equiv 0, \ 8(t - t_0)\widetilde{X}_7 \equiv \widetilde{X}_3$$

which are true in $\widetilde{\Omega}$, thus in $\widetilde{M}$. By the previous argument, we obtain global $\widetilde{\mathscr{F}}$-adapted coordinates such that $H$ is defined by

$$dy = \varphi(y)\{[-u(x - x_0) + 6u^2(t - t_0)]dx + [(p + 6u^2)(x - x_0) + (-12up + 6z^2 - 48u^3)(t - t_0) - z]dt\}, \tag{17}$$

where $\lambda \in \mathbb{R}$ and $\varphi$ is an arbitrary analytic function

(b) If $\widetilde{X}_3 \equiv 0$, $\widetilde{X}_2 \not\equiv 0$, let $\widetilde{\Omega} = \{\tilde{m} \in \widetilde{M} / \widetilde{X}_{2\tilde{m}} \neq 0\}$. Here, $\widetilde{\Omega}$ is an open dense set in $\widetilde{M}$. If $\tilde{m} \in \widetilde{\Omega}$, we use $\widetilde{\mathscr{F}}$-adapted local coordinates in a neighborhood of $\tilde{m}$ such that $\widetilde{X}_2 \equiv \partial/\partial y$. Then, by (9d), we obtain

$$\tilde{X}_1 \equiv \frac{1}{2}\frac{\partial}{\partial x} + [\alpha(x,t)y^2 + \beta(x,t)y + \gamma(x,t)]\frac{\partial}{\partial y}, \quad (18)$$

with either $\alpha \equiv \beta \equiv 0$ or $\alpha \equiv \frac{1}{2}$.

If $\alpha \equiv \beta \equiv 0$, we have $\tilde{X}_5 \equiv \tilde{X}_6 \equiv \tilde{X}_7 \equiv 0$ and by the previous argument we obtain global $\tilde{\mathscr{F}}$-adapted coordinates such that $H$ is defined by

$$dy = \varphi(y)[2u\,dx + (-2p - 12u^2)dt]. \quad (19)$$

If $\alpha \equiv \frac{1}{2}$, by a change of local coordinates of the form $y' = y + \beta(x,t)$, we obtain the local reduced expressions

$$\tilde{X}_1 \equiv \frac{1}{2}\frac{\partial}{\partial x} + \left(\frac{1}{2}y^2 + \lambda\right)\frac{\partial}{\partial y}, \quad \tilde{X}_2 \equiv \tilde{X}_6 \equiv \frac{\partial}{\partial y}, \quad \tilde{X}_7 = y\frac{\partial}{\partial y},$$
$$\tilde{X}_4 \equiv \frac{1}{8}\frac{\partial}{\partial t} + (\lambda y^2 + 2\lambda^2)\frac{\partial}{\partial y}, \quad \tilde{X}_5 \equiv \left(-\frac{1}{2}y^2 + \lambda\right)\frac{\partial}{\partial y}, \quad (20)$$

and a slightly more sophisticated version of the previous argument (see Ref. 4) shows that there exist global $\tilde{F}$-adapted coordinates such that $H$ is defined by

$$dy = \varphi_1(y)[(2\lambda + 2u)dx + (-2p - 8u^2 + 8u\lambda + 16\lambda^2)dt] + \varphi_2(y)[-4zdt] + \varphi_3(y)[dx + (-4u + 8\lambda)dt], \quad (21)$$

where $\lambda \in \mathbb{R}$ and $X_i = \varphi_i(y)\partial/\partial y$ are analytic vector fields on $\mathbb{R}$ such that

$$[X_1, X_2] = X_1, \quad [X_1, X_3] = 2X_2, \quad [X_2, X_3] = X_3.$$

Finally, we obtain the following theorem.

**Theorem II**: Let $(\tilde{M}, \tilde{\pi}, M)$ be an analytic locally trivial fiber bundle with $\mathbb{R}$ as a typical fiber. If $H$ is an analytic CE connection on $(\tilde{M}, \tilde{\pi}, M)$ which is adapted to (1), then there exists a global analytic trivialization $\tilde{M} = M \times \mathbb{R}$ such that, in the corresponding coordinates $(x, t, u, z, p, y)$, $H$ is defined by one of the following equations.

(i) $$dy = \varphi(y)\{[\lambda_1 u + \lambda_2 u^2 + \lambda_3(-ux + 6u^2 t)]dx + [\lambda_1(-p - 6u^2) + \lambda_2(-2up + z^2 - 8u^3) + \lambda_3(px + 6u^2 x - 12upt + 6z^2 t - 48u^3 t - z)]dt\},$$

where $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and $\varphi$ is an arbitrary analytic function on $\mathbb{R}$.

(ii) $$dy = \varphi_1(y)[(2\lambda + 2u)dx + (-2p - 8u^2 + 8u\lambda + 16\lambda^2)dt] + \varphi_2(y)[-4zdt] + \varphi_3(y)[dx + (-4u + 8\lambda)dt],$$

where $\lambda \in \mathbb{R}$ and $X_i = \varphi_i(y)\partial/\partial y$, $i = 1, 2, 3$ are analytic vector fields on $\mathbb{R}$ such that

$$[X_1, X_2] = X_1, \quad [X_1, X_3] = 2X_2, \quad [X_2, X_3] = X_3.$$

In the case (i), if $\varphi(y) \equiv 1$, we obtain a potential with three independent parameters.

In the case (ii), if $\varphi_1(y) \equiv 1$, $\varphi_2(y) \equiv y$, and $\varphi_3(y) \equiv y^2$, we obtain essentially the pseudopotential discovered by Wahlquist–Estabrook.[1]


[1]H. D. Wahlquist and F. B. Estabrook, J. Math. Phys. **16**, 1 (1975).
[2]E. Cartan, *Les systèmes différentiels extérieurs et leurs applications géométriques* (Hermann, Paris, 1945).
[3]E. Fedida, C. R. Acad. Sci., Paris, **272**, 999 (1971).
[4]P. Molino, "Simple pseudopotentials for the KdV equation," in *Lecture Notes in Mathematics* (Springer, Berlin, 1982), Vol. 926.
[5]R. Hermann, "Geometric theory of nonlinear differential equations, Bäcklund transformations and solitons," in *Interdisciplinary Mathematics*, (Math. Sci. Press, Brookline, MA, 1976), Vols. XII and XIV.

# The sine-Gordon equations: Complete and partial integrability

John Weiss
*La Jolla Institute, La Jolla, California 92037 and Institute for Pure and Applied Physical Science, University of California, San Diego, La Jolla, California 92093*

(Received 30 November 1983; accepted for publication 24 February 1984)

The sine–Gordon equation in one space-one time dimension is known to possess the Painlevé property and to be completely integrable. It is shown how the method of "singular manifold" analysis obtains the Bäcklund transform and the Lax pair for this equation. A connection with the sequence of higher-order KdV equations is found. The "modified" sine–Gordon equations are defined in terms of the singular manifold. These equations are shown to be identically Painlevé. Also, certain "rational" solutions are constructed iteratively. The double sine–Gordon equation is shown not to possess the Painlevé property. However, if the singular manifold defines an "affine minimal surface," then the equation has integrable solutions. This restriction is termed "partial integrability." The sine–Gordon equation in $(N+1)$ variables ($N$ space, 1 time) where $N$ is greater than one is shown not to possess the Painlevé property. The condition of partial integrability requires the singular manifold to be an "Einstein space with null scalar curvature." The known integrable solutions satisfy this constraint in a trivial manner. Finally, the coupled KdV, or Hirota–Satsuma, equations possess the Painlevé property. The associated "modified" equations are derived and from these the Lax pair is found.

PACS numbers: 03.40.Kf, 02.30. + g

## I. INTRODUCTION

In Ref. 1 the Painlevé property for partial differential equations was defined. Briefly, we say that a partial differential equation has the Painlevé property when the solutions of the pde are "single valued" about the movable singularity manifold. To be precise, if the singularity manifold is determined by

$$\varphi(z_1,z_2,...,z_n)=0, \tag{1.1}$$

and $u=u(z_1,...,z_n)$ is a solution of the pde, then we require that

$$u=\varphi^{\alpha}\sum_{j=0}^{\infty}u_j\varphi^j, \tag{1.2}$$

where $u_0\neq 0$, $\varphi=\varphi(z_1,...,z_n)$, $u_j=u_j(z_1,...,z_n)$ are analytic functions of $(z_j)$ in a neighborhood of the manifold (1.1) and $\alpha$ is a negative, rational number. Substitution of (1.2) into the pde determines the allowed values of $\alpha$, and defines the recursion relations for $u_j$, $j=0,1,2,...$ . When the anzatz (1.2) is correct the pde is said to possess the Painlevé property and is conjectured to be integrable.

In Ref. 2 Bäcklund transformations were obtained by truncating the expansion (1.2) at the "constant" level term. That is, we set

$$u=u_0\varphi^{-N}+u_1\varphi^{-N+1}+\cdots+u_N, \tag{1.3}$$

and find, from the recursion relations for $u_j$, an overdetermined system of equations for $(\varphi,u_j,j=0,1,...,N)$, where $u_N$ will satisfy the (original) pde. Upon solving the overdetermined system it was found, for those equations considered, that $\varphi$ satisfied an equation formulated in terms of the Schwarzian derivative:

$$\{\varphi;x\}=\frac{\partial}{\partial x}\left(\frac{\varphi_{xx}}{\varphi_x}\right)-\frac{1}{2}\left(\frac{\varphi_{xx}}{\varphi_x}\right)^2. \tag{1.4}$$

The invariance of (1.4) under the Moebius group,

$$\varphi=(a\psi+b)/(c\psi+d),\quad \{\varphi;x\}=\{\psi;x\}, \tag{1.5}$$

motivates the substitution

$$\varphi=V_1/V_2, \tag{1.6}$$

where $V_1$ and $V_2$ satisfy the same linear equation. From the resulting Wronskian relations the Lax pair may be found.

In Ref. 3 it is shown how study of the Caudrey–Dodd–Gibbon equation leads to the formulation of a class of equations, in terms of the Schwarzian derivative, that identically possess the Painlevé property. This class of equations contains the higher-order KdV, Caudrey–Dodd–Gibbon, and Kuperschmidt equations.

In this paper various equations of sine–Gordon type are considered. These equations are somewhat different from those studied previously in that they have a symmetric dependence on the independent variables (under Lorenz transformation). Only the $(1+1)$ sine–Gordon (one space-one time variable) equation is found to identically posssess the Painlevé property. The method of "singular manifold" analysis, i.e., Bäcklund transform and formulation in terms of the Schwarzian derivative, obtains, for this equation, the Lax pair. In addition, a connection to the sequence of higher-order KdV equations is found. That is, the $(1+1)$ sine–Gordon equation is formulated in terms of "minus one" functional of the Lenard recursion relations, where positive functionals determine the sequence of higher-order KdV equations. For the sine–Gordon equation we define a system of "modified" equations that identically possess the Painlevé property. These "modified" equations are related to the "characteristic" initial value problem. Furthermore, we find, using the discrete symmetries of the modified equations, certain rational solutions of the sine–Gordon equation.

The double sine–Gordon and $(N+1)$ sine–Gordon equations are found not to possess the Painlevé property. This would seem to answer various questions concerning the integrability of these equations.[4–10] However, if the "singular manifold," $\varphi$, in the *Ansatz* (1.2) is restricted (to satisfy a subsidiary constraint) a type of "partial" integrability can be defined for these equations. The known, exact solutions appear to satisfy the appropriate constraint in a more or less trivial manner. We conjecture that the class of exact solutions (for these equations) is more general. Hopefully, study of the "constrained" dynamics wil lead to their discovery.

In a recent paper, Oevel[11] states that the coupled KdV, or Hirota–Satsuma, equations "do not seem to be 'completely integrable' in the usual sense." Analysis reveals that these equations identically possess the Painlevé property. Thus, if these equations are "partially integrable" it is in a different sense from that defined above. The Painlevé ("singular manifold") analysis is presented in the Appendices.

We note that "partial" integrability (of various types) for ordinary differential equations has been considered by several authors, i.e., Segur[12] and Tabor and Weiss.[13]

## II. THE (1 + 1) SINE–GORDON EQUATION

An interesting discussion of the long history of the (1 + 1) sine–Gordon equation

$$u_{xt} = \sin u \tag{2.1}$$

can be found in Chap. 1 of Ref. 14. Suffice it to say that the original Bäcklund transformation[15] was defined for this equation, while the Lax pair is contained in the inverse scattering transforms of Zakharov and Shabat[16] and Ablowitz *et al.*[17]

In Ref. 1 the sine–Gordon equation was shown to possess the Painlevé property. For reference, we present part of the analysis here.

Since the nonlinearity of (2.1) is nonalgebraic it is convenient to transform Eq. (2.1) into a different form. That is, let

$$V = e^{iu}, \tag{2.2}$$

and find

$$VV_{xt} - V_x V_t = \tfrac{1}{2}(V^3 - V). \tag{2.3}$$

By a leading order and resonance analysis this equation has an expansion

$$V = \varphi^{-2} \sum_{j=0}^{\infty} V_j \varphi_j, \tag{2.4}$$

where the "resonances" occur at

$$j = -1, 2, \tag{2.5}$$

and

$$V_0 = 4\varphi_x\varphi_t, \quad V_1 = -4\varphi_{xt}. \tag{2.6}$$

The compatibility condition at $j = 2$ is satisfied identically ($u_2$ is arbitrary) and (2.3) and (2.1) possesses the Painlevé property.[1]

To proceed further, we now define the transform

$$V = \varphi^{-2}V_0 + \varphi^{-1}V_1 + V_2, \tag{2.7}$$

or, using (2.6),

$$V = -4\frac{\partial^2}{\partial x\,\partial t}\ln\varphi + V_2. \tag{2.8}$$

Substitution of (2.7) and (2.8) into Eq. (2.3) obtains an overdetermined system of equations for $(\varphi_0, V_2)$. This system arises from the recursion relations for the $V_j$ and the requirement that

$$V_3 = V_4 = V_5 = V_6 = 0, \tag{2.9}$$

where $V_0$ and $V_1$ are defined by (2.6) and the condition $V_6 = 0$ requires $V_2$ to satisfy Eq. (2.2). There is no condition when $j = 2$ since this is a resonance of the recursion relations.

To effect the reduction of the system (2.9) of four equations in two unknowns to the Lax pair for Eq. (2.2) involves extensive calculation. To simplify the calculation it is convenient to let

$$Y_2 = W + \varphi_{xt}^2/\varphi_x\varphi_t. \tag{2.10}$$

The reason for this is as follows. Under the inversion,

$$\varphi = 1/\psi, \tag{2.11}$$

$$V_2 = -4\frac{\partial^2}{\partial x\,\partial t}\ln\psi + V, \tag{2.12}$$

and the form

$$W = V_2 - \varphi_{xt}^2/\varphi_x\varphi_t \tag{2.13}$$

becomes

$$W = V - \psi_{xt}^2/\psi_x\psi_t. \tag{2.14}$$

This invariance of $W$ under (2.11) is a useful check on the calculation.

We then recast the overdetermined (2.9) in the variables $(W,\varphi)$ into a form that is, insofar as possible, invariant under the transformation (2.11). The resulting equations involve $W$, $W_x$, $W_t$, etc. and the expressions

$$\varphi_x\frac{\partial}{\partial x}\Omega_1 + \varphi_t\frac{\partial}{\partial t}\Omega_2, \tag{2.15}$$

and

$$\Omega_1\Omega_2 - \tfrac{1}{4}, \tag{2.16}$$

where

$$\Omega_1 = \frac{\varphi_{xtt}}{\varphi_x} - \frac{\varphi_{tt}\varphi_{xt}}{\varphi_x\varphi_t} - \frac{1}{2}\frac{\varphi_{xt}^2}{\varphi_x^2}, \tag{2.17}$$

$$\Omega_2 = \frac{\varphi_{xxt}}{\varphi_t} - \frac{\varphi_{xx}\varphi_{xt}}{\varphi_x\varphi_t} - \frac{1}{2}\frac{\varphi_{xt}^2}{\varphi_t^2}. \tag{2.18}$$

The forms $\Omega_1$ and $\Omega_2$ are similar to the Schwarzian derivative (1.4) in that they are invariant under the Moebius group (1.5).

Now, from the system (2.9) we find the "reduced" system of equations

$$W = 0 \quad\text{or}\quad V_2 = \frac{\varphi_{xt}^2}{\varphi_x\varphi_t}, \tag{2.19}$$

$$\varphi_x\frac{\partial}{\partial x}\Omega_1 + \varphi_t\frac{\partial}{\partial t}\Omega_2 = 0, \tag{2.20}$$

and

$$\Omega_1\Omega_2 = \tfrac{1}{4}. \tag{2.21}$$

The system of two equations [(2.20) and (2.21)] in one unknown ($\varphi$) can be reduced further by using the identity

$$\varphi_x\frac{\partial}{\partial x}\Omega_1 = \varphi_t\frac{\partial}{\partial t}\Omega_2. \tag{2.22}$$

Thus, there results

$$\Omega_1 = \alpha, \quad \Omega_2 = \beta, \tag{2.23}$$

where

$$\alpha\beta = \tfrac{1}{4}. \tag{2.24}$$

We now let

$$Z^2 = \varphi_x/\varphi_t, \tag{2.25}$$

$$W^2 = \varphi_t/\varphi_x, \tag{2.26}$$

and find

$$\Omega_1 = \{\varphi;t\} + 2Z_{tt}/Z = \alpha, \tag{2.27}$$

$$\Omega_2 = \{\varphi;x\} + 2W_{xx}/W = \beta, \tag{2.28}$$

where $\alpha\beta = \frac{1}{4}$ and $\{\varphi;x\}$, $\{\varphi;t\}$ are Schwarzian derivatives.

To find the Lax pair we now assume that

$$\varphi = Y_1/Y_2, \tag{2.29}$$

where $Y_1$ and $Y_2$ satisfy

$$Y_{xx} = aY,$$

and (2.30)

$$Y_t = bY_x + cY.$$

By the condition

$$Y_{xxt} = Y_{txx}, \tag{2.31}$$

it is found that

$$2c_x + b_{xx} = 0, \tag{2.32}$$

$$a_t = -b_{xxx}/2 + 2ab_x + ba_x. \tag{2.33}$$

By the Wronskian relation for (2.30),

$$W^2 = Z^{-2} = b, \tag{2.34}$$

and

$$\{\varphi;x\} = -2a. \tag{2.35}$$

Evaluating Eq. (2.28), we find

$$a = \frac{1}{2}\left(\frac{b_{xx}}{b} - \frac{1}{2}\frac{b_x^2}{b^2}\right) - \frac{\beta}{2}, \tag{2.36}$$

and substitution into Eq. (2.33) obtains

$$a_t = -\beta b_x. \tag{2.37}$$

On the other hand, evaluation of (2.27 obtains

$$b_{xt} + bb_{xx} - b_t b_x/b - \tfrac{1}{2}b_x^2 - 2b^2 a = \alpha. \tag{2.38}$$

Using Eq. (2.36),

$$b_{xt} - b_t b_x/b = \alpha - \beta b^2. \tag{2.39}$$

We now let

$$\alpha = -\lambda^{-1}/4, \quad \beta = -\lambda, \quad b = (\lambda^{-1}/2)\Theta, \tag{2.40}$$

and find that $\Theta$ satisfies the equation

$$\Theta_{xt}/\Theta - \Theta_x\Theta_t/\Theta^2 = \tfrac{1}{2}(\Theta - \Theta^{-1}), \tag{2.41}$$

which is Eq. (2.2).

Now substitution of (2.36) into (2.37) produces

$$\frac{\partial}{\partial x}\left(\frac{b_{xt}}{b} - \frac{b_x b_t}{b^2}\right) + \frac{b_x}{b}\left(\frac{b_{xt}}{b} - b_x\frac{b_t}{b^2}\right) = -2\beta b_x, \tag{2.42}$$

or, by (2.39),

$$\frac{\partial}{\partial x}\left(\frac{b_{xt}}{b} - \frac{b_x b_t}{b^2} - \alpha b^{-1} + \beta b\right) = 0. \tag{2.43}$$

Thus, Eqs. (2.39) and (2.37) are consistent, and (2.30), (2.36), and (2.40) define the Lax pair for Eq. (2.41) or (2.2).

Having reduced (2.9) to the Lax pair for Eq. (2.2) and, thus, effectively defining the Bäcklund transform (2.8), we next consider some consequence for this reduction.

Taking into account the various scalings,

$$a = \frac{1}{2}\left(\frac{\Theta_{xx}}{\Theta} - \frac{1}{2}\frac{\Theta_x^2}{\Theta^2}\right) + \frac{\lambda}{2}. \tag{2.44}$$

In the scattering problem (2.30) $\lambda$ is the spectral parameter and

$$d = \frac{1}{2}\left(\frac{\Theta_{xx}}{\Theta} - \frac{1}{2}\frac{\Theta_x^2}{\Theta^2}\right), \tag{2.45}$$

where $\lim_{|x|\to\infty} d = 0$, is the (in general, complex) "potential."

From (2.45),

$$2\Theta\Theta_{xx} - \Theta_x^2 - 4\,d\Theta^2 = 0, \tag{2.46}$$

and differentiating with respect to $x$,

$$\Theta_{xxx} - 4\,d\Theta_x - 2\,d_x\Theta = 0. \tag{2.47}$$

Now formally, the Lenard recursion relations[18] are

$$\psi_{n+1,x} = -\psi_{n,xxx} + 4\,d\psi_{n,x} + 2\,d_x\psi_n, \tag{2.48}$$

where

$$\psi_0 = 1, \quad \psi_1 = d, \quad \psi_2 = -d_{xx} + 3d^2 \tag{2.49}$$

are obtained from the generating function $\psi$, where

$$2\psi\psi_{xx} - \psi_x^2 - 4\,d\psi^2 + 2\lambda\psi^2 - 2\lambda = 0, \tag{2.50}$$

and

$$\psi = \sum_{n=0}^{\infty}\frac{\psi_n}{\lambda^n}. \tag{2.51}$$

From (2.48) and (2.47),

$$\Theta = \psi_{-1}, \tag{2.52}$$

and the sine–Gordon equation is, with the scaling employed,

$$d_t = \frac{\partial}{\partial x}\psi_{-1}. \tag{2.53}$$

The sequence of higher-order KdV equations are

$$d_t = \frac{\partial}{\partial x}\psi_n, \tag{2.54}$$

for $n = 0,1,2,\ldots$ .

It seems appropriate that

$$d_t = \frac{\partial}{\partial x}\psi_{-n}, \tag{2.55}$$

for $n = 1,2,3,4,\ldots$ be termed the higher-order sine–Gordon equations. The results of Ref. 19 demonstrate that the flows of (2.54) and (2.55) "commute" in the sense of Hamiltonian systems. This result is essentially equivalent to that found in Ref. 20.

Next, we note that Eqs. (2.27) and (2.28) are, in effect, the "classical" Bäcklund transformation for the sine–Gordon equation. Let

$$H^2 = \varphi_{xt}^2/\varphi_x\varphi_t, \tag{2.56}$$

then

$$\Omega_1 = WH_t - HW_t - \tfrac{1}{2}W^2H^2 = \alpha,$$

$$\Omega_2 = ZH_x - HZ_x - \tfrac{1}{2}Z^2H^2 = \beta. \tag{2.57}$$

With

$$\alpha = -\tfrac{1}{2}e^{-i\omega_0},$$
$$\beta = -\tfrac{1}{2}e^{i\omega_0},$$
$$H^2 = e^{iu},$$
$$W^2 = \varphi_t/\varphi_x = e^{i\omega}, \tag{2.58}$$

the Eqs. (2.57) become

$$\left(\frac{u-\omega-\omega_0}{2}\right)_t = e^{-i\omega_0/2}\sin\left(\frac{u+\omega+\omega_0}{2}\right), \tag{2.59}$$

and

$$\left(\frac{u+\omega+\omega_0}{2}\right)_x = e^{i\omega_0/2}\sin\left(\frac{u-\omega-\omega_0}{2}\right), \tag{2.60}$$

where

$$u_{xt} = \sin u, \quad (\omega+\omega_0)_{xt} = \sin(\omega+\omega_0). \tag{2.61}$$

Now, Eqs.(2.57) may be reduced by the substitution

$$\Theta = -\frac{H}{W} = -\frac{\varphi_{xt}}{\varphi_t}, \quad \Phi = -\frac{H}{Z} = -\frac{\varphi_{xt}}{\varphi_x}, \tag{2.62}$$

to the form

$$\Theta_t + \frac{1}{2}\Theta\Phi + \frac{\lambda}{2}\frac{\Theta}{\Phi} = 0,$$
$$\Phi_x + \frac{1}{2}\Theta\Phi + \frac{\lambda^{-1}}{2}\frac{\Phi}{\Theta} = 0, \tag{2.63}$$

where

$$V = e^{iu} = \Theta\Phi, \quad \alpha = \lambda/2, \quad \beta = \lambda^{-1}/2. \tag{2.64}$$

We term Eqs. (2.63), the "modified" sine–Gordon equations. (See Appendix B.)

## III. THE DOUBLE SINE–GORDON EQUATION

An extensive discussion of the physics of the double sine–Gordon equation,

$$u_{xt} = 4a\sin(u/2) + 4\sin u, \tag{3.1}$$

is contained in Chap. 3 of Ref. 14.

To apply the Painlevé analysis we set

$$V = e^{iu/2}, \tag{3.2}$$

and find

$$VV_{xt} - V_xV_t = a(V^3 - V) + V^4 - 1. \tag{3.3}$$

The expansion about the singular manifold takes the form

$$V = \varphi^{-1}\sum_{j=0}^{\infty} V_j\varphi^j, \tag{3.4}$$

with resonances at

$$j = -1,2. \tag{3.5}$$

From the recursion relations

$$V_0^2 = \varphi_x\varphi_t, \tag{3.6}$$

$$V_1 = -\frac{1}{2}\frac{\varphi_{xt}}{\varphi_x\varphi_t}V_0 - \frac{a}{2}. \tag{3.7}$$

The compatibility condition at the resonance $j = 2$ is not satisfied identically. Instead, there is found the following "constraint" on $\varphi$:

$$a\left[\frac{\partial}{\partial t}\left(\frac{\varphi_x}{\varphi_t}\right)^{1/2} + \frac{\partial}{\partial x}\left(\frac{\varphi_t}{\varphi_x}\right)^{1/2}\right] = 0. \tag{3.8}$$

Thus, unless $a = 0$, Eq. (3.3) does not possess the Painlevé property. However, if

$$V = f(x+ct), \tag{3.9}$$

condition (3.8) is satisfied and the resulting ode for $f(\epsilon)$,

$$ff_{\epsilon\epsilon} - f_\epsilon^2 = a(f^3 - f) + f^4 - 1, \tag{3.10}$$

is identically Painlevé; and can be solved by quadrature. So far as we have been able to determine, this is the only known exact solution of Eq. (3.1). Concerning this problem, we note the following observations.

(1) If $\varphi$ is a solution of (3.8) then $\psi = f(\varphi)$ is a solution of (3.8) for arbitrary (differentiable) $f$.

(2) Condition (3.8) is

$$\frac{\partial}{\partial t}\left(\frac{\varphi_x}{\sqrt{\varphi_x\varphi_t}}\right) + \frac{\partial}{\partial x}\left(\frac{\varphi_t}{\sqrt{\varphi_x\varphi_t}}\right) = 0, \tag{3.11}$$

which is the "Euler equation"[20] for the functional

$$I_1(\varphi) = \iint \sqrt{\varphi_x\varphi_t}\,dx\,dt. \tag{3.12}$$

However, the identity

$$\frac{\partial}{\partial t}\left(\frac{\varphi_t}{\sqrt{\varphi_x^2+\varphi_t^2}}\right) + \frac{\partial}{\partial x}\left(\frac{\varphi_x}{\sqrt{\varphi_x^2+\varphi_t^2}}\right) = \frac{\partial}{\partial t}\left(\frac{\varphi_x}{\sqrt{\varphi_x\varphi_t}}\right) + \frac{\partial}{\partial x}\left(\frac{\varphi_t}{\sqrt{\varphi_x\varphi_t}}\right) = 0 \tag{3.13}$$

demonstrates that conditions (3.11) or (3.13) are simultaneously the Euler equations of

$$I_2(\varphi) = \iint_D \sqrt{\varphi_x^2+\varphi_t^2}\,dx\,dt. \tag{3.14}$$

Since the "minimal surfaces"[1] are the "minima" of the functional

$$I_3 = \int_D\int \sqrt{1+\varphi_x^2+\varphi_y^2}\,dx\,dt. \tag{3.15}$$

we term the solutions of (3.13) "affine minimal surfaces," i.e., affine in the sense that (3.13) is invariant under the scalings

$$\varphi\to\lambda\varphi, \quad x\to\alpha x, \quad y\to\alpha y. \tag{3.16}$$

(3) The similarity solution of (3.3),

$$v = f(\epsilon), \tag{3.17}$$
$$\epsilon = xt,$$
$$\epsilon ff_{\epsilon\epsilon} - \epsilon f_\epsilon^2 + ff_\epsilon = a(f^3 - f) + f^4 - 1 \tag{3.18}$$

is not Painlevé ($a\neq 0$) since $\varphi = \varphi(xt)$ does not satisfy (3.11).

(4) Letting

$$b = \varphi_t/\varphi_x, \tag{3.19}$$

condition (3.11) becomes

$$b_t = bb_x, \tag{3.20}$$

which is the inviscid Burgers equation. The well-known theory of this equation[21] demonstrates that general, analytic initial data becomes singular "multiple-valued" in a finite time (loss of regularity). Consequently, smooth, "global" solutions of Eq. (3.11) do not exist.

The simple (Painlevé), traveling wave solution (3.9) corresponds to the trivial, $b = \text{const}$, solution of (3.20).

(5) Condition (3.8) can be linearized by a Legendre transformation and the complete solution found. That is, we write (3.8) as

$$\varphi_x^2\varphi_{tt} - 2\varphi_x\varphi_t\varphi_{xt} + \varphi_t^2\varphi_{xx} = 0. \tag{3.21}$$

Then, the Legendre transformation[22]

$$\epsilon = \varphi_x, \quad x = W_\epsilon, \quad \eta = \varphi_t, \quad t = W_n, \tag{3.22}$$

$$\varphi(x,y) + W(\epsilon,\eta) = x\epsilon + t\eta, \tag{3.23}$$

obtains from (3.21) the linear equation

$$\epsilon^2 W_{\epsilon\epsilon} + 2\epsilon n W_{\epsilon n} + \eta^2 W_{nn} = 0. \tag{3.24}$$

Letting

$$\frac{d}{ds} = \epsilon\frac{\partial}{\partial\epsilon} + \eta\frac{\partial}{\partial\eta}, \tag{3.25}$$

we find

$$\frac{d^2}{ds^2}W = \frac{d}{ds}W. \tag{3.26}$$

The complete solution of (3.26) is

$$W = W_0 + W_1, \tag{3.27}$$

where

$$\frac{d}{ds}W_0 = 0,$$

and

$$\frac{d}{ds}W_1 = W_1. \tag{3.28}$$

Here, $W_0$ and $W_1$ are "homogeneous" functions of degree zero and one, respectively. Their general representations are[22]

$$W_0 = G(\epsilon/\eta), \quad W_1 = \eta H(\epsilon/\eta), \tag{3.29}$$

where $G(z)$ and $H(z)$ are arbitrary (smooth) functions. Thus,

$$W = G(\epsilon/\eta) + \eta H(\epsilon/\eta) \tag{3.30}$$

represents the general solution of (3.26). We find, from the above, that

$$\varphi(x,y) = -W_0(\epsilon,\eta), \tag{3.31}$$

and

$$\epsilon x + \eta t = W_1(\epsilon,\eta). \tag{3.32}$$

The Legendre transform is inverted by (3.22). We note that the above goes through when

$$\Omega = \varphi_{xx}\varphi_{tt} - \varphi_{xt}^2 \neq 0. \tag{3.33}$$

If $\Omega = 0$, (3.21) implies

$$\varphi = f(x + ct); \tag{3.34}$$

or, the Legendre transform is defined when $\varphi$ is not a traveling wave.

A few simple solutions can be easily found. For instance,

$$W_1 = 0 \tag{3.35}$$

obtains

$$x\varphi_x + t\varphi_t = 0, \tag{3.36}$$

or

$$\varphi = f(x/t), \tag{3.37}$$

and

$$b = \varphi_t/\varphi_x = -x/t, \tag{3.38}$$

where $f(z)$ is arbitrary and $b$ is a solution of (3.20), the inviscid Burgers equation. From

$$W_0 = \epsilon/\eta, \quad W_1 = \eta^2/\epsilon, \tag{3.39}$$

it is found that

$$\varphi(x,t) = -\frac{\epsilon}{\eta} = -\frac{\varphi_x}{\varphi_t} = -b^{-1} = \frac{t \pm \sqrt{t^2+4x}}{2x}. \tag{3.40}$$

Obviously, algebraic solutions of the inviscid Burgers equation can be constructed by the above method. In a sense, Eq. (3.26) is the linearization of Eq. (3.20). At this point, however, it is not clear how these specific functional forms of $\varphi$, i.e., (3.37) or (3.40), relate to (possible) exact solutions of the double sine–Gordon equation (3.3).

However, it is possible to define a Bäcklund transformation for Eq. (3.3) by letting

$$V = \varphi^{-1}V_0 + V_1, \tag{3.41}$$

where

$$V_0^2 = \varphi_x\varphi_t, \tag{3.42}$$

$$V_1 = -\frac{1}{2}\frac{\varphi_{xt}}{\varphi_x\varphi_t}V_0 - \frac{a}{2}.$$

There is obtained an overdetermined system of three equations in one unknown $\varphi$:

(i) $$a\left[\frac{\partial}{\partial t}\left(\frac{\varphi_x}{\varphi_t}\right)^{1/2} + \frac{\partial}{\partial x}\left(\frac{\varphi_t}{\varphi_x}\right)^{1/2}\right] = 0, \tag{3.43}$$

(ii) $$\left(\frac{\varphi_x}{\varphi_t}\right)^{1/2}\frac{\partial}{\partial x}\Omega_1 + \left(\frac{\varphi_t}{\varphi_x}\right)^{1/2}\frac{\partial}{\partial t}\Omega_2 + a\left[\frac{\varphi_x}{\varphi_t}\Omega_1 + \frac{\varphi_t}{\varphi_x}\Omega_2 + \frac{1}{2}\frac{\partial}{\partial x}\left(\frac{\varphi_x}{\varphi_t}\right) \times \frac{\partial}{\partial t}\left(\frac{\varphi_t}{\varphi_x}\right) + a^2 - 4\right] = 0, \tag{3.44}$$

(iii) $$\frac{a}{8}\left[\left(\frac{\varphi_t}{\varphi_x}\right)^{1/2}\frac{\partial}{\partial x}\left(\frac{\varphi_x}{\varphi_t}\Omega_1\right) + \left(\frac{\varphi_x}{\varphi_t}\right)^{1/2}\frac{\partial}{\partial t}\left(\frac{\varphi_t}{\varphi_x}\Omega_2\right)\right] - \frac{1}{4}\Omega_1\Omega_2 + 1 - \frac{a^2}{8}\left(4 - \frac{a^2}{2}\right) = 0, \tag{3.45}$$

where $\Omega_1$ and $\Omega_2$ are defined in Sec. II.

When $a = 0$ these equations reduce to (2.20) and (2.21). When $a \neq 0$, we conjecture that integrable solutions of the double sine–Gordon equation correspond to the solutions of the system (3.43)–(3.45).

## IV. THE ($N + 1$) SINE–GORDON EQUATION

Herein, we consider the $N$ space-one time ($N + 1$) dimension sine–Gordon equation (SGE). For the (2 + 1) SGE explicit soliton type solutions were obtained by Hirota,[5] while a Bäcklund transform was found by Leibbrandt.[7] Basically, the $n$-soliton solution found by these authors consists of a superposition of $n$ plane, traveling waves.[8] The parameter (directions) of these waves (soliton) are required to satisfy a certain set of compatibility conditions for the solutions to exist.[5,6,8,9] For the $(1+1)$ SGE these conditions are trivial. For the two-soliton solution of the $(2+1)$ SGE, Gibbon and Zambotti[6] have shown the compatability conditions to be trivial; while, for the three-soliton solution, the area of the triangle formed by the three plane waves is time invariant. All the known exact solutions of the $(N+1)$ have an infinite energy since they are constructed from plane waves. It is not known if there exist exact solutions with finite energy.

In what follows we apply the Painlevé analysis to the $(N+1)$ SGE and find that (for $N>1$) this equation is not identically Painlevé. In addition, it can be shown that the directions of the $n$-plane waves must lie in the same plane if the compatibility conditions are to be satisfied for solutions of this type. Hence, these solutions can be obtained by a Lorenz transformation of the solutions of the $(1+1)$ SGE.

Without loss of generality and for notational convenience, we consider the $(N+1)$ elliptic SGE

$$-\Box u = \sin u, \tag{4.1}$$

where

$$\Box = \partial_{xj}^2 = \nabla^t\nabla, \tag{4.2}$$

and

$$\nabla_j = \frac{\partial}{\partial x_j}.$$

By the substitution

$$V = e^{iu}, \tag{4.3}$$

we find

$$-V\Box V + \nabla V\cdot\nabla V = \tfrac{1}{2}(V^3 - V). \tag{4.4}$$

The Painlevé representation

$$V = \varphi^{-2}\sum_{j=0}^{\infty} V_j\varphi^j, \tag{4.5}$$

with resonances at

$$j = -1, 2, \tag{4.6}$$

will be valid if $\varphi = \varphi(x_1,\ldots,x_{n+1})$ satisfies a compatibility condition. Using the expressions

$$V_0 = -4\nabla\varphi\cdot\nabla\varphi, \quad V_1 = 4\Box\varphi, \tag{4.7}$$

the compatibility condition is found to be

$$\nabla\varphi\cdot D\nabla\varphi = 0, \tag{4.8}$$

where

$$D_{ii} = \sum_{\substack{l=1\\ l\neq i}}^{N+1}\sum_{\substack{m=1\\ m\neq i}}^{N+1}(\varphi_{lm}^2 - \varphi_{ll}\varphi_{mm}), \tag{4.9}$$

and

$$D_{ij} = \sum_{k=1}^{N+1}(\varphi_{ij}\varphi_{kk} - \varphi_{ik}\varphi_{jk}). \tag{4.10}$$

We note the following observations.

(1) The matrix $D$ is symmetric ($D_{ij} = D_{ji}$) and Eq. (4.8) is trivial when $N=1$ [$(1+1)$ SGE].

(2) Equation (4.8) is invariant under the change of variables, $x_j \to ix_j$ (hyperbolic SGE).

(3) Equation (4.8) is translation invariant, i.e., $x_j \to x_j + c_j$.

(4) Equation (4.8) is invariant under orthogonal changes of independent variables,

$$\nabla = B\nabla', \tag{4.11}$$

where

$$B^t = B^{-1}.$$

Observation (4) follows from the orthogonal invariance of (4.4) and (4.7).

Therefore, consider the hypersurface $M$ defined by

$$M = \{\hat{x}:\varphi(\hat{x}) = \varphi_0\}, \tag{4.12}$$

where $\hat{x} = (x_1, x_2, \ldots, x_{n+1})$ and $\nabla\varphi|M$.

By translation and rotation we may locate the origin of the coordinate system at a point $\hat{x}_0 \in M$ so that

$$\frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_{n+1}}$$

provide an orthogonal basis for the tangent space of $M$ at $\hat{x}_0$. Since $M$ is a hypersurface there is a unique normal to $M$ at $\hat{x}_0$:

$$\hat{N} \approx \begin{pmatrix}\varphi_{x_1}\\ 0\\ \vdots\\ 0\end{pmatrix}. \tag{4.13}$$

By observations (3) and (4) and (4.13), Eq. (4.8) reduces to

$$\varphi_{x_1}^2 \sum_{l=2}^{N+1}\sum_{m=2}^{N+1}(\varphi_{lm}^2 - \varphi_{ll}\varphi_{mm}) = 0, \tag{4.14}$$

at the "arbitrary point" $\hat{x}_0$.

In terms of the hypersurface $M$, Eq. (4.14) states[23] that the elementary symmetric function of the principal curvatures of $M$ vanishes. That is,

$$K_1K_2 + K_1K_3 + \cdots + K_{n-1}K_n = 0, \tag{4.15}$$

where $K_j, j = 1,\ldots,n$ are the principal curvatures of $M$. In effect, Eq. (4.14) is the sum of the principal minors of order 2 of the second fundamental form of $M$.[23]

Now, let $N=2$ [the $(2+1)$ SGE] and find

$$K_1K_2 = 0 \tag{4.16}$$

or $K = K_1K_2$ (the Gaussian curvature) vanishes, defining a "developable surface."[23] Condition (4.8) becomes, in the variables $(t,x,y)$,

$$\begin{aligned}&\varphi_t^2(\varphi_{xx}\varphi_{yy} - \varphi_{xy}^2) + \varphi_x^2(\varphi_{tt}\varphi_{yy} - \varphi_{yt}^2)\\ &\quad + \varphi_y^2(\varphi_{tt}\varphi_{xx} - \varphi_{xt}^2)\\ &\quad + 2\varphi_x\varphi_t(\varphi_{ty}\varphi_{yx} - \varphi_{xt}\varphi_{yy}) + 2\varphi_y\varphi_t(\varphi_{tx}\varphi_{xy} - \varphi_{yt}\varphi_{xx})\\ &\quad + 2\varphi_x\varphi_y(\varphi_{xt}\varphi_{yt} - \varphi_{xy}\varphi_{tt}) = 0.\end{aligned} \tag{4.17}$$

As noted in observation (1), Eq. (4.17) is trivial when $\varphi$ is a function of two variables, i.e., $\varphi = \varphi(t,x)$.

Now, let $\varphi$ be a product of plane, traveling waves:

$$\varphi = \prod_{j=1}^{m} f_j(a_jt + b_jx + c_jy - d_j), \tag{4.18}$$

where the $f_j(z)$ are *arbitrary*.

If $m=2$ (two waves), a rotation of the coordinates can be devised so that $\varphi$ depends (effectively) on two variables, and condition (4.17) will be trivial.[6]

For any $m$ a similar argument demonstrates (4.17) will be satisfied identically if *all* of the wave directions, $\hat{a}_j = (a_j, b_j, c_j)$, lie in the same plane. Furthermore, the necessity of this condition can be proven by direct substitution of (4.18) into (4.17) and using the requirement that the $f_j(z)$ be arbitrary.

For three waves the co-planar condition may be written

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = 0. \tag{4.19}$$

This is the condition found in Ref. 6 for the existence of the three soliton solution. It indicates that the area of the triangle formed by the plane waves is time invariant.

From the above it appears that the class of known, exact solutions for the (2 + 1) SGE is trivial in that they can be reduced to solutions of the (1 + 1) SGE. If nontrivial solutions of (4.17) (developable surfaces) correspond to exact solutions of (4.4) this class may contain solutions with nonreducible behavior.

As in Sec. III the compatibility condition (4.17) may be "linearized" and the complete solution found by a Legendre transformation. That is,

$$\begin{aligned} \epsilon_1 &= \varphi_t, \quad t = W_{\epsilon_1}, \\ \epsilon_2 &= \varphi_x, \quad x = W_{\epsilon_2}, \\ \epsilon_3 &= \varphi_y, \quad y = W_{\epsilon_3}, \end{aligned} \tag{4.20}$$

$$\varphi(t,x,y) + W(\epsilon_1,\epsilon_2,\epsilon_3) = t\epsilon_1 + x\epsilon_2 + y\epsilon_3 \tag{4.21}$$

obtains from (4.17) the linear equation (with summation convention)

$$\epsilon_i \epsilon_j \frac{\partial^2}{\partial\epsilon_i\,\partial\epsilon_j} W = 0. \tag{4.22}$$

Letting

$$\frac{d}{ds} = \epsilon_i \frac{\partial}{\partial \epsilon_i}, \tag{4.23}$$

we find

$$\frac{d^2}{ds^2} W = \frac{d}{ds} W. \tag{4.24}$$

The complete solution of (4.24) is

$$W = W_0 + W_1, \tag{4.25}$$

where

$$\frac{d}{ds} W_0 = 0, \quad \frac{d}{ds} W_1 = W_1. \tag{4.26}$$

Here $W_0$ and $W_1$ are "homogeneous" functions of degree zero and one, respectively. (See Sec. III.) Again, we find

$$\varphi(t,x,y) = -W_0(\epsilon_1,\epsilon_2,\epsilon_3), \tag{4.27}$$

and

$$t\epsilon_1 + x\epsilon_2 + y\epsilon_3 = W_1(\epsilon_1,\epsilon_2,\epsilon_3). \tag{4.28}$$

We note that the Legendre transformation is defined when $\varphi$ depends, effectively, on three independent variables.

## ACKNOWLEDGMENTS

This work was supported by the Department of Energy Contract No. DOE DE-AC03-81ER10923 and AFOSR Grant No. AFOSR 83-0095.

## APPENDIX A: THE COUPLED KdV, OR HIROTA–SATSUMA EQUATIONS

The Hirota–Satsuma, or coupled KdV, equations[24]

$$u_t = \frac{\partial}{\partial x}(-3\omega^2 + a u_{xx} + 3au^2), \tag{A1}$$

$$\omega_t + \omega_{xxx} + 3\omega_x u = 0 \tag{A2}$$

have a Lax pair[25] and an infinite sequence of conserved quantities.[11] Although this strongly indicates their "complete integrability," Oevel[11] states that (A1) and (A2) are not "completely integrable" in the usual sense since the "symmetries" of these equations are not "dense" in the "space of vector fields." We note that the fourth-order scattering theory associated with the Lax pair for the KdV equations has not been developed. (See Ref. 4.)

In this Appendix we find the CKdV equations identically possess the Painlevé property if and only if $a = \frac{1}{2}$; consistent with the results of Refs. 24 and 25. Additionally, the "singular manifold" analysis is applied to obtain the Bäcklund transform/Lax-pair structure.

(i) There are found to be two types of singularities.

Branch 1:

$$u = \varphi^{-2} \sum_{j=0}^{\infty} u_j \varphi^j, \tag{A3}$$

$$\omega = \varphi^{-1} \sum_{j=0}^{\infty} \omega_j \varphi^j, \tag{A4}$$

with resonances

$$j = -1,0,1,4,5,6. \tag{A5}$$

(ii) Branch 2:

$$u = \varphi^{-2} \sum_{j=0}^{\infty} u_j \varphi^j, \tag{A6}$$

$$\omega = \varphi^{-2} \sum_{j=0}^{\infty} \omega_j \varphi^j, \tag{A7}$$

with resonances

$$j = -2, -1, -3,4,6,8. \tag{A8}$$

Calculation obtains that both branches have the Painlevé property if and only if

$$a = \tfrac{1}{2}. \tag{A9}$$

Branch 1 depends on six, and branch 2 on five, arbitrary functions. In what follows, we consider only branch 1, and define the Bäcklund transform

$$u = u_0/\varphi^2 + u_1/\varphi + u_2, \tag{A10}$$

$$\omega = \omega_0/\varphi + \omega_1, \tag{A11}$$

where

$$u_0 = -2\varphi_x^2, \quad u_1 = 2\varphi_{xx}. \tag{A12}$$

Hence,

$$u = 2\frac{\partial^2}{\partial x^2} \ln \varphi + u_2. \tag{A13}$$

The resulting overdetermined system consists of six equations for four unknowns $(\varphi, u_2, \omega_0, \omega_1)$. The somewhat tedious reduction of this system is facilitated by the substitution (see Sec. II)

$$u_2 = V - \tfrac{1}{2}\varphi_{xx}^2/\varphi_x^2, \tag{A14}$$

and reformulation of the system of equations in terms of Schwarzian derivatives. Eventually, we arrive at the following consistent reduction:

$$\varphi_t + \varphi_{xxx} + 3\varphi_x u_2 = 2\varphi_x \Theta, \tag{A15}$$

$$\varphi_t/\varphi_x - \tfrac{1}{2}\{\varphi;x\} - \tfrac{3}{4}H^2 = \Theta, \tag{A16}$$

$$\omega_0 = \varphi_x H, \tag{A17}$$

$$\Theta_x^2 = (\lambda^2 + \Theta^2)H^2, \tag{A18}$$

$$\omega_1 = -\tfrac{1}{2}\omega_{0x}/\varphi_x - \tfrac{1}{3}(\lambda^2 + \Theta^2)^{1/2}, \tag{A19}$$

and

$$H_t + \frac{\partial}{\partial x}\left(H_{xx} + \frac{H^3}{4} + \Theta H + \frac{3}{2}\{\varphi;x\}H\right) = 0. \tag{A20}$$

From (A16) and (A17)

$$\tfrac{3}{4}\omega_0^2 = \varphi_x\varphi_t - (\varphi_x^2/2)\{\varphi;x\} - \varphi_x^2\Theta. \tag{A21}$$

The relevant equations in the above system are (A16), (A18), and (A20). These equations define, implicitly, an equation for $\varphi$, invariant under the Moebius group. From this, we can, as in Ref. 2, find the Lax pair for (A1) and (A2) from the Wronskian relations. However, here it is more convenient to proceed differently. That is, we let

$$W = \varphi_{xx}/\varphi_x, \tag{A22}$$

and find the "modified" Hirota–Satsuma equations

$$H_t + \frac{\partial}{\partial x}\left[H_{xx} + \tfrac{1}{4}H^3 + \Theta H + \tfrac{3}{2}(W_x - \tfrac{1}{2}W^2)H\right] = 0, \tag{A23}$$

$$W_t = \frac{1}{2}\frac{\partial}{\partial x}\left[W_{xx} - \frac{W^3}{2} + 3\left(H_x + \frac{WH}{2}\right) \times H + 2(\Theta_x + W\Theta)\right], \tag{A24}$$

where

$$\Theta_x = (\lambda^2 + \Theta^2)^{1/2}H.$$

We intend to find the Lax pair by "linearizing" the Miura type transformation relating (A23) and (A24) to (A1) and (A2). From (A15) to (A19) and (A22), the "Miura transformations" are

$$-2u_2 = W_x + \tfrac{1}{2}W^2 + \tfrac{1}{2}H^2 - \tfrac{2}{3}\Theta, \tag{A25}$$

$$-2\omega_1 = H_x + WH + \tfrac{2}{3}(\lambda^2 + \Theta^2)^{1/2}, \tag{A26}$$

were $(u_2,\omega_1)$ satisfy (A1) and (A2) and $(H,W)$ satisfy (A23) and (A24).

Now letting

$$W + H = 2\psi_x/\psi, \tag{A27}$$

$$W - H = 2\beta_x/\beta, \tag{A28}$$

and

$$\Theta = \lambda \sinh \alpha, \tag{A29}$$

we find from the above that

$$\alpha = \ln(\psi/\beta), \tag{A30}$$

$$H = \alpha_x, \tag{A31}$$

and

$$\psi_{xx} + (u_2 + \omega_1)\psi = -(\lambda/3)\beta,$$
$$\beta_{xx} + (u_2 - \omega_1)\beta = (\lambda/3)\psi. \tag{A32}$$

Equations (A32) are the spatial (scattering) part of the Lax pair[25] for (A1) and (A2). The time-dependent operator is found from (A23), (A24), (A27), and (A28). That is,

$$\psi_t + \tfrac{1}{2}(u_{2_x} - 2\omega_{1x})\psi + (u_2 - 2\omega_1)\psi_x = -\tfrac{2}{3}\lambda\beta_x, \tag{A33}$$

$$\beta_t + \tfrac{1}{2}(u_{2_x} + 2\omega_{1x})\beta + (u_2 + 2\omega_1)\beta_x = \tfrac{2}{3}\lambda\psi_x. \tag{A34}$$

We now consider the singularities of the modified Hirota–Satsuma system, i.e., Eqs. (A16), (A18), and (A20). It is convenient to use the substitution (A29) with

$$H = \alpha_x = h_x/h, \tag{A35}$$

to obtain the system

$$\frac{\varphi_t}{\varphi_x} = \frac{1}{2}\{\varphi;x\} + \frac{3}{4}\left(\frac{h_x}{h}\right)^2 + \frac{\lambda}{2}\left(h - \frac{1}{h}\right), \tag{A36}$$

$$\frac{h_t}{h} + \left(\frac{h_x}{h}\right)_{xx} + \frac{1}{4}\left(\frac{h_x}{h}\right)^3 + \frac{3}{2}\frac{h_x}{h}\{\varphi;x\} + \frac{\lambda}{2}\left(h - \frac{1}{h}\right)\frac{h_x}{h} = 0. \tag{A37}$$

A leading-order analysis with

$$\varphi_{xx}/\varphi_x \approx a/\epsilon, \quad h_x/h \approx b/\epsilon, \tag{A38}$$

discovers the following possibilities:

(i) $b = 0, \quad a = -2;$ (A39)

(ii) $b^2 = 1, \quad a = 1$ or $-3;$ (A40)

and, if $\lambda \neq 0$,

(iii) $b = -2, \quad a = 0, -2;$ (A41)

(iv) $b = 2, \quad a = 0, -2.$ (A42)

We proceed to investigate in detail singularities of the form

$$\varphi = \sum_{j=0}^{\infty}\varphi_j\epsilon^j, \quad h = \sum_{j=0}^{\infty}h_j e^{j-1}, \tag{A43}$$

where we employ the "reduced" *Ansätze*,[1]

$$\epsilon = x - \psi(t), \quad \varphi_j = \varphi_j(t), \quad h_j = h_j(t), \tag{A44}$$

and

$$\varphi_1 \equiv 0 \tag{A45}$$

is required by the condition $\varphi_x = 0$. A calculation finds the resonances to occur at

$$j = -2, -1, 0, 0, 2, 4, \tag{A46}$$

which corresponds to the "arbitrary" functions $\varphi_0, \epsilon, \varphi_2, h_0, \varphi_4, h_4$, respectively. Nontrivial compatibility conditions occur when $j = 2,4$.

A direct calculation determines that the compatibility conditions at $j = 2$ and $j = 4$ are satisfied identically. Thus Eqs. (A34) and (A35) have the Painlevé property about singularities of form (A42). Although we have not checked all the singularities of Eqs. (A36) and (A37), it is probably true that they identically possess the Painlevé property. However, it is interesting to observe that the modified equations for the Hirota–Satsuma system [(A23) and (A24)] do not seem to have any discrete symmetries (other than the trivial $H \to -H$). A connection between the discrete symmetries (or modified equations) and the Painlevé property of sequences of higher-order equations is examined in Ref. 3. (See also Appendix B.) In Ref. 26 it is shown how Miura-type transformations from modified (to original) equations allows the definition of, among other things, the recursion operators producing the sequences of higher order equations. It is possible that, when the modified equations (defined in terms of the "singular manifold" Bäcklund transformation) are "missing" discrete symmetries, the associated Hamiltonian structures are "degenerate" (e.g., Ref. 27).

## APPENDIX B: THE MODIFIED SINE–GORDON EQUATIONS

In Sec. II we have defined the modified sine–Gordon equations to be

$$\Theta_t + \frac{1}{2}\Theta\Phi + \frac{\lambda}{2}\frac{\Theta}{\Phi} = 0, \quad \Phi_x + \frac{1}{2}\Theta\Phi + \frac{\lambda^{-1}}{2}\frac{\Phi}{\Theta} = 0, \tag{B1}$$

where

$$V = e^{iu} = \Theta\Phi, \quad \alpha = \lambda/2, \quad \beta = \lambda^{-1}/2. \tag{B2}$$

These equations have singularities of the form

$$\Theta \sim \Theta_0\epsilon^\alpha, \quad \Phi \sim \Phi_0\epsilon^\beta, \tag{B3}$$

where

$$\text{(i)}\ \alpha = \beta = -1, \quad \Theta_0 = 2\epsilon_x, \quad \Phi_0 = 2\epsilon_t; \tag{B4}$$

$$\text{(ii)}\ \alpha = -1, \quad \beta = 1, \quad \Theta_0 = -2\epsilon_x, \quad \Phi_0 = \lambda/2\epsilon_t; \tag{B5}$$

$$\text{(iii)}\ \alpha = 1, \quad \beta = -1, \quad \Theta_0 = 1/2\lambda\epsilon_x, \quad \Phi_0 = -2\epsilon_t. \tag{B6}$$

The resonances, in all cases, occur at

$$j = -1, 1. \tag{B7}$$

Equations (B1) have the following discrete symmetries:

$$\text{(i)}\ \Theta = (1/\lambda)\tilde{\Theta}^{-1}, \quad \Phi = -\tilde{\Phi}; \tag{B8}$$

$$\text{(ii)}\ \Theta = -\tilde{\Theta}, \quad \Phi = \lambda\tilde{\Phi}^{-1}. \tag{B9}$$

Thus, by composition of (B8) and (B8), the following four solutions of (B1) are related:

$$[\Theta,\Phi], [(1/\lambda)\Theta^{-1}, -\Phi], \quad [-\Theta, \lambda\Phi^{-1}], [-(1/\lambda)\Theta^{-1}, -\lambda\Phi^{-1}]. \tag{B10}$$

Direct calculation obtains the Painlevé property for singularities of the form (B4), while the above symmetry implies that (B5) and (B6) are Painlevé as well. Thus, (B1) has the Painlevé property.

Now, let

$$\Theta = (\lambda/2)V(b_t/b), \quad \Phi = (\lambda^{-1}/2)V(h_x/h), \tag{B11}$$

and, using (B2), find

$$b_{tt} + (V_t/V)b_t + \lambda^{-1}b = 0, \quad h_{xx} + (V_x/V)h_x + \lambda h = 0, \tag{B12}$$

where

$$V(h_x/h)(b_t/b) = 4, \tag{B13}$$

and $V$ satisfies Eq. (2.3).

The pair of linear equations (B12) (essentially Schrödinger equations) would seem to be related to the "characteristic" initial value problem for Eq. (2.3). The initial conditions for a problem of this might be

$$V(x,0), \quad 0 < x < \infty, \quad V(0,t), \quad 0 < t < \infty, \tag{B14}$$

where it is required to find $V(x,t)$.

We recall that the modified sine–Gordon equations (B1), are obtained from the equations

$$\Omega_1 = \{\varphi;t\} + 2(Z_{tt}/Z) = \lambda/2, \quad \Omega_2 = \{\varphi;x\} + 2(W_{xx}/W) = 1/2\lambda, \tag{B15}$$

where

$$Z^2 = W^{-2} = \varphi_x/\varphi_t, \tag{B16}$$

by the substitution

$$\Theta = -\varphi_{xt}/\varphi_t, \quad \Phi = -\varphi_{xt}/\varphi_x. \tag{B17}$$

Equations (B15) allow three types of singularities:

$$\text{(i)}\ \varphi = \epsilon^{-1}\sum_{j=0}^{\infty}\varphi_j\,\epsilon^j, \tag{B18}$$

$$\text{(ii)}\ \varphi = \varphi_0(t) + \varphi_3\epsilon^3 + \varphi_5\epsilon^5 + \ldots, \tag{B19}$$

$$\text{(iii)}\ \varphi = \varphi_0(x) + \varphi_3\epsilon^3 + \varphi_5\epsilon^5 + \cdots. \tag{B20}$$

These are all of the Painlevé type.

Now, with

$$\tilde{\Theta} = -\psi_{xt}/\psi_t, \quad \tilde{\Phi} = -\psi_{xt}/\psi_x, \tag{B21}$$

the symmetries (B8)–(B10) become

$$\text{(i)}\ \frac{\psi_{xt}}{\psi_t}\frac{\varphi_{xt}}{\varphi_t} = \frac{1}{\lambda}, \quad \psi_x\varphi_x = 1; \tag{B22}$$

$$\text{(ii)}\ \frac{\psi_{xt}}{\psi_x}\frac{\varphi_{xt}}{\varphi_x} = \lambda, \quad \psi_t\varphi_t = 1; \tag{B23}$$

$$\text{(iii)}\ \frac{\psi_{xt}}{\psi_t}\frac{\varphi_{xt}}{\varphi_t} = -\frac{1}{\lambda}, \quad \frac{\psi_{xt}}{\psi_x}\frac{\varphi_{xt}}{\psi_x} = -\lambda. \tag{B24}$$

These, along with the invariance under the Moebius group,

$$\psi = (a\varphi + b)/(c\varphi + d), \tag{B25}$$

constitute Bäcklund transformations for Eqs. (B15). For instance, consider (B23), which is equivalent to

$$\psi_t = \varphi_t^{-1}, \quad \psi_x = -(1/\lambda)(\varphi_{xt}^2/\varphi_t^2\varphi_x). \tag{B26}$$

The consistency condition

$$\psi_{tx} = \psi_{xt},$$

requires $\varphi$ to satisfy Eqs. (B15). We note that (B23) is symmetric in $(\varphi,\psi)$. Thus, (B23) implies that both $(\varphi,\psi)$ satisfy (B15).

Following the method of Ref. 3 we iteratively construct the "rational" solutions of the sine–Gordon equation (using the symmetries of the "modified equations"). In this case, by rational, we mean rational in $(x,t,e^x,e^t)$. To proceed let the "meromorphic" functions $(\psi,\varphi)$ be expressed as the ratio of entire functions

$$\psi = P/Q, \quad \varphi = R/S, \tag{B27}$$

and substitute into, say (B23). The resulting expressions may be reduced to the equations

$$ST_t - RS_t = Q^2,$$
$$QP_t - PQ_t = S^2, \tag{B28}$$
$$\sigma^2(QP_x - PQ_x)(SR_x - RS_x) = 4\{SQ_x - QS_x\}^2, \tag{B29}$$

where $\lambda = -\sigma^2$. Equations (B28), (B29), and (B27) define solutions of (B23), consistent with the assumption that $(P,Q,R,S)$ are entire, if the terms $(QP_x - PQ_x)$ and $(SR_x - RS_x)$ "divide" the term $\{SQ_x - QS_x\}^2$. For instance, let it be required to solve (B29) for $(P,Q)$. Then we must have

$$\{SQ_x - QS_x\}^2 = a(SR_x - RS_x), \tag{B30}$$

where $a$ is entire. Using (B15), (B23), and (B26)–(B28) it is found that

$$4a = \sigma^2(\psi_x/\psi_t)S^2. \tag{B31}$$

Since $S$ is entire, singularities of $a$ can only occur when [see (B20)]

$$\psi = \psi_0(x) + \psi_3\epsilon^3 + \cdots. \tag{B32}$$

By (B27), locally, with $\epsilon = t + f(x)$,

$$P = \psi_0(x) + P_3\epsilon^3 + \ldots, \quad Q = 1 + Q_3\epsilon^3 + \ldots, \tag{B33}$$

and, by (B28)

$$S^2 = QP_t - PQ_t. \tag{B34}$$

Thus,

$$4a = O(\epsilon^0) \tag{B35}$$

is entire.

We now compose (B23) iteratively with the transformation

$$\varphi \to -1/\varphi, \tag{B36}$$

which is to, effectively, identify

$$\psi = P_{n+1}/P_{n-1}, \quad \varphi = -P_{n-2}/P_n, \tag{B37}$$

thereby obtaining the recursion relations

$$P_{n-1}P_{n+1,t} - P_{n+1}P_{n-1,t} = P_n^2, \tag{B38}$$

and

$$\sigma^2(P_{n-1}P_{n+1,x} - P_{n+1}P_{n-1,x})(P_{n-2}P_{n,x} - P_nP_{n-2,x}) = 4\{P_{n-1}P_{n,x} - P_nP_{n-1,x}\}^2. \tag{B39}$$

From Eqs. (B15), with $\lambda = -\sigma^2$, the simplest nontrivial solution seems to be

$$\varphi_0 = e^{\sigma t + x/\sigma}. \tag{B40}$$

By (B25), the solution

$$\varphi_1 = \tanh\left(\frac{\sigma}{2}t + \frac{x}{2\sigma}\right) \tag{B41}$$

is found. From (B38) and (B39), we find

$$\varphi_2 = \frac{1}{\sigma^2}\left[\sinh\left(\sigma t + \frac{x}{\sigma}\right) + \sigma t - \frac{x}{\sigma}\right],$$
$$\varphi_3 = \frac{1}{\sigma^2}\left[\sinh\left(\sigma t + \frac{x}{\sigma}\right) - \left(\sigma t - \frac{x}{\sigma}\right) + \left(\sigma t - \frac{x}{\sigma}\right)^2 \tanh\left(\frac{\sigma}{2}t + \frac{x}{2\sigma}\right)\right], \tag{B42}$$

which define "rational" solutions of the sine–Gordon (modified sine–Gordon) equations.

We note that Eq. (B38) is identical to that found in Ref. 3 for KdV equation [Eq. (B39) here determines certain constants of integration]. However, unlike for the KdV equation, there are no solutions rational in $(x,t)$ only, since in Eqs. (B15) the limit when $\lambda \to 0$ is not defined. Of course, the Bäcklund transformations (B22)–(B25) may be iteratively applied to create different sequences of "rational" solutions.


[1]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[2]J. Weiss, J. Math. Phys. **24**, 1405 (1983).
[3]J. Weiss, J. Math. Phys. **25**, 13 (1984).
[4]J. B. McLeod and P. J. Olver, SIAM J. Math. Anal. **14**, 488 (1983).
[5]R. Hirota, J. Phys. Soc. Jpn. **35**, 1566 (1973).
[6]J. D. Gibbon and G. Zambotti, Nuovo Cimento B **28**, 1 (1975).
[7]G. Liebbrandt, Phys. Rev. Lett. **41**, 435 (1978).
[8]P. L. Christiasen, *On Bäcklund Transformations and Solutions to the 2+1 and 3+1-Dimensional Sine–Gordon Equation,* Lecture Notes in Mathemetics (Springer, New York, 1980), Vol. 810, p. 43.
[9]G. Liebbrandt, R. Morf, and S. Wang, J. Math. Phys. **21**, 1613 (1980).
[10]J. Zagrodzinski, J. Math. Phys. **24**, 46 (1983).
[11]W. Oevel, "On the Integrability of the Hirota-Satsuma System," Phys. Lett. A **94**, 404 (1983).
[12]H. Segur, "Solitons and the Inverse Scattering Transform," Lectures given at International School of Physics "Enrico Fermi," Varenna, Italy, 7–19 July 1980.
[13]M. Tabor and J. Weiss, Phys. Rev. A **24**, 2157 (1981).
[14]*Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, New York, 1980).
[15]A. V. Bäcklund, Lunds Univ. Arsskr. Avd. 2, **19** (1883).
[16]V. E. Zahkharov and A. B. Shabat, Sov. Phys. JETP **34**, 62 (1972).
[17]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[18]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Commun. Pure Appl. Math. **27**, 97 (1974).
[19]P. D. Lax, SIAM Rev. **18**, 351 (1976).
[20]K. M. Case and A. M. Roos, J. Math. Phys. **23**, 392 (1983).
[21]G. B. Whitham, *Linear and Nonlinear Waves* (Interscience, New York, 1974).
[22]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1962), Vol. II.
[23]D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Wiley–Interscience, New York, 1975).
[24]R. Hirota and J. Satsuma, Phys. Lett. A **85**, 407 (1981).
[25]R. Dodd and A. Fordy, Phys. Lett. A **89**, 168 (1982).
[26]A. S. Fokas and R. L. Anderson, J. Math. Phys. **23**, 1066 (1982).
[27]B. Fuchssteiner, Prog. Theor. Phys. **68**, 1082 (1982).

# An exact nonghost solution for a plane-symmetric cosmology containing a classical spinor field

Richard A. Matzner
*Relativity Center, The University of Texas, Austin, Texas 78712*
Michael P. Ryan, Jr.[a)]
*Universidad Autonóma Metropolitana-Iztapalapa, 09340, México, D.F., Mexico*

(Received 17 January 1984; accepted for publication 24 February 1984)

An exact solution (up to quadratures) of the Einstein–Dirac system is presented for cosmological models that depend only on one temporal and one space coordinate. Four solutions to the Dirac equation, all with zero helicity, are given.

PACS numbers: 03.40.Kf, 04.20.Jb, 04.40. + c

## I. INTRODUCTION

Plane symmetric cosmologies have been the subject of a number of studies.[1–8] These models have a particulary simple type of inhomogeneity that leads, in many cases, to field equations in which there is a "key" equation that is linear and easily soluble. The remaining nonlinear equations then can be solved by quadratures; their integrability is assured if the key equation is satisfied. A number of "folded" versions of these cosmologies are known, in which points a certain coordinate distance apart in certain directions are identified to give the manifold a cylindrical or toroidal topology. The cylindrical topology corresponds to the Einstein–Rosen[6] solution, and the toroidal topology to the Gowdy model.[7] Any results in any one of the above models can be directly transferred to the plane-symmetric models and vice versa provided proper attention is paid to the effects of the different boundary conditions dictated by the different topologies. Taking into account all work on this subject regardless of the topology chosen, there exists a large body of results for various types of matter in the plane-symmetric case. For instance, there exist a number of results valid for the Gowdy model[8] which can with minor modification be transferred to the plane-symmetric case.

One possibility that apparently has not been considered is that of a classical spinor field in these cosmologies. Classical spinor fields have a number of strange properties that make it worthwhile and interesting to consider models filled with this type of matter. Most interesting is the strong tendency for simplified metrics to allow at most "ghost"[9] solutions in which the spinor field is nonzero, but the stress-energy tensor vanishes. This feature arises, for instance, when homogeneous spinor fields are introduced in the homogeneous-isotropic Robertson–Walker cosmologies. It seems in large part to be due to the existence of a spinor momentum flux $T_{0i}$ which cannot in many cases be put equal to zero without forcing the rest of $T_{\mu\nu}$ to vanish. As a result, since the Robertson–Walker cosmologies require by symmetry that $G_{0i} = 0$, nonghost homogeneous neutrino fields are excluded in them. (There do exist non-Robertson–Walker cases where $T_{0i}$ is automatically zero as a result of the Dirac equation.) To avoid ghosts then, the major requirement for those cases in which $T_{0i}$ does not automatically vanish by the Dirac equation is that the geometry admit nonzero $G_{0i}$. This is exemplified in the Bianchi type IX cosmological models; in the diagonal and FRW cases the fact that $G_{0i}$ is zero forces the models to allow only ghost solutions,[10–11] while the symmetric case[12] in which $G_{0i} \neq 0$ allows nonghost solutions. Isham and Nelson[11] suggest that allowing inhomogeneity would help considerably in finding nonghost solutions. In this paper we show that even inhomogeneous models impose strong restrictions on the spinor fields that are allowed. In the plane-symmetric case there are nonghost solutions, but for which $G_{0i}$ is nonzero, so it is not clear that the inhomogeneity alone is sufficient to allow nonghost solutions. The existence of ghost solutions certainly adds interest to the study of spinor fields in cosmology. Realistically, one must admit that the existence of ghost solutions is almost certainly another indication of the fact that Dirac theory is a quantum theory with no real classical limit, and the attempt to force it into a classical mold results in strange behavior.[13]

In this paper we will solve the Dirac equation in the metric of a plane-symmetric model subject to the constraints imposed by the fact that some of the $G_{\mu\nu}$ are zero while the corresponding $T_{\mu\nu}$ are not automatically zero. The equations for the metric components then reduce to the *same* key equation as found in the vacuum case, and a set of equations that, in principle, can be solved by quadratures. The integrability conditions for this set of equations are one that is satisfied if the key equation is satisfied, and another that we show is satisfied given the solution of the Dirac equation. In this way we have an exact, nonghost solution to the problem up to quadratures.

The paper is organized as follows: In Sec. II we give the equations of motion for the metric and the spinor field. In Sec. III we solve the equations up to quadratures, and in Sec. IV we give conclusions and discuss some of the properties of the solutions.

## II. EQUATION TO MOTION

We write the metric of a plane-symmetric model in the form (see, for example, Ref. 5)

[a)] Permanent address: Centro de Estudios Nucleares, Universidad Nacional Autónoma de México, Apdo. Postal 70-543, 04510, México, D.F., Mexico.

  

$$ds^2 = e^{2(\gamma - \psi)}(-dT^2 + dZ^2) + e^{2\psi} dX^2 + T^2 e^{-2\psi} dY^2, \tag{2.1}$$

where $\gamma$ and $\psi$ are functions of $T$ and $Z$ only. Matter in this model will be described by a Dirac spinor $\Psi$ which is also a function only of $Z$ and $T$. The spinor $\Psi$ obeys the Dirac equation in curved space,

$$i\gamma^\mu \Psi_{|\mu} - m\Psi = 0, \tag{2.2}$$

where the $\gamma^\mu$ are the flat-space $\gamma$ matrices in the standard representation, that is,

$$\gamma^i = \begin{bmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{bmatrix}, \quad \gamma^0 = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \tag{2.3}$$

where the $\sigma^i$ are the Pauli matrices. Note that $\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = -2\eta^{\mu\nu}$ since our metric has signature $(- + + +)$. The use of the flat-space $\gamma$ matrices implies that Eq. (2.2) is written in an orthonormal frame, which we choose to be defined by the following one-forms:

$$\omega^0 = e^{\gamma - \psi} dT, \quad \omega^X = e^\psi dX,$$
$$\omega^Y = Te^{-\psi} dY, \quad \omega^Z = e^{\gamma - \psi} dZ. \tag{2.4}$$

In the notation of Misner, Thorne, and Wheeler,[14] $\Psi_{|\mu}$ is $\nabla_\mu \Psi - \Gamma_\mu \Psi$, where $\nabla_\mu$ means $e^\alpha_{(\mu)}\ \partial/\partial X^\alpha$ with $\{e_{(\mu)}\}$ the basis vectors dual to the $\{\omega^\sigma\}$, where $\Gamma_\mu$ is a spin connection defined by[12] $\Gamma_\mu = \frac{1}{4}\Gamma_{\rho\nu\mu}\gamma^\rho\gamma^\nu$, and where the $\Gamma_{\rho\nu\mu}$ are the connection coefficients of the orthonormal basis (2.4), again in the notation of Misner, Thorne, and Wheeler. In our case

$$\Gamma_0 = -\tfrac{1}{2}(\gamma - \psi)' e^{-(\gamma - \psi)} \gamma^0 \gamma^Z, \quad \Gamma_Z = -\tfrac{1}{2}(\gamma - \psi)^\bullet \gamma^0 \gamma^Z,$$
$$\Gamma_X = \tfrac{1}{2}(-\dot\psi e^{-(\gamma - \psi)}\gamma^0\gamma^X - \psi' e^{-(\gamma - \psi)}\gamma^Z\gamma^X),$$
$$\Gamma_Y = \tfrac{1}{2}[-(1/T - \dot\psi)e^{-(\gamma - \psi)}\gamma^0\gamma^Y + \psi' e^{-(\gamma - \psi)}\gamma^Z\gamma^Y], \tag{2.5}$$

where a dot means $\partial/\partial T$ and a prime is $\partial/\partial Z$. For our signature of the metric the stress-energy tensor associated with $\Psi$ is

$$T_{\mu\nu} = -(i/4)(\bar\Psi\gamma_{(\nu}\Psi_{|\mu)} - \bar\Psi_{(|\mu}\gamma_{\nu)}\Psi), \tag{2.6}$$

where the components are in the orthonormal frame defined by (2.4), $\bar\Psi \equiv \Psi^\dagger\gamma^0$, and the parentheses mean symmetrization on $\mu$ and $\nu$. For the metric (2.1) $T_{\mu\nu}$ becomes

$$T_{00} = A_{00}, \tag{2.7a}$$
$$T_{0X} = A_{0X} - \tfrac{1}{8}(\bar\Psi\gamma^Y\gamma_5\Psi)(\gamma - 2\psi)' e^{-(\gamma - \psi)}, \tag{2.7b}$$
$$T_{0Y} = A_{0Y} + \tfrac{1}{8}(\bar\Psi\gamma^X\gamma_5\Psi)\gamma' e^{-(\gamma - \psi)}, \tag{2.7c}$$
$$T_{0Z} = A_{0Z}, \tag{2.7d}$$
$$T_{ZZ} = A_{ZZ}, \tag{2.7e}$$
$$T_{ZX} = A_{ZX} - \tfrac{1}{8}(\bar\Psi\gamma^Y\gamma_5\Psi)(\gamma - 2\psi)^{\cdot} e^{-(\gamma - \psi)}, \tag{2.7f}$$
$$T_{ZY} = A_{ZY} - \tfrac{1}{8}(\bar\Psi\gamma^X\gamma_5\Psi)(1/T - \dot\gamma)e^{-(\gamma - \psi)}, \tag{2.7g}$$
$$T_{XY} = \tfrac{1}{8}e^{-(\gamma - \psi)}[(1/T - 2\dot\psi)(\bar\Psi\gamma^Z\gamma_5\Psi) - 2\psi'(\bar\Psi\gamma^0\gamma_5\Psi)], \tag{2.7h}$$
$$T_{XX} = 0, \tag{2.7i}$$
$$T_{YY} = 0, \tag{2.7j}$$

where $A_{\mu\nu} \equiv -(i/4)[\bar\Psi\gamma_{(\nu}\nabla_{\mu)}\Psi - \nabla_{(\mu}\bar\Psi\gamma_{\nu)}\Psi]$, and $\gamma_5 \equiv i\gamma^0\gamma^X\gamma^Y\gamma^Z$. For the metric (2.1) the Einstein equations are

$$\ddot\psi + \dot\psi/T - \psi'' = 1/2T(\mathscr{T}^X{}_X - \mathscr{T}^Y{}_Y), \tag{2.8a}$$
$$\gamma' = \mathscr{T}^Z{}_0 + 2T\dot\psi\psi', \tag{2.8b}$$
$$\dot\gamma = -\mathscr{T}^0{}_0 + T(\dot\psi^2 + \psi'^2), \tag{2.8c}$$
$$\ddot\gamma - \gamma'' = \psi'^2 - \dot\psi^2 - (1/T)\mathscr{T}^Y{}_Y, \tag{2.8d}$$
$$0 = T_{00} - T_{ZZ}. \tag{2.8e}$$

[Here $\mathscr{T}^\mu{}_\nu = Te^{2(\gamma - \psi)}T^\mu{}_\nu$ and indices are raised and lowered with $\eta_{\mu\nu}$.] In addition there is a set of conditions that come from the fact that $G_{\mu\nu} = 0$ for some values of $\mu$ and $\nu$, namely

$$T_{0X} = T_{0Y} = T_{ZX} = T_{ZY} = T_{XY} = 0. \tag{2.9}$$

Equation (2.8e) is a result of our choice of Eq. (2.1) for the form of the metric. It restricts us to consider only neutrino fields. For, supposing we have satisfied $T_{00} - T_{ZZ} = 0$, the final equations of motion for the Einstein–Dirac system will be (2.8), (2.9), and the Dirac equation and its conjugate. These latter are

$$i\gamma^0\nabla_0\Psi + i\gamma^Z\nabla_Z\Psi + (i/2)e^{-(\gamma - \psi)}[(\gamma - \psi)'\gamma^Z + (1/T + \dot\gamma - \dot\psi)\gamma^0]\Psi - m\Psi = 0, \tag{2.10a}$$

and

$$i\nabla_0\bar\Psi\gamma^0 + i\nabla_Z\bar\Psi\gamma^Z + (i/2)\bar\Psi[(\gamma - \psi)'\gamma^Z + (1/T + \dot\gamma - \dot\psi)\gamma^0]e^{-(\gamma - \psi)} + m\bar\Psi = 0. \tag{2.10b}$$

Calculating $T_{\mu\nu}$ and using Eqs. (2.10), we find that

$$T_{00} - T_{ZZ} = \tfrac{1}{2} im\bar\Psi\Psi. \tag{2.11}$$

So the condition $T_{00} - T_{ZZ} = 0$ means that $m$ must be zero, which is consistent with the behavior of other massless fields that are plane symmetric with equal $T_{00}$ and $T_{ZZ}$. Henceforth we consider only massless neutrino fields.

## III. A PARTICULAR SOLUTION

To find a particular nonghost solution to (2.8)–(2.10) with $m = 0$, we begin by imposing the conditions (2.9) on $\Psi$. We use the Dirac equation or its conjugate to express $Z$ derivatives in terms of $T$ derivatives and vice versa. For instance, $T_{0X} = 0$ and $T_{ZX} = 0$ give

$$\frac{\partial}{\partial Z}(\bar\Psi\gamma^Y\gamma_5\Psi Te^\psi) = \frac{\partial}{\partial T}(\bar\Psi\gamma^Y\gamma_5\Psi Te^\psi) = 0, \tag{3.1}$$

or

$$\bar\Psi\gamma^Y\gamma_5\Psi = (B/T)e^{-\psi}, \tag{3.2}$$

and the $T_{0Y} = T_{ZY} = 0$ condition gives

$$\bar\Psi\gamma^X\gamma_5\Psi = (A/T^2)e^\psi, \tag{3.3}$$

where $A$ and $B$ are constants independent of $T$ and $Z$. In order to reduce these equations and the condition $T_{XY} = 0$ to conditions on the components of $\Psi$ we write

$$\Psi = \begin{bmatrix} a_1 e^{i\theta_1} \\ a_2 e^{i\theta_2} \\ b_1 e^{i\phi_1} \\ b_2 e^{i\phi_2} \end{bmatrix} = \begin{bmatrix} \tilde a \\ \tilde b \end{bmatrix}, \tag{3.4}$$

where $\tilde a$ and $\tilde b$ are two-spinors, and (3.2) and (3.3) become

$$2a_1a_2\cos(\theta_1-\theta_2)+2b_1b_2\cos(\phi_1-\phi_2)=(A/T^2)e^{\psi}, \tag{3.5a}$$

$$2a_1a_2\sin(\theta_1-\theta_2)+2b_1b_2\sin(\phi_1-\phi_2)=(B/T)e^{-\psi}, \tag{3.5b}$$

while $T_{XY}=0$ gives

$$(1/T-2\dot\psi)(a_1^2-a_2^2+b_1^2-b_2^2) = -4\psi'(a_1b_1+a_2b_2)\cos(\theta_1-\phi_1). \tag{3.6}$$

The Dirac equation takes the form

$$\frac{\partial}{\partial T}(T^{1/2}e^{(1/2)(\gamma-\psi)}\tilde a)+\sigma^Z\frac{\partial}{\partial Z}(T^{1/2}e^{(1/2)(\gamma-\psi)}\tilde b)=0, \tag{3.7a}$$

$$\frac{\partial}{\partial T}(T^{1/2}e^{(1/2)(\gamma-\psi)}\tilde b)+\sigma^Z\frac{\partial}{\partial Z}(T^{1/2}e^{(1/2)(\gamma-\psi)}\tilde a)=0. \tag{3.7b}$$

If we let $\theta_1=\theta_2=\phi_1=\phi_2=\theta$ and take $\theta=\theta(Z\pm T)$ and $a_A=a_{0A}T^{-1/2}e^{-(1/2)(\gamma-\psi)}$, $b_A=b_{0A}T^{-1/2}e^{-(1/2)(\gamma-\psi)}$, where $A=1,2$ and $a_{0A}$ and $b_{0A}$ are constants, Eqs. (3.7) reduce to

$$a_{01}=\mp b_{01},\quad a_{02}=\pm b_{02}. \tag{3.8}$$

Inserting (3.8) and the fact that all the phases are equal into (3.5) we find that these solutions correspond to $A=B=0$. Equation (3.6) becomes

$$4(\pm\psi'+\dot\psi-1/2T)(a_1^2-a_2^2)=0. \tag{3.9}$$

The coefficient of $(a_1^2-a_2^2)$ in this equation is the shear of the hypersurfaces $\theta=\text{const}$ if $\theta=\theta(Z\pm T)$. This quantity is not zero unless the metric is isotropic, so we can take the solution to (3.9) to be $a_1=\pm a_2$. We now have four possible solutions to the Dirac equations: (1) $\theta=\theta(Z-T)$, $a_1=b_1=a_2=-b_2$; (2) $\theta=\theta(Z-T)$, $a_1=b_1=-a_2=b_2$; (3) $\theta=\theta(Z+T)$, $a_1=-b_1=a_2=b_2$; and (4) $\theta=\theta(Z+T)$, $a_1=-b_1=-a_2=-b_2$. In Sec. IV we will discuss these solutions in more detail.

We must now return to Eqs. (2.8) and show that the solution to the Einstein equations with the above spinor solutions can be reduced to quadratures. We find $\mathscr{T}^X{}_X=\mathscr{T}^Y{}_Y=0$. Thus Eq. (2.8a) for $\psi$ reduces to a linear equation that is the same as that of the vacuum metric. This equation is the key equation mentioned in the Introduction. Since $\mathscr{T}^Y{}_Y=0$, and since $T^{\mu\nu}{}_{;\nu}=0$ by the Dirac equation, (2.8d) is automatically valid if the key equation is solved for $\psi$. For Eqs. (2.8b) and (2.8c), we must check the integrability condition. They can be integrated if the partial derivative with respect to $T$ (2.8b) equals the partial derivative of (2.8a) with respect to $Z$. If the key equation is satisfied this condition reduces to $(\mathscr{T}^Z{}_0)^\bullet-(\mathscr{T}^0{}_0)'=0$. It is not difficult to show that this condition is satisfied for the four solutions given above.

We now have the complete solution, at least up to quadratures. The solutions for $\Psi$ given above are complete except for the functional values of $\psi$ and $\gamma$ which can be obtained from (2.8). Since $\mathscr{T}^X{}_X=\mathscr{T}^Y{}_Y=0$, there exist solutions for (2.8a),[5] and these solutions can be inserted in (2.8b) and (2.8c) to give $\gamma$ by quadratures. As we will show in the next section $T^0{}_0$ and $T^Z{}_0$ are proportional to $(\sqrt{-g})^{-1}$, so $\mathscr{T}^0{}_0$ and $\mathscr{T}^Z{}_0$ have no explicit dependence on metric terms, so (2.8b) and (2.8c) are integrable directly, without the need of an integrating factor which would be necessary if $\mathscr{T}^0{}_0$ and $\mathscr{T}^Z{}_0$ contained $\gamma$ explicitly. This solution is the solution mentioned in the Introduction.

## IV. DISCUSSION AND CONCLUSIONS

We need first to show that our solution is not a ghost, that is, that $T_{\mu\nu}$ is not identically zero. It is easy to show that for all four of the solutions given above

$$T_{00}=2e^{-(\gamma-\psi)}a^2\dot\theta, \tag{4.1a}$$

$$T_{0Z}=2e^{-(\gamma-\psi)}a^2\dot\theta, \tag{4.1b}$$

where $a=a_1$, so our solution is not a ghost. We can use these expressions to show that $\mathscr{T}^0{}_0$ and $\mathscr{T}^Z{}_0$ do not depend explicitly on the metric components. Taking $\mathscr{T}^0{}_0$ as a paradigm, we find that

$$\mathscr{T}^0{}_0=Te^{2(\gamma-\psi)}T^0{}_0 = -2Te^{2(\gamma-\psi)}e^{-(\gamma-\psi)}a_{01}^2T^{-1}e^{-(\gamma-\psi)}\dot\theta=-2a_{01}^2\dot\theta, \tag{4.2}$$

where the constant $a_{01}$ is defined in Sec. III.

We conclude our discussion of $\Psi$ by classifying the four solutions given in Sec. III according to current and helicity. We calculate $j^\mu=\bar\Psi\gamma^\mu\Psi$ for our four solutions, and find that for (1) and (2), $j^\mu=(4a^2,0,0,4a^2)$, while (3) and (4) give $j^\mu=(4a^2,0,0,-4a^2)$. Since the momentum of these solutions is obviously in the $\pm Z$ direction, the helicity operator is proportional to $\Sigma_3$, where

$$\Sigma_3=\begin{pmatrix}\sigma^3 & 0\\ 0 & \sigma^3\end{pmatrix}, \tag{4.3}$$

and if we calculate $\bar\Psi\Sigma_3\Psi$, we find that it is zero for all four solutions. According to the notation of Schweber,[15] the solutions (1) and (3) are positive energy, and (2) and (4) are negative energy, and all of the solutions are the appropriate sums of states of helicity $+1$ and helicity $-1$ to give zero helicity. The final classification is (1) positive energy, current in the $+Z$ direction, zero helicity; (2) negative energy, current in the $+Z$ direction, zero helicity; (3) positive energy, current in the $-Z$ direction, zero helicity; and (4) negative energy, current in the $-Z$ direction, zero helicity.

The solutions for $\Psi$ and solutions of Eq. (2.8a) and the integration of (2.8b) and (2.8c) give us the exact nonghost solution to the Einstein–Dirac field equations promised in the Introduction. There should exist other solutions to the problem if the condition that all the phases in (3.4) be equal is relaxed.


## ACKNOWLEDGMENTS

We thank Mr. H. Harleston for very helpful reading and criticism of this manuscript.

This work was supported by a grant from CONACyT, by NSF grants PHY81-07381, PHY84-04931, and INT78-22553.

[1]J. Centrella, Ph.D. thesis, Cambridge Univeristy, 1978.
[2]J. Centrella, Astrophys. J. **241**, 875 (1980).
[3]J. Centrella and R. Matzner, Astrophys. J. **230**, 361 (1979).
[4]J. Centrella and R. Matzner, Phys. Rev. D **25**, 930 (1982).
[5]R. Matzner, M. Rosenbaum, and M. Ryan, Jr., J. Math. Phys. **23**, 1984 (1982).
[6]A. Einstein and N. Rosen, J. Franklin Inst. **223** 43 (1937).
[7]R. Gowdy, Ann. Phys. (NY) **83**, 203 (1974).
[8]Ch. Charach, Phys. Rev. D **19**, 3516 (1979); **19**, 1058 (1979).
[9]See, for example, T. Davis and J. Ray, Phys. Rev. D**9**, 331 (1974); and C. Collinson and P. Morris, J. Phys. A **6**, 915 (1973).
[10]M. Henneaux, Gen. Rel. Grav. **12**, 137 (1980); Phys. Rev. D **21**, 857 (1980); Ann. Inst. Henri Poincaré A **34**, 329 (1981).
[11]C. Isham and J. Nelson, Phys. Rev. D **10**, 3226 (1974).
[12]O. Obregon and M. Ryan, J. Math. Phys. **22**, 623 (1981).
[13]R. Jantzen, J. Math. Phys. **23**, 1137 (1982).
[14]C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[15]S. Schweber, *Introduction to Relativistic Quantum Field Theory* (Princeton U. P., Princeton, NJ, 1961).

# Positivity properties of phase-plane distribution functions

A. J. E. M. Janssen
*Philips Research Laboratories, 5600 JA Eindhoven, The Netherlands*

(Received 1 November 1983; accepted for publication 16 February 1984)

The aim of this paper is to compare the members of Cohen's class of phase-plane distributions with respect to positivity properties. It is known that certain averages (which are in a sense compatible with Heisenberg's uncertainty principle) of the Wigner distribution over the phase-plane yield non-negative values for all states. It is shown in this paper that the Wigner distribution is unique in this respect among the members of Cohen's class that have correct marginals or that satisfy Moyal's formula for all states. The subset of members of Cohen's class (not necessarily satisfying one of these two conditions) with positivity properties comparable with those for the Wigner distribution is shown to be rather small.

PACS numbers: 03.65. − w, 02.30. + g

## I. INTRODUCTION

In this Introduction we present in a rather informal way some known facts about Cohen's class of phase-plane distribution functions, and we indicate what we are aiming at in this paper. Cohen's class is parametrized by means of a function $\Phi$ of two variables[1]: for any such $\Phi$ we have the family of phase-plane distribution functions

$$C_f^{(\Phi)}(q,p) = \int\int\int \exp[-2\pi i(\theta q + \tau p - \theta u)]\, \Phi(\theta,\tau) \times f(u + \tfrac{1}{2}\tau)\overline{f(u - \tfrac{1}{2}\tau)}\, d\theta\, d\tau\, du \quad [(q,p)\in\mathbb{R}^2], \tag{1.1}$$

where $f$ is an arbitrary state (all integrations are over the real line, unless indicated otherwise). Of course, in order for this definition to make sense certain assumptions on $\Phi$ as well as on $f$ should be made. In Sec. II a convenient mathematical setting for dealing with rather general $\Phi$'s in (1.1) is presented. Any family $C_f^{(\Phi)}$ ($f$ arbitrary state) can be used to give a formulation of quantum mechanics in the phase plane of position $q$ and momentum $p$. In fact, it can be shown that any bilinear map $f\to C_f$, mapping states $f$ onto functions $C_f$ of the phase-plane variables $(q,p)$, satisfying

$$C_f(q+a,p+b) = C_{T_aR_bf}(q,p) \quad [(q,p)\in\mathbb{R}^2] \tag{1.2}$$

for all states $f$ and all $(a,b)\in\mathbb{R}^2$ can be brought into the form (1.1). Here $T_a$ and $R_b$ are the shift operators, defined, respectively, by

$$(T_af)(q) = f(q+a),$$
$$(R_bf)(q) = e^{-2\pi ibq}f(q) \quad (q\in\mathbb{R}), \tag{1.3}$$

for all $f$ and all $(a,b)\in\mathbb{R}^2$. It is easily verified that any $C_f = C_f^{(\Phi)}$ as in (1.1) satisfies (1.2) for all $f$ and all $(a,b)\in\mathbb{R}^2$.

The choice $\Phi(\theta,\tau) = 1$ in (1.1) yields the Wigner distribution[2] of $f$, viz.

$$W_f(q,p) = \int e^{-2\pi ipt} f\left(q + \frac{1}{2}t\right)\overline{f\left(q - \frac{1}{2}t\right)}dt \quad [(q,p)\in\mathbb{R}^2]. \tag{1.4}$$

In a way one can consider the Wigner distribution as the basic distribution of Cohen's class from which all others can be derived[3]: one has

$$C_f^{(\Phi)}(q,p) = \int\int \varphi(q-a,p-b)W_f(a,b)\,da\,db \quad [(q,p)\in\mathbb{R}^2], \tag{1.5}$$

where $\varphi$ is the double Fourier transform of $\Phi$, given by

$$\varphi(q,p) = \int\int e^{-2\pi i(\theta q + \tau p)}\Phi(\theta,\tau)\,d\theta\,d\tau \quad [(q,p)\in\mathbb{R}^2]. \tag{1.6}$$

This $\varphi$ must be treated as a generalized function, e.g., $\varphi(q,p) = \delta(q)\delta(p)$ for the Wigner distribution case, whereas $\Phi$ is usually smooth.

The class of all possible phase-plane distributions can be restricted considerably by imposing certain "natural" requirements. We consider in this paper four additional conditions.

(a) $C_f^{(\Phi)}$ yields the "correct" marginal distributions for all states $f$ [see (1.7)].

(b) $C_f^{(\Phi)}$ has finite support properties [see (1.11) and (1.12)].

(c) $C_f^{(\Phi)}$ is such that Moyal's formula holds for all states $f$ and $g$ [see (1.15)].

(d) $C_f^{(\Phi)}$ is a non-negative distribution for all states $f$.

Each of the requirements (a), (b), (c), and (d) has consequences for $\Phi$ (and $\varphi$); it is well known that not all four conditions are compatible. However, the Wigner distribution satisfies (a), (b), and (c), while also certain positivity properties hold.

The condition (a) means that for all states $f$ we should have

$$\int C_f^{(\Phi)}(q,p)dp = |f(q)|^2 \quad (q\in\mathbb{R}),$$
$$\int C_f^{(\Phi)}(q,p)dq = |(\mathscr{F}f)(p)|^2 \quad (p\in\mathbb{R}). \tag{1.7}$$

Here $\mathscr{F}$ denotes the Fourier transform, given for all $f$ by

$$(\mathscr{F}f)(p) = \int e^{-2\pi iqp}f(q)dq \quad (p\in\mathbb{R}). \tag{1.8}$$

It can be shown[1,3,4] that (1.7) holds for all states $f$ if and only if

 0022-2488/84/072240-13$02.50  

$$\Phi(0,\tau) = \Phi(\theta,0) = 1 \quad [(\theta,\tau)\in\mathbb{R}^2], \tag{1.9}$$

or, equivalently,

$$\int \varphi(q,p)dp = \delta(q) \quad (q\in\mathbb{R}), \qquad \int \varphi(q,p)dq = \delta(p) \quad (p\in\mathbb{R}). \tag{1.10}$$

For condition (b) it is required[3] that for all states $f$ and all $(Q,P)\in\mathbb{R}^2$

$$f(q) = 0 \quad (|q| > Q) \Rightarrow C_f^{(\Phi)}(q,p) = 0 \quad (|q| > Q), \tag{1.11}$$

and

$$(\mathscr{F}f)(p) = 0 \quad (|p| > P) \Rightarrow C_f^{(\Phi)}(q,p) = 0 \quad (|p| > P). \tag{1.12}$$

It can be shown[3] that validity of (1.11) for all $f$ is equivalent to

$$\int e^{-2\pi i\theta q}\Phi(\theta,\tau)d\theta = 0 \quad (|q| > |\tau|/2), \tag{1.13}$$

for all $\tau\in\mathbb{R}$; similarly, validity of (1.12) for all $f$ is equivalent to

$$\int e^{-2\pi i\tau p}\Phi(\theta,\tau)d\tau = 0 \quad (|p| > |\theta|/2), \tag{1.14}$$

for all $\theta\in\mathbb{R}$. That is, $\Phi(\cdot,\tau)$, $\Phi(\theta,\cdot)$ are functions of the Paley–Wiener kind[5] with type $\leqslant|\tau|/2$, $\leqslant|\theta|/2$, respectively, for $(\theta,\tau)\in\mathbb{R}^2$ when the finite support properties are satisfied.

For property (c) to hold, we must have that Moyal's formula[6–8]

$$\int\int C_f^{(\Phi)}(q,p)\,\overline{C_g^{(\Phi)}(q,p)}dq\,dp = |(f,g)|^2 \tag{1.15}$$

is valid for all states $f$ and $g$. It has been shown[9] that validity of (1.15) for all $f$ and $g$ is equivalent to

$$|\Phi(\theta,\tau)| = 1 \quad [(\theta,\tau)\in\mathbb{R}^2], \tag{1.16}$$

or

$$(\varphi * \tilde{\varphi})(q,p) = \delta(q)\delta(p) \quad [(q,p)\in\mathbb{R}^2], \tag{1.17}$$

where $\tilde{\varphi}(q,p) = \overline{\varphi(-q,-p)}$, and $*$ denotes convolution over $\mathbb{R}^2$. A further result[9] is that validity of (1.15) for all $f$ and $g$, together with validity of (1.7), (1.11), and (1.12) for all $f$, implies that $\Phi$ takes the special form

$$\Phi(\theta,\tau) = \Phi_\alpha(\theta,\tau) = \exp(2\pi i\alpha\theta\tau) \quad [(\theta,\tau)\in\mathbb{R}^2] \tag{1.18}$$

for some $\alpha\in\mathbb{R}$ with $|\alpha|\leqslant\frac{1}{2}$. In that case $\varphi$ is given by

$$\varphi(q,p) = \varphi_\alpha(q,p) = \alpha^{-1}\exp(-2\pi iqp/\alpha) \text{ or } \delta(q)\delta(p) \quad [(q,p)\in\mathbb{R}^2] \tag{1.19}$$

according as $\alpha\neq 0$ or $\alpha = 0$, and $C_f^{(\Phi)}$ takes the special form[9]

$$C_f^{(\Phi)}(q,p) = C_f^{(\Phi_\alpha)}(q,p) = \int e^{-2\pi ipt}f\left(q + t\left(\frac{1}{2}-\alpha\right)\right) \times \overline{f\left(q - t\left(\frac{1}{2}+\alpha\right)\right)}dt \quad [(q,p)\in\mathbb{R}^2]. \tag{1.20}$$

It is interesting to note that for any state $f$ and any $(a,b)\in\mathbb{R}^2$ the global spread

$$\int\int [(q-a)^2 + (p-b)^2]\,|C_f^{(\Phi_\alpha)}(q,p)|^2\,dq\,dp \tag{1.21}$$

of $C_f^{(\Phi_\alpha)}$ around $(a,b)$ is minimal for $\alpha = 0$, the Wigner distribution case. Choosing for $(a,b)$ the center of gravity[8] of $C_f^{(\Phi_\alpha)}$, which is independent of $\alpha$ and equals[9]

$$(a,b) = \left(\int q|f(q)|^2\,dq,\ \int p|(\mathscr{F}f)(p)|^2\,dp\right), \tag{1.22}$$

we see that the Wigner distribution behaves, in some sense, best with respect to spread among the members of Cohen's class that satisfy conditions (a), (b), and (c). This is some indication that the Wigner distribution is to be preferred over the other members of Cohen's class. One may find this argument not entirely convincing yet, for one has to restrict oneself to distributions satisfying the strong condition that Moyal's formula is satisfied and this excludes, for example, the family of distributions ($f$ arbitrary state)

$$\mathrm{Re}[e^{2\pi iqp}\,\overline{f(q)}(\mathscr{F}f)(p)] \quad [(q,p)\in\mathbb{R}^2], \tag{1.23}$$

which was considered by Margenau and Hill.[10]

We finally discuss condition (d). This condition says that for all $f$ it should hold that[11]

$$C_f^{(\Phi)}(q,p)\geqslant 0 \quad [(q,p)\in\mathbb{R}^2]. \tag{1.24}$$

It has been shown[12] that validity of (1.7) and (1.24) for all states $f$ is not possible. This does not contradict the result of Ref. 13 where to every state a non-negative function of $(q,p)$ with correct marginal distributions is assigned in a nonbilinear way.

With respect to positivity properties only the Wigner distribution has been studied in some detail[14–16] as far as we know. It is exactly the purpose of this paper to compare the general phase-plane distribution functions on this point with the Wigner distribution. The best known positivity property of the Wigner distribution[17–21] reads: for all states $f$, all $\gamma > 0$, $\delta > 0$ with $\gamma\delta\leqslant 1$, and all $(a,b)\in\mathbb{R}^2$ we have

$$\int\int \exp[-2\pi\gamma(q-a)^2 - 2\pi\delta(p-b)^2]\,W_f(q,p)dq\,dp\geqslant 0. \tag{1.25}$$

This paper concentrates on finding out for what $\Phi$ and what $\gamma$, $\delta$ inequality (1.25) still holds for all $f$, $(a,b)$ when $W_f$ is replaced by the more general phase-plane distribution $C_f^{(\Phi)}$. In connection with (1.25) we note that the following has been proved for the Wigner distribution. Hudson[17] has shown that $W_f$ takes negative values unless $f$ is a Gaussian. The argument used by Hudson was augmented[21] to show that, if $\gamma\delta > 1$, any $f$ for which (1.25) is non-negative for all $(a,b)\in\mathbb{R}^2$ must be a (possibly degenerate) Gaussian (in Ref. 21 certain generalized functions are allowed; we turn to these in Sec. II). It is not clear how a result of similar strength can be shown to hold generally for the distributions of Cohen. We have, e.g., with $\Phi(\theta,\tau) = \cos\pi\theta\tau$ [which yields (1.23)] that $C_f^{(\Phi)}(q,p)\geqslant 0$ for $f(q) = \cos 2\pi q$. Nevertheless the following results will be proved in this paper. Assume that $\Phi$ is such that (1.7) is satisfied for all $f$. Under a mild smoothness and growth condition[22] on $\Phi$ we have the following.

(1) If $\gamma\delta > 1$, then there is no $\Phi$ such that (1.25) (with $C_f^{(\Phi)}$ instead of $W_f$) holds for all $f$ and all $(a,b)\in\mathbb{R}^2$.

(2) If $\gamma\delta = 1$, then the only $\Phi$ for which (1.25) (with $C_f^{(\Phi)}$ instead of $W_f$) holds for all $f$ and all $(a,b)\in\mathbb{R}^2$ equals $\Phi(\theta,\tau) = 1$ (Wigner distribution case).

We shall prove that a similar result holds when validity of (1.7) is replaced by validity of (1.15) for all $f$ and $g$. We shall in addition show that validity of (1.25) (with $C_f^{(\Phi)}$ instead of $W_f$) imposes severe restrictions on $\Phi$ if $\gamma\delta < 1$ and (1.7) is satisfied for all $f$, or if $\gamma\delta \geqslant 1$.

The further plan of this paper is as follows. In Sec. II we give a mathematical setting that allows us to consider functions $\Phi$ with mild restrictions on growth. We furthermore recall in Sec. II the main results of Ref. 16, and we extend these results somewhat. In Ref. 16 conditions for a function $K(q,p)$ are given that ensure that

$$\int\int K(q,p)W_f(q,p)dq\,dp \tag{1.26}$$

is non-negative for all $f$. It is clear that these results will be useful, since (1.5) and (1.25) show that non-negativity of (1.25) [with $C_f^{(\Phi)}$ instead of $W_f$ and $(a,b) = (0,0)$] for all $f$ is equivalent to non-negativity of (1.26) for all $f$, where $K$ is the convolution of $\varphi(q,p)$ and $\exp(-2\pi\gamma q^2 - 2\pi\delta p^2)$. In Sec. III we consider the case that no other condition than non-negativity of (1.25) [with $C_f^{(\Phi)}$ instead of $W_f$ and $(a,b) = (0,0)$] for all $f$ is imposed; in Sec. IV we require in addition correct marginals or validity of Moyal's formula.

## II. MATHEMATICAL SETTING AND RESULTS ON POSITIVITY FOR THE WIGNER DISTRIBUTION

As we have to discuss rather general functions $\Phi$ it is convenient to restrict the states $f$ to a certain space of test functions. We consider the space $S$ of smooth functions; this function space has been proposed in Ref. 8 as a setting suited for doing Wigner distribution analysis. It is the same space as the one used in Refs. 16, 21, and 23. To describe it briefly we denote, for $n = 0,1,\ldots$, by $\psi_n$ the $n$th Hermite function,

$$\psi_n(q) = \frac{(-1)^n 2^{1/4} e^{\pi q^2} (d/dq)^n e^{-2\pi q^2}}{n!(4\pi)^{n/2}} \quad (q\in\mathbb{R}); \tag{2.1}$$

the normalization has been chosen in such a way that

$$e^{\pi q^2 - 2\pi(q-w)^2} = 2^{-1/4} \sum_{n=0}^{\infty} \frac{(2w\sqrt{\pi})^n}{\sqrt{n!}} \psi_n(q) \quad (q\in\mathbb{R}, w\in\mathbb{C}). \tag{2.2}$$

The space $S$ consists of all functions $f$ whose Hermite coefficients $(f,\psi_n)$ satisfy an estimate

$$(f,\psi_n) = O(e^{-n\alpha}) \quad (n = 0,1,\ldots), \tag{2.3}$$

for some $\alpha > 0$. It can be shown that the space $S$ is identical to the set of (restrictions to the real axis of) entire functions $g$ for which there are $M > 0$, $A > 0$, $B > 0$ such that

$$|g(x+iy)| \leqslant M\exp(-\pi Ax^2 + \pi By^2) \quad [(x,y)\in\mathbb{R}^2]. \tag{2.4}$$

A sequence $(f_k)_k$ in $S$ is said to converge to zero when, for some $\alpha > 0$, $\sup_{n=0,1,\ldots} e^{n\alpha}|(f_k,\psi_n)| \to 0$ when $k\to\infty$.

The space $S^*$ consists of all continuous linear functionals on $S$. It can be shown that for $F\in S^*$

$$(F,\psi_n) = O(e^{n\alpha}) \quad (n = 0,1,\ldots), \tag{2.5}$$

for all $\alpha > 0$. The smoothing operators $N_\alpha$ with $\text{Re}\,\alpha > 0$ play an important role; they map $S^*$ into $S$ and are defined by

$$(N_\alpha F)(q) = \sum_{n=0}^{\infty} (F,\psi_n)e^{-(n+1/2)\alpha}\psi_n(q) \quad (F\in S^*, q\in\mathbb{C}). \tag{2.6}$$

As an integral operator of $L^2(\mathbb{R})$, $N_\alpha$ has the kernel $K_\alpha$ given by

$$K_\alpha(q,p) = \left(\frac{1}{\sinh\alpha}\right)^{1/2} \exp\left(-\frac{\pi}{\sinh\alpha} \times [(q^2+p^2)\cosh\alpha - 2qp]\right) = \sum_{n=0}^{\infty} e^{-(n+1/2)\alpha}\psi_n(q)\psi_n(p) \quad [(q,p)\in\mathbb{R}^2]. \tag{2.7}$$

The identity in (2.7) is just one way to write Mehler's formula

$$\left(\frac{2}{1-w^2}\right)^{1/2} \exp\left(-\pi(q^2+p^2)\frac{1+w^2}{1-w^2} + 4\pi\frac{qpw}{1-w^2}\right) = \sum_{n=0}^{\infty} w^n\psi_n(q)\psi_n(p) \quad [(q,p)\in\mathbb{C}^2, |w|<1]. \tag{2.8}$$

The spaces $S^2$ and $S^{2*}$ of smooth and generalized functions of two variables can be defined in a similar fashion. An important formula, relating smoothing operators and Wigner distributions,[24] reads

$$(N_{\alpha,2}V_f)(q,p) = V_{N_\alpha f}(q,p) \quad [(q,p)\in\mathbb{R}^2, \text{Re}\,\alpha > 0] \tag{2.9}$$

for $f\in L^2(\mathbb{R})$. Here $N_{\alpha,2}$ is the smoothing operator for functions of two variables [whose kernel $K_{\alpha,2}(q,p;x,y)$ equals $K_\alpha(q,x)K_\alpha(p,y)$], and

$$V_f(q,p) = \frac{1}{\sqrt{2}} W_f\left(\frac{q}{\sqrt{2}}, \frac{p}{\sqrt{2}}\right) \quad [(q,p)\in\mathbb{R}^2] \tag{2.10}$$

for $f\in L^2(\mathbb{R})$. We note[25] that $V_F$ (and hence $W_F$) can be defined for $F\in S^*$ and that $V_F\in S^{2*}$.

Another useful formula[26] is

$$W_{N_{i\theta}f}(q,p) = W_f(q\cos\theta + p\sin\theta, p\cos\theta - q\sin\theta) \quad [(q,p)\in\mathbb{R}^2], \tag{2.11}$$

which holds for all real $\theta$ and all $f\in S$.

In spite of the rather heavy machinery we have developed here, we shall usually manipulate with generalized functions in a rather carefree manner; we shall give details only in cases where the verification are not straightforward.

We now turn to positivity properties of the Wigner distribution. We have, for $n = 0,1,\ldots$,[27]

$$W_{\psi_n}(q,p) = 2(-1)^n \exp[-2\pi(q^2+p^2)] \times L_n[4\pi(q^2+p^2)] \quad [(q,p)\in\mathbb{R}^2]. \tag{2.12}$$

Here $L_n$ is the $n$th Laguerre polynomial,

$$L_n(x) = \sum_{j=0}^{n} \binom{n}{j} \frac{(-x)^j}{j!} \quad (x\geqslant 0; n = 0,1,\ldots), \tag{2.13}$$

for which a generating formula[28] is given by

$$(1-w)^{-1}\exp[-xw(1-w)^{-1}] = \sum_{n=0}^{\infty} w^n L_n(x) \quad (|w|<1, x\geqslant 0). \tag{2.14}$$

Formula (2.12) can be used to show the identity[29]

$$\int\int W_f(q,p)K[2\pi(q^2+p^2)]dq\,dp = \sum_{n=0}^{\infty} (-1)^n|(f,\psi_n)|^2 \int_0^\infty e^{-r}K(r)L_n(2r)dr, \tag{2.15}$$

where $f\in S$ and $K$: $[0,\infty)\to\mathbb{C}$ is measurable and satisfies

$$\int_0^\infty |K(x)|^2 e^{-\epsilon x}\,dx<\infty \quad (\epsilon>0). \tag{2.16}$$

Now positivity properties of the Wigner distribution result on taking non-negative functions $K$ with the property that

$$(-1)^n\int_0^\infty e^{-r}K(r)L_n(2r)dr\geqslant 0 \quad (n=0,1,\ldots). \tag{2.17}$$

In Ref. 16 a large number of examples of such $K$'s have been given. We mention in particular the choices

$$K(r)=r^n e^{-\rho r} \quad (0\leqslant\rho\leqslant 1, \quad n=0,1,\ldots), \tag{2.18}$$

$$K(r)=r^\alpha \quad (\alpha\geqslant -\tfrac{1}{2}). \tag{2.19}$$

The following positivity property is new as far as we know.

**Theorem 2.1:** Let $K$: $[0,\infty)\to[0,\infty)$ be nondecreasing, and assume that $K(x)=O[\exp(\epsilon x)]$ for some $\epsilon<1$. Then (2.17) holds.

*Proof*: It follows from Bonnet's theorem[30] that for all $A>0$ there is an $x_0(A)\in[0,A]$ such that

$$(-1)^n\int_0^A e^{-r}L_n(2r)K(r)dr = (-1)^n K(A-)\int_{x_0(A)}^A e^{-r}L_n(2r)dr. \tag{2.20}$$

It is easy to check from formula (2.14) that

$$(-1)^n\int_r^\infty e^{-s}L_n(2s)ds=S_n(r)+S_{n-1}(r) \quad (r\geqslant 0), \tag{2.21}$$

where

$$S_n(r)=\sum_{k=0}^n(-1)^k e^{-r}L_k(2r) \quad (n\geqslant -1, r\geqslant 0). \tag{2.22}$$

Since in Ref. 28, Problem 100, p. 392, shows that $S_n(r)\geqslant 0$ for $n\geqslant -1$, $r\geqslant 0$, it follows that

$$(-1)^n K(A-)\int_{x_0(A)}^\infty e^{-r}L_n(2r)dr\geqslant 0 \quad (A\geqslant 0, n=0,1,\ldots). \tag{2.23}$$

The proof is easily completed by noting that, for $n=0,1,\ldots,$

$$K(A-)\int_A^\infty e^{-r}L_n(2r)dr\to 0 \quad (A\to\infty). \tag{2.24}$$

*Notes:* (1) Assume that $K$ is infinitely many times differentiable, and that $K(r)$ and all its derivatives are $O(e^{\epsilon r})$ for some $\epsilon<1$. Then (2.17) holds if and only if

$$\int_0^\infty r^n e^{-r}\left(\frac{d}{dr}\right)^n\left[e^{r/2}K\left(\frac{r}{2}\right)\right]dr\geqslant 0 \quad (n=0,1,\ldots). \tag{2.25}$$

This follows on using $e^{-r}L_n(r)=1/n!(d/dr)^n(e^{-r}r^n)$ and performing $n$ partial integrations in (2.17).

(2) Since both $K(r)=r^\alpha(\alpha\geqslant -\frac{1}{2})$ and $K(r)=e^{-\rho r}(0\leqslant\rho\leqslant 1)$ satisfy (2.17), one may ask whether $K(r)=r^\alpha e^{-\rho r}$ satisfies (2.17). Well, it does not unless $\alpha$ is an integer. It can be shown from the formula (2.14) that, for $n=0,1,\ldots,$

$$(-1)^n\int_0^\infty e^{-r}L_n(2r)r^\alpha e^{-\rho r}dr =(1-\rho)^{-\alpha-1}\Gamma(\alpha+1)C_{w^n}\left[(1+w)^\alpha\times\left(\frac{1+\rho}{1-\rho}-w\right)^{-\alpha-1}\right]. \tag{2.26}$$

Here $C_{w^n}$ denotes "coefficient of $w^n$ in." Now Darboux's method[31] can be used to find the asymptotic behavior of the coefficients of the function $(1+w)^\alpha[(1+\rho)/(1-\rho)-w]^{-\alpha-1}$. We get $[a=(1+\rho)/(1-\rho)]$

$$(-1)^n\int_0^\infty e^{-r}L_n(2r)r^\alpha e^{-\rho r}dr =\frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}\times\left[1+\frac{(\alpha+1)^2}{(a+1)(\alpha-n+1)}+O\left(\frac{1}{n^2}\right)\right] \quad (n=0,1,\ldots), \tag{2.27}$$

and this oscillates for large $n$ when $\alpha$ is noninteger. This example shows that the condition (2.17) is rather intricate.

(3) We give an application of formula (2.15) which has nothing to do with the main subject of this paper. In the context of the Weyl quantization map we can express the left-hand side of (2.15) as $(T_K f, f)$, where $T_K$ is the linear operator whose Weyl symbol[32] equals $K[2\pi(q^2+p^2)]$. Denote by $H$ the Hermite operator $-(1/4\pi^2)(d^2/dq^2)+q^2$, whose Weyl symbol equals $q^2+p^2$. One can now ask how well $f(q^2+p^2)$ is an approximation to the Weyl symbol of $f(H)$. As an example we consider $f(r)=r^{1/2}$, and to that end we choose $K(r)=(r/2\pi)^{1/2}$ in (2.17). Now $T_K$ is an operator whose matrix relative to the basis $(\psi_n)_{n=0,1,\ldots}$ of Hermite functions is a diagonal matrix, with diagonal elements

$$(T_K\psi_n,\psi_n)=\frac{(-1)^n}{\sqrt{2\pi}}\int_0^\infty e^{-r}L_n(2r)r^{1/2}dr =2^{-3/2}C_{w^n}[(1-w)^{1/2}/(1+w)^{1/2}]. \tag{2.28}$$

By using Darboux's method, one can show that

$$(T_K\psi_n,\psi_n)=\pi^{-1/2}(n+\tfrac{1}{2})^{1/2}[1+O(1/n)] \quad (n=0,1,\ldots). \tag{2.29}$$

At the same time $(\sqrt{H}\psi_n,\psi_n)=\pi^{-1/2}(n+\frac{1}{2})^{1/2}$ for $n=0,1,\ldots$. Hence $T_K-\sqrt{H}$ is a diagonal operator (relative to the $\psi_n$'s) with diagonal elements that are $O(n^{-1/2})$. This shows that $T_K-\sqrt{H}$ is of Schatten's $p$ class with $p>2$. Of course, all sorts of generalizations are possible here.

## III. PHASE-PLANE DISTRIBUTION FUNCTIONS WITH NON-NEGATIVE GAUSSIAN AVERAGES

Let $\gamma>0$. In this section we want to find out for which $\Phi$ as in (1.1) or $\varphi$ as in (1.6) we have

$$\int\int C_f^{(\Phi)}(q,p)\exp[-2\pi\gamma(q^2+p^2)]dq\,dp\geqslant 0 \tag{3.1}$$

for all $f\in S$. We require here that $\Phi\in S^{2*}$ or $\varphi\in S^{2*}$, for then formula (1.5) shows that $C_f^{(\Phi)}$ is the convolution of $\varphi\in S^{2*}$

and $W_f \in S^2$, and this is a smooth function that can be integrated against any Gaussian as in (3.1). For the details concerning convolution theory in the spaces $S, S^2, S^*, S^{2*}$, one may consult Ref. 33. We consider here only radially symmetric Gaussian weight functions since the more general Gaussians $\exp[-2\pi(\gamma q^2 + \delta p^2)]$ can be dealt with by considering $\Phi(\alpha^{-1}\theta, \alpha\tau)$ instead of $\Phi(\theta,\tau)$ $[\alpha = (\delta/\gamma)^{1/2}]$. We can write (3.1) as

$$\int\int G(a,b) W_f(a,b) da\, db, \tag{3.2}$$

with $G$ the convolution of $\varphi$ and $\exp[-2\pi\gamma(q^2+p^2)]$, i.e.,

$$G(a,b) = \int\int \varphi(q-a, p-b) \times \exp[-2\pi\gamma(q^2+p^2)] dq\, dp \quad [(a,b)\in\mathbb{R}^2]. \tag{3.3}$$

The following results show that a $G$ for which (3.2) is non-negative for all $f\in S$ cannot decay too rapidly.

*Lemma 3.1:* Assume that $G:\mathbb{R}^2\to\mathbb{R}$ is bounded and measurable and satisfies $G(a,b) = o(\exp[-2\pi(a^2+b^2)])$ $(a^2+b^2\to\infty)$. Then (3.2) is negative for some $f\in S$, unless

$$\int_0^{2\pi} G(R\cos\theta, R\sin\theta) d\theta = 0 \quad (R\geqslant 0). \tag{3.4}$$

*Proof:* Part of the argument given here can also be found in Ref. 16. Suppose that (3.2) is non-negative for all $f\in S$, and let

$$K(r) = \frac{1}{2\pi}\int_0^{2\pi} G\left(\sqrt{\frac{r}{2\pi}}\cos\theta, \sqrt{\frac{r}{2\pi}}\sin\theta\right) d\theta. \quad (r\geqslant 0). \tag{3.5}$$

We have for any $f\in S$ by (2.11)

$$\int\int K[2\pi(q^2+p^2)] W_f(q,p) dq\, dp = \frac{1}{2\pi}\int_0^{2\pi}\left(\int\int G(q,p) W_{N_{\theta}f}(q,p) dq\, dp\right) d\theta \geqslant 0. \tag{3.6}$$

Therefore, by (2.15), we have, for all $n$,

$$a_n := (-1)^n \int_0^\infty e^{-r} K(r) L_n(2r) dr \geqslant 0. \tag{3.7}$$

It follows from the formula[34]

$$r^\alpha = 2^{-\alpha}\sum_{n=0}^\infty (-1)^n \frac{\Gamma^2(\alpha+1)}{n!\,\Gamma(\alpha-n+1)} L_n(2r) \quad (\alpha > -1, r>0) \tag{3.8}$$

that

$$\int_0^\infty r^\alpha e^{-r} K(r) dr = 2^{-\alpha}\sum_{n=0}^\infty \frac{\Gamma^2(\alpha+1)}{n!\,\Gamma(\alpha-n+1)} a_n. \tag{3.9}$$

The left-hand side of (3.9) can be shown to be $o[2^{-\alpha}\Gamma(\alpha+1)]$ as $\alpha\to\infty$. Indeed, this follows from the assumptions on $G$ implying that $K(r) = o(e^{-r})$ as $r\to\infty$. The sum on the right-hand side of (3.9) has, for integer $\alpha$, non-negative terms only. Hence, for any $m = 0,1,\ldots$, we have

$$2^{-\alpha}\sum_{n=0}^\infty \frac{\Gamma^2(\alpha+1)}{n!\,\Gamma(\alpha-n+1)} a_n \geqslant 2^{-\alpha}\frac{\Gamma^2(\alpha+1)}{m!\,\Gamma(\alpha-m+1)} a_m = 2^{-\alpha}\Gamma(\alpha+1)(\alpha-m+1)\cdots(\alpha+1) a_m/m! \quad (\alpha = m, m+1,\ldots). \tag{3.10}$$

This is certainly not $o[2^{-\alpha}\Gamma(\alpha+1)]$ as $\alpha\to\infty$, unless all $a_m$ are 0. Since the functions $e^{-r}L_n(2r), n = 0,1,\ldots$ are complete in $L^2([0,\infty))$, we see that $K=0$, and the proof is finished.

*Note:* With a similar proof one can show that if $G$ is radially symmetric and satisfies

$$G(a,b) = O((a^2+b^2)^p \exp[-2\pi(a^2+b^2])\in\mathbb{R}^2, \tag{3.11}$$

for some $p\geqslant 0$, and (3.2) is non-negative for all $f\in S$, then $G$ is of the form

$$G(a,b) = \sum_{n<p} (-1)^n a_n \exp[-2\pi(a^2+b^2)] \times L_n[4\pi(a^2+b^2)] \quad [a,b)\in\mathbb{R}^2], \tag{3.12}$$

with $a_n\geqslant 0$ $(n<p)$.

**Theorem 3.1:** Assume that $G:\mathbb{R}^2\to\mathbb{R}$ is continuous and that

$$G(a,b) = O(\exp[-2\pi\delta(a^2+b^2)]) \quad [(a,b)\in\mathbb{R}^2], \tag{3.13}$$

for some $\delta > 1$. If (3.2) is non-negative for all $f\in S$, then $G=0$.

*Proof:* Let $(a_0,b_0)\in\mathbb{R}^2$, and let

$$G_0(a,b) := G(a-a_0, b-b_0) \quad [(a,b)\in\mathbb{R}^2]. \tag{3.14}$$

We see from (1.2) that (3.2) holds for all $f$ (with $G_0$ instead of $G$). Furthermore

$$G_0(a,b) = O(\exp[-2\pi\epsilon(a^2+b^2)]) \quad [(a,b)\in\mathbb{R}^2], \tag{3.15}$$

for any $\epsilon$ between 1 and $\delta$. Now Lemma 3.1 shows that

$$\int_0^{2\pi} G_0(R\cos\theta, R\sin\theta) d\theta = 0 \quad (R\geqslant 0). \tag{3.16}$$

It then follows from continuity of $G$ that $G_0(0,0) = G(a_0,b_0) = 0$. This completes the proof.

*Note:* It is clear that the conditions on $G$ can be weakened somewhat.

**Theorem 3.2:** Let $\gamma > 1$ and let $\delta > \gamma(\gamma-1)^{-1}$. Assume that $\varphi:\mathbb{R}^2\to\mathbb{R}$ satisfies

$$\varphi(q,p) = O(\exp[-2\pi\delta(q^2+p^2)]) \quad [(q,p)\in\mathbb{R}^2]. \tag{3.17}$$

Then there is an $f\in S$ for which (3.1) is negative, unless $\varphi = 0$. In particular, there is no compactly supported $\varphi\neq 0$ such that (3.1) is non-negative for all $f\in S$.

*Proof:* Let $G$ be as in (3.3). Then $G$ is smooth and satisfies

$$G(a,b) = O\left[\exp\left(-2\pi\frac{\delta\gamma}{\delta+\gamma}(a^2+b^2)\right)\right] \quad [(a,b)\in\mathbb{R}^2]. \tag{3.18}$$

As $\delta\gamma/(\delta+\gamma) > 1$, the theorem follows from Theorem 3.1.

*Note:* We can allow $\varphi$ to be an element of $S^{2*}$ if we have a substitute for condition (3.17). The theorem also holds, for instance, when $N_{\alpha,2}\varphi$ (instead of $\varphi$) satisfies (3.17) for some $\alpha > 0$. This is a consequence of (2.9). The theorem also holds when one requires that $\Phi$ be an entire function of two variables with

$$\Phi(\theta,\tau) = O\left[\exp\left(\frac{\pi\epsilon}{2}(|\theta|^2 + |\tau|^2)\right)\right] \quad [(\theta,\tau)\in\mathbb{C}^2] \tag{3.19}$$

for some $\epsilon < (\gamma - 1)/\gamma$, for then the $G$ of (3.3) also satisfies (3.13) with a $\delta > 1$. All these matters can be proved rigorously within the framework of the theory in Ref. 33.

*Example:* Let $\gamma > 0$ and consider the choice $\Phi_0(\theta,\tau) = \exp(2\pi i\alpha\theta\tau)$ with $\alpha\in\mathbb{R}$, $\alpha\neq 0$. Now $\varphi_0$ is given by

$$\varphi_0(q,p) = \alpha^{-1}\exp(-2\pi i\alpha^{-1}qp) \quad [(q,p)\in\mathbb{R}^2], \tag{3.20}$$

and the $G = G_0$ of (3.3) can be shown to equal

$$G_0(a,b) = (1 + 4\gamma^2\alpha^2)^{-1/2} \times\exp\left(-\frac{2\pi\gamma(a^2+b^2)}{1+4\gamma^2\alpha^2} - \frac{8\pi i\alpha\gamma^2 ab}{1+4\gamma^2\alpha^2}\right) \quad [(a,b)\in\mathbb{R}^2]. \tag{3.21}$$

Let $g$ be the Gaussian $2^{1/4}\exp[-\pi(1+i)q^2]$ whose Wigner distribution equals

$$W_g(q,p) = 2\exp(-2\pi[q^2 + (q+p)^2]) \quad [(q,p)\in\mathbb{R}^2]. \tag{3.22}$$

The convolution of $W_g$ and $G_0$ is a function of the form

$$(W_g * G_0)(q,p) = \exp[-\pi P_1(q,p) + \pi i P_2(q,p)] \quad [(q,p)\in\mathbb{R}^2], \tag{3.23}$$

with $P_1$ a positive definite quadratic and $P_2$ a real nonconstant quadratic. Letting $\varphi(q,p) = \mathrm{Re}[\varphi_0(q,p)] = \alpha^{-1}\cos 2\pi\alpha^{-1}qp$, so that $G(a,b) = \mathrm{Re}[G_0(a,b)]$ and $\Phi(\theta,\tau) = \cos 2\pi\alpha\theta\tau$, we get an example of a $\Phi$ such that (3.2) takes negative values for certain $f$'s. This is so since the real part of (3.23) does so. Note that this example works for any $\gamma > 0$ while Theorem 3.1 and (3.21) predict trouble only for $\gamma/(1+4\gamma^2\alpha^2) > 1$.

We consider the case $\gamma = 1$, which has our prime interest, in some more detail. The next theorem shows that a $\varphi$ yielding non-negative averages in (3.1) must be of positive type in a certain weak sense.

**Theorem 3.3:** Assume that $\varphi:\mathbb{R}^2\to\mathbb{R}$ satisfies

$$\varphi(q,p) = O(\exp[\pi\epsilon(q^2+p^2)]) \quad [(q,p)\in\mathbb{R}^2], \tag{3.24}$$

for all $\epsilon > 0$. A necessary condition that (3.1) with $\gamma = 1$ is non-negative for all $f\in S$ is that

$$\int_0^\infty r^n e^{-r}\varphi_{\underline{a}}\left[\left(\frac{r}{\pi}\right)^{1/2}\right]dr \geqslant 0 \quad (n = 0,1,\ldots,\underline{a}\in\mathbb{R}^2). \tag{3.25}$$

Here $\varphi_{\underline{a}}(R)$ is the average of $\varphi$ over the circle of radius $R$ with center $\underline{a}$, i.e.,

$$\varphi_{\underline{a}}(R) = \frac{1}{2\pi}\int_0^{2\pi}\varphi[\underline{a} + R(\cos\theta,\sin\theta)]d\theta \quad (R\geqslant 0). \tag{3.26}$$

*Proof:* Assume that (3.1) is non-negative for all $f\in S$. By (1.2) it is sufficient to consider the case $\underline{a} = \underline{0}$. Insert formula (1.5) into (3.1) and interchange integrals. We get, for all $f\in S$,

$$\int\int\varphi(a,b)\left(\int\int\exp[-2\pi(q^2+p^2)] \times W_f(q-a,p-b)dq\,dp\right)da\,db. \tag{3.27}$$

The expression between the large parentheses equals $\frac{1}{2}|(f,G_1(-a,-b))|^2$, where for all $(a,b)\in\mathbb{R}^2$

$$G_1(-a,-b)(q) = 2^{1/4}\exp[-\pi(q+a)^2 - 2\pi ibq - \pi iab] \quad (q\in\mathbb{R}). \tag{3.28}$$

This follows from the fact that, for all $(a,b)\in\mathbb{R}^2$,

$$W_{G_1(-a,-b)}(q,p) = 2\exp[-2\pi(q+a)^2 - 2\pi(p+b)^2] \quad [(q,p)\in\mathbb{R}^2], \tag{3.29}$$

and Moyal's formula. The choice $f = \psi_n$ gives[35]

$$|(\psi_n,G_1(-a,-b))|^2 = [(a^2+b^2)^n/n!]\exp[-\pi(a^2+b^2)] \quad [(a,b)\in\mathbb{R}^2]. \tag{3.30}$$

Hence

$$\int\int\varphi(a,b)\exp[-\pi(a^2+b^2)](a^2+b^2)^n da\,db = \frac{1}{4\pi^{n+2}}\int_0^\infty r^n e^{-r} \times\left[\int_0^{2\pi}\varphi\left(\sqrt{\frac{r}{\pi}}(\cos\theta,\sin\theta)\right)d\theta\right]dr \geqslant 0, \tag{3.31}$$

for all $n = 0,1,\ldots$, and the theorem follows.

*Note:* Observe that $r^n e^{-r}\sqrt{2\pi n}/n!$ has its maximum for $r = n$ and that this maximum tends to 1 as $n\to\infty$. Also, if $\epsilon > 0$, the set of $r$ with $r^n e^{-r}\sqrt{2\pi n}/n! \geqslant \epsilon$ is an interval around $r = n$ with length of the order $\sqrt{2n\log\epsilon^{-1}}$.

## IV. PHASE-PLANE DISTRIBUTIONS, CORRECT MARGINALS AND MOYAL'S FORMULA

Let $\gamma > 0$. In this section we aim at characterizing all functions $\Phi$ (or $\varphi$) as in (1.1) [(or 1.6)] such that (3.1) holds for all $f\in S$ and such that the corresponding phase-plane distribution functions have correct marginals or satisfy Moyal's formula [see (1.7) and (1.15)]. In the case $\gamma\geqslant 1$ we shall show that, under certain mild conditions on $\Phi$, the situation is very simple: for $\gamma > 1$ no such $\Phi$ exists, for $\gamma = 1$ we must have $\Phi(\theta,\tau) = 1$ (correct marginals) or $\Phi(\theta,\tau) = \exp[-2\pi i(\theta a + \tau b)]$ for some $(a,b)\in\mathbb{R}^2$ (Moyal). And in the case where $\gamma < 1$ and (1.7) is satisfied for all $f\in S$, we are still able to derive certain properties of $\Phi$.

We start with a lemma.

*Lemma 4.1:* Let $H\in L^1(\mathbb{R}^2)\cap L^2(\mathbb{R}^2)$, and assume that

$$\int\int H(q,p)W_f(q,p)dq\,dp \geqslant 0 \quad (f\in S). \tag{4.1}$$

There exists $c_n\geqslant 0$ with $\Sigma_n c_n < \infty$ and orthonormal $f_n\in L^2(\mathbb{R})$ such that

$$H(q,p) = \sum_n c_n W_{f_n}(q,p) \quad [(q,p)\in\mathbb{R}^2], \tag{4.2}$$

with convergence in the $L^2(\mathbb{R}^2)$ sense.

*Proof:* Let $T$ be the linear operator defined for $K\in L^2(\mathbb{R}^2)$ by

$$(TK)(q,p)=\int e^{-2\pi ipt}K\left(q+\frac{1}{2}t,q-\frac{1}{2}t\right)dt \quad [(q,p)\in\mathbb{R}^2]. \tag{4.3}$$

This $T$ maps $L^2(\mathbb{R}^2)$ unitarily onto $L^2(\mathbb{R}^2)$ as can be seen from Moyal's formula.[36] And, letting $(f\otimes\bar{f})(q_1,q_2) = f(q_1)\overline{f(q_2)}$, we have $T(f\otimes\bar{f})=W_f$, for all $f\in S$. Hence, if $T^*$ is the adjoint of $T$,

$$(T^*H,f\otimes\bar{f})\geqslant 0 \quad (f\in S), \tag{4.4}$$

where ( , ) denotes the inner product in $L^2(\mathbb{R}^2)$. Formula (4.4) extends to all $f\in L^2(\mathbb{R})$ since $T^*H\in L^2(\mathbb{R}^2)$ and $S$ is dense in $L^2(\mathbb{R})$. We conclude that $T^*H$ has a representation[37]

$$(T^*H)(q_1,q_2)=\sum_n c_n f_n(q_1)\overline{f_n(q_2)} \quad [(q_1,q_2)\in\mathbb{R}^2], \tag{4.5}$$

with $f_n\in L^2(\mathbb{R})$ orthonormal, $c_n\geqslant 0$, $\Sigma_n c_n{}^2<\infty$ and convergence in the $L^2(\mathbb{R}^2)$-sense. Taking $T$ at both sides of (4.5) we arrive at

$$H(q,p)=\sum_n c_n W_{f_n}(q,p) \quad [(q,p)\in\mathbb{R}^2], \tag{4.6}$$

with convergence in the $L^2(\mathbb{R}^2)$ sense.

We still have to prove that $\Sigma_n c_n<\infty$. To that end we consider $H_1(q,p)=(1/\sqrt{2})H(q/\sqrt{2},p/\sqrt{2})$. We have [see (2.10)]

$$H_1(q,p)=\sum_n c_n V_{f_n}(q,p) \quad [(q,p)\in\mathbb{R}^2]. \tag{4.7}$$

Let $\alpha>0$, and apply to both sides of (4.7) the smoothing operator $N_{\alpha,2}$ (see Sec. I). We get by (2.9)

$$(N_{\alpha,2}H_1)(q,p)=\sum_n c_n V_{N_\alpha f_n}(q,p) \quad [(q,p)\in\mathbb{R}^2], \tag{4.8}$$

with convergence in the $S^2$ sense.[38] If we integrate this identity over all $(q,p)\in\mathbb{R}^2$, we obtain by (1.7)

$$\int\int (N_{\alpha,2}H_1)(q,p)dq\,dp=\sqrt{2}\sum_n c_n\|N_\alpha f_n\|^2, \tag{4.9}$$

where $\|\ \|$ denotes the $L^2(\mathbb{R})$ norm. Now $\|N_\alpha f_n\|$ increases to $\|f_n\|=1$ for all $n$ [see (2.6)], and[39] $N_{\alpha,2}H_1\to H_1$ in the $L^1(\mathbb{R}^2)$ sense if $\alpha\downarrow 0$ since $H\in L^1(\mathbb{R}^2)$, and whence $H_1\in L^1(\mathbb{R}^2)$. We conclude that

$$\sum_n c_n=\int\int H(q,p)dq\,dp<\infty, \tag{4.10}$$

and this completes the proof.

*Note:* Since $\|f_n\|=1$, we have $|W_{f_n}(q,p)|\leqslant 2$ for $(q,p)\in\mathbb{R}^2$. Hence, the convergence of the series in (4.2) is uniform. Since $W_{f_n}$ is continuous for every $n$, we furthermore see that the $H$ of Lemma 4.1 is continuous.

We are now ready to prove the following theorem.

**Theorem 4.1:** Assume that the $G$ of (3.3) is in $L^1(\mathbb{R}^2)\cap L^2(\mathbb{R}^2)$, and that (3.1) holds for all $f\in S$. Then, (a) if $\gamma>1$, $C_f^{(\Phi)}$ cannot have correct marginals for all $f\in S$; and (b) if $\gamma=1$, and $C_f^{(\Phi)}$ has correct marginals for all $f\in S$, then $\Phi=1$, and $C_f^{(\Phi)}$ is the Wigner distribution of $f$ for all $f\in S$.

*Proof:* Assume that $C_f^{(\Phi)}$ has correct marginals for all $f\in S$. This means that

$$\int\varphi(q,p)dp=\delta(q) \quad (q\in\mathbb{R}),$$
$$\int\varphi(q,p)dq=\delta(p) \quad (p\in\mathbb{R}). \tag{4.11}$$

Hence, if we integrate the $G$ of (3.3) over all $b$ and $a$, we get, respectively,

$$\int G(a,b)db=\left(\frac{1}{2\gamma}\right)^{1/2}\exp(-2\pi\gamma a^2) \quad (a\in\mathbb{R}), \tag{4.12}$$

and

$$\int G(a,b)da=\left(\frac{1}{2\gamma}\right)^{1/2}\exp(-2\pi\gamma b^2) \quad (b\in\mathbb{R}). \tag{4.13}$$

Our $G$ satisfies the conditions of Lemma 4.1 and therefore we have the representation (4.2) for $H=G$. With an argument similar to the one used for proving convergence of $\Sigma_n c_n$ in Lemma 4.1 we can show that

$$\sum_n c_n|f_n(a)|^2=\int G(a,b)db \quad (\text{a.e. } a\in\mathbb{R}), \tag{4.14}$$

and

$$\sum_n c_n|(\mathscr{F}f_n(b)|^2=\int G(a,b)da \ (\text{a.e. } b\in\mathbb{R}). \tag{4.15}$$

Since all $c_n\geqslant 0$, we conclude that, for all $n$ by (4.12) and (4.13),

$$c_n^{1/2}|f_n(a)|\leqslant(1/2\gamma)^{1/4}\exp(-\pi\gamma a^2) \quad (\text{a.e. } a\in\mathbb{R}), \tag{4.16}$$

and

$$c_n^{1/2}|(\mathscr{F}f_n)(b)|\leqslant(1/2\gamma)^{1/4}\exp(-\pi\gamma b^2) \quad (\text{a.e. } b\in\mathbb{R}). \tag{4.17}$$

As we shall show in Lemma 4.2, the conditions (4.16) and (4.17) are incompatible when $\gamma>1$ (unless $c_n=0$). This completes the proof for the case $\gamma>1$. When $\gamma=1$, it follows from Lemma 4.2 that every $c_n^{1/2}f_n$ is a multiple of the Gaussian $\exp(-\pi a^2)$. Therefore, $c_n\neq 0$ for only one $n$, and it easily follows that

$$G(a,b)=\exp[-2\pi(a^2+b^2)] \quad [(a,b)\in\mathbb{R}^2]. \tag{4.18}$$

Hence, as $G$ is the convolution of $\varphi$ and $\exp[-2\pi(a^2+b^2)]$, we get $\varphi(q,p)=\delta(q)\delta(p)$. This completes the proof.

*Notes*: (1) Since the $G$ of (3.3) is the double inverse Fourier transform of $(1/2\gamma)\Phi(\theta,\tau)\exp[-(\pi/2\gamma)(\theta^2+\tau^2)]$ it is clear that one should impose certain conditions on smoothness and growth on $\Phi$ to get $G\in L^1(\mathbb{R}^2)\cap L^2(\mathbb{R}^2)$. For instance, conditions of type (1.13) and (1.14) guarantee[40] that $G\in L^1(\mathbb{R}^2)\cap L^2(\mathbb{R}^2)$.

(2) As the proof shows, the theorem can be proved equally well with the Gaussian $\exp[-2\pi\gamma(q^2+p^2)]$ in (3.1) replaced by certain smooth functions $K(q,p)$ with $\int K(q,p)dp=O[\exp(-2\pi\gamma q^2)]$ and $\int K(q,p)dq=O[\exp(-2\pi\gamma p^2)]$.

In the next theorem we replace the condition of having correct marginals by the condition that Moyal's formula holds. We restrict the class of allowed $\varphi$'s a little further since we need some results from Ref. 33 about convolution theory in $S^2$ and $S^{2*}$. Of course, if one chooses a different mathematical setting (e.g., a setting based on Schwartz' theory of tempered distributions), one can still prove a theorem as the one below.

**Theorem 4.2:** Assume that [41]$\Phi(\theta,\tau)\exp[-\pi\epsilon(\theta^2+\tau^2)]\in S^2$ for all $\epsilon>0$, and that (3.1) holds for all $f\in S$. Then, (a) if $\gamma>1$, Moyal's formula (1.15) cannot hold for all $f\in S$, $g\in S$; and (b) if $\gamma=1$ and Moyal's formula holds for all $f\in S$, $g\in S$, then there is an $(a,b)\in\mathbb{R}^2$ with $C_f^{(\Phi)}(q,p) = W_f(q-a,p-b)$ for all $f\in S$ $[(q,p)\in\mathbb{R}^2]$.

*Proof:* Assume that Moyal's formula holds for all $f$ and $g$. Then

$$|\Phi(\theta,\tau)|=1 \quad [(\theta,\tau)\in\mathbb{R}^2]. \tag{4.19}$$

In terms of $\varphi$ this condition can be written as

$$\int\int \varphi(q+a,p+b)\overline{\varphi(a,b)}da\,db = (\varphi*\tilde{\varphi})(q,p)=\delta(q)\delta(p) \quad [(q,p)\in\mathbb{R}^2]. \tag{4.20}$$

Here $\tilde{\varphi}(a,b)=\overline{\varphi(-a,-b)}$ for all $(a,b)\in\mathbb{R}^2$, and $*$ denotes the convolution product for (generalized) functions of two variables.

By the definition of $G$ and the representation (4.2) we have, with $K(q,p)=\exp[-2\pi\gamma(q^2+p^2)]$,

$$\varphi*K=G=\sum_n c_n W_{f_n}. \tag{4.21}$$

It will be demonstrated in Appendix A that $c_n = O(e^{-n\beta})$ for some $\beta>0$, that $f_n\in S$ and that the right-hand series converges in the $S^2$ sense to $\varphi*K\in S^2$. Taking convolution with $\tilde{\varphi}$ at both sides and interchanging the convolution and summation signs at the right-hand side (this is allowed[42]), we get

$$K=\tilde{\varphi}*\varphi*K=\sum_n c_n\tilde{\varphi}*W_{f_n}, \tag{4.22}$$

by (4.20) and (4.21).

We now observe that the Fourier transform of $\tilde{\varphi}$ equals $\overline{\Phi(\theta,\tau)}$. Hence, Moyal's formula is valid with $\Phi$ as well as with $\bar{\Phi}$. Since $C_f^{(\Phi)}=\varphi*W_f$ we have

$$\int\int C_f^{(\bar{\Phi})}(q,p)dq\,dp=\int\int(\tilde{\varphi}*W_f)(q,p)dq\,dp = \overline{\Phi(0,0)}\int\int W_f(q,p)dq\,dp=d\,\|f\|^2, \tag{4.23}$$

where $d=\overline{\Phi(0,0)}$ is a number of modulus 1. Hence, if we integrate identity (4.22) over the phase plane, we get by (4.23)

$$\frac{1}{2\gamma}=\int\int K(q,p)dq\,dp=d\sum_n c_n\|f_n\|^2=d\sum_n c_n. \tag{4.24}$$

We conclude from $c_n\geqslant 0$ (all $n$) and $|d|=1$ that $d=1$.

On the other hand, (4.22) provides an expansion of $K$ in a series of orthogonal functions, and we have by Parseval's formula

$$\frac{1}{4\gamma}=\int\int|K(q,p)|^2dq\,dp=\sum_n c_n^2. \tag{4.25}$$

Now, if we let $d_n=2\gamma c_n$, then $d_n\geqslant 0$, $\Sigma_n d_n=1$, $\Sigma_n d_n^2=\gamma$. This is not possible when $\gamma>1$, whence the case $\gamma>1$ has been dealt with.

We shall give two proofs for the case $\gamma=1$, one directly hereafter, and one in Appendix B. When $\gamma=1$, we see that exactly one $d_n$ equals 1; the others are 0. Hence,

$$\varphi*K=\tfrac{1}{2}W_f \tag{4.26}$$

for some $f\in S$ with $\|f\|=1$. Take the double inverse Fourier transform of (4.26). We get the identity

$$(\mathscr{F}^{(1)*}\mathscr{F}^{(2)*}W_f)(\theta,\tau) = \Phi(\theta,\tau)\exp[-(\pi/2)(\theta^2+\tau^2)] \quad [(\theta,\tau)\in\mathbb{R}^2]. \tag{4.27}$$

The expression at the left-hand side of (4.27) can be written as

$$(\mathscr{F}^{(1)*}\mathscr{F}^{(2)*}W_f)(\theta,\tau)=\int e^{2\pi i\theta q}f\Big(q+\frac{1}{2}\tau\Big)\overline{f\Big(q-\frac{1}{2}\tau\Big)}dq = \mathrm{Amb}_f(-\tau,-\theta) \quad [(\theta,\tau)\in\mathbb{R}^2]; \tag{4.28}$$

here $\mathrm{Amb}_f$ is the ambiguity function of $f$ which is well known in radar analysis.[43,44] From a result of Ref. 44 the following inequality can be derived for ambiguity functions. If $p=1,2,\ldots$, then for any $g$,

$$\int\int|\mathrm{Amb}_g(\tau,\theta)|^{2p}d\tau\,d\theta\leqslant\frac{1}{p}\|g\|^{2p}; \tag{4.29}$$

if $p=2,3,\ldots$, the only functions $g$ that never vanish, that are twice differentiable, and that achieve equality in (4.29) are of the form

$$g(q)=\exp(-\pi\alpha q^2+2\pi\beta q-\pi\epsilon) \quad (q\in\mathbb{R}), \tag{4.30}$$

with arbitrary complex $\alpha$, $\beta$, $\epsilon$, and $\mathrm{Re}\,\alpha>0$.

It is easily verified from the fact that $|\Phi(\theta,\tau)|=1$ and $\|f\|=1$ that $f$ achieves equality in (4.29) for $p=2,3,\ldots$. However, our $f$ is allowed to have zeros. What the argument of the proof in Ref. 44 shows, though, is that if a smooth $g$ achieves equality in (4.29) and $g(q_1)\neq 0$, then $g$ has the special form (4.30) in a neighborhood of $q_1$. And as our $f$ is an entire function, the conclusion that $f$ has the special form (4.30) remains equally valid.

If we calculate $\mathrm{Amb}_g$ for the $g$ of (4.30), we find

$$\begin{aligned}\mathrm{Amb}_g(\theta,\tau) &= (1/2\,\mathrm{Re}\,\alpha)^{1/2}\exp(-2\pi[\mathrm{Re}\,\gamma-(\mathrm{Re}\,\beta)^2/\mathrm{Re}\,\alpha])\\ &\times\exp[-\tfrac{1}{2}\pi\tau^2\,\mathrm{Re}\,\alpha-\tfrac{1}{2}\pi(\omega-\tau\,\mathrm{Im}\,\alpha)^2/\mathrm{Re}\,\alpha\\ &-(2\pi i/\mathrm{Re}\,\alpha)(\omega\,\mathrm{Re}\,\beta+\tau\,\mathrm{Im}\,\beta\bar{\alpha})].\end{aligned} \tag{4.31}$$

It is now easy to check from (4.27) that $|\Phi(\theta,\tau)|=1$ implies that $\alpha=1$, $\beta\in\mathbb{C}$ arbitrary, $\gamma\in\mathbb{C}$ such that $\mathrm{Amb}_f(0,0)=1$. Then $\Phi$ becomes

$$\Phi(\theta,\tau)=\exp[-2\pi i(\tau\,\mathrm{Im}\,\beta+\theta\,\mathrm{Re}\,\beta)] \quad [(\theta,\tau)\in\mathbb{R}^2], \tag{4.32}$$

and

$$\varphi(q,p)=\delta(q+\mathrm{Re}\,\beta)\delta(p+\mathrm{Im}\,\beta) \quad [(q,p)\in\mathbb{R}^2]. \tag{4.33}$$

This completes the proof.

We shall now prove the claim made in connection with (4.16) and (4.17). It is likely that the results of the lemma below for $\gamma\geqslant 1$ are known, but we could not find appropriate references. In addition, we get useful information for the case that $0<\gamma<1$.

*Lemma 4.2:* Let $\gamma > 0$, and assume that $f \in L^2(\mathbb{R})$ satisfies

$$f(q) = O[\exp(-\pi\gamma q^2)] \quad (\text{a.e. } q \in \mathbb{R}),$$
$$(\mathscr{F}f)(p) = O[\exp(-\pi\gamma p^2)] \quad (\text{a.e. } p \in \mathbb{R}). \tag{4.34}$$

Then, (a) if $\gamma > 1$, we have $f = 0$, (b) if $\gamma = 1$, we have $f(q) = c \exp(-\pi q^2)$ for some $c \in \mathbb{C}$, (c) if $0 < \gamma < 1$, we have, with $r = (1+\gamma)^{1/2}(1-\gamma)^{-1/2}$,

$$\sum_{n=0}^{N} |(f,\psi_n)|^2 r^n = O(N) \quad (N = 0,1,...). \tag{4.35}$$

*Proof:* We obtain from Mehler's formula (2.8), with $-iw$ instead of $w$,

$$\left(\frac{2}{1+w^2}\right)^{1/2} \exp\left(-\pi(q^2+p^2)\frac{1-w^2}{1+w^2} - \frac{4\pi i q p w}{1+w^2}\right)$$
$$= \sum_{n=0}^{\infty} (-iw)^n \psi_n(q)\psi_n(p) \quad [(q,p) \in \mathbb{R}^2, |w| < 1]. \tag{4.36}$$

Noting that $\mathscr{F}\psi_n = (-i)^n \psi_n$, multiplying (4.36) by $f(q)\,\overline{(\mathscr{F}f)(p)}$ and integrating the result over the phase plane, we obtain for $|w| < 1$

$$\sum_{n=0}^{\infty} w^n |(f,\psi_n)|^2$$
$$= \left(\frac{2}{1+w^2}\right)^{1/2} \int\int f(q)\,\overline{(\mathscr{F}f)(p)}$$
$$\times \exp\left(-\pi(q^2+p^2)\frac{1-w^2}{1+w^2} - \frac{4\pi i q p w}{1+w^2}\right) dq\, dp. \tag{4.37}$$

We let $w > 0$, we insert the estimates (4.34) in the integral at the right-hand side of (4.37), and we take the modulus. The integral that turns up can be evaluated explicitly, and we obtain

$$\sum_{n=0}^{\infty} w^n |(f,\psi_n)|^2 \leqslant K \frac{(1+w^2)^{1/2}}{\gamma + 1 + (\gamma - 1)w^2} \quad (0 < w < 1), \tag{4.38}$$

for some constant $K \geqslant 0$. The integral in (4.37) thus converges absolutely as long as $\gamma + 1 + (\gamma - 1)w^2 > 0$.

Since the left-hand side of (4.37) is a power series with non-negative coefficients, we see by Pringsheim's theorem[45] that the radius of convergence of the power series is at least equal to $r$ when $0 < \gamma < 1$, and $\infty$ when $\gamma \geqslant 1$. In the first case we have in addition that

$$\limsup_{w \to r-} (r - w) \sum_{n=0}^{\infty} w^n |(f,\psi_n)|^2 < \infty. \tag{4.39}$$

It is not hard to see then that

$$\sum_{n=0}^{N} |(f,\psi_n)|^2 r^n = O(N) \quad (N = 0,1,...). \tag{4.40}$$

In the case $\gamma > 1$ we see that the right-hand side of (4.38) tends to zero when $w \to \infty$. This implies that $(f,\psi_n) = 0$ for all $n$, whence $f = 0$. Finally, if $\gamma = 1$, we see that the right-hand side of (4.38) is $O(|w|)$, $w \to \infty$, whence $(f,\psi_n) \neq 0$ is only possible for $n = 0,1$. Since $\psi_0(q) = 2^{1/4}\exp(-\pi q^2)$, $\psi_1(q) = 2\pi^{1/2} q\, \psi_0(q)$ we see from (4.34) that $(f,\psi_1) = 0$. This completes the proof.

In the remainder of this paper we let $0 < \gamma < 1$. We shall find conditions on the Wigner distributions of the $f_n$'s as in (4.2) and on $G$ that must be satisfied in order that (3.1) is non-negative for any $f$ while $C_f^{(\Phi)}$ has the correct marginals for any $f$. There exist $\Phi \neq 1$ with these two properties, viz. $\Phi(\theta,\tau) = \exp(\pi\delta\theta\tau)$ $[(\theta,\tau) \in \mathbb{R}^2]$ with $\delta = \pm\gamma^{-1}(1-\gamma^2)^{1/2}$. (In fact, this example is not quite proper since $\Phi$ cannot be tested against all elements of $S^2$.) It can be shown that the $G$ of (3.3) equals in this case

$$G(q,p) = W_f(q,p)$$
$$= \frac{1}{\gamma}\exp\left(-\frac{2\pi\gamma(q^2+p^2)}{1+\sqrt{1-\gamma^2}} - \frac{2\pi}{\gamma}\sqrt{1-\gamma^2}(q+p)^2\right)$$
$$[(q,p) \in \mathbb{R}^2], \tag{4.41}$$

where

$$f(q) = (1/2\gamma)^{1/4}\exp(-\pi[\gamma + i(1-\gamma^2)^{1/2}]q^2) \quad (q \in \mathbb{R}). \tag{4.42}$$

Since the collection of all $\Phi$'s with (3.1) non-negative and (1.7) valid for all $f$ is closed under taking convex combinations, it does not seem easy to describe this collection.

The $f$ in (4.42) satisfies

$$|f(q)| = (1/2\gamma)^{1/4}\exp(-\pi\gamma q^2) \quad (q \in \mathbb{R}),$$
$$|(\mathscr{F}f)(p)| = (1/2\gamma)^{1/4}\exp(-\pi\gamma p^2) \quad (p \in \mathbb{R}), \tag{4.43}$$

while its Wigner distribution satisfies

$$W_f(q,p) = O\left[\exp\left(-\frac{2\pi\gamma(q^2+p^2)}{1+\sqrt{1-\gamma^2}}\right)\right] \quad [(q,p) \in \mathbb{R}^2], \tag{4.44}$$

and its Hermite coefficients are given by ($w = \gamma + i\sqrt{1-\gamma^2}$)

$$(f,\psi_k) = 0 \quad \text{or} \quad \frac{\sqrt{2n!}}{(2\gamma)^{1/4}2^n n!}\left(\frac{w-1}{w+1}\right)^n, \tag{4.45}$$

according as $k$ is odd or $k = 2n$ is even. Hence

$$(f,\psi_k) = O\left(\left|\frac{w-1}{w+1}\right|^{k/2}\right) = O\left(\left(\frac{1-\gamma}{1+\gamma}\right)^{k/4}\right).$$

See also Theorem 4.3 below.

To find a condition on the $W_{f_n}$'s and on $G$, we recall from the proof of Theorem 4.1 that $(K(q,p) = \exp[-2\pi\gamma(q^2+p^2)])$

$$G = \varphi * K = \sum_n c_n W_{f_n}, \tag{4.46}$$

with $f_n$ orthonormal, $c_n \geqslant 0$, $\Sigma_n c_n < \infty$ and, for $(q,p) \in \mathbb{R}^2$,

$$\sum_n c_n |f_n(q)|^2 = \left(\frac{1}{2\gamma}\right)^{1/2}\exp(-2\pi\gamma q^2), \tag{4.47}$$

$$\sum_n c_n |(\mathscr{F}f_n)(p)|^2 = \left(\frac{1}{2\gamma}\right)^{1/2}\exp(-2\pi\gamma p^2). \tag{4.48}$$

We shall show that for any $n = 0,1,...$ and for any $\epsilon < \gamma/(1+\sqrt{1-\gamma^2})$,

$$W_{f_n}(q,p) = O(\exp[-2\pi\epsilon(q^2+p^2)])$$
$$[(q,p) \in \mathbb{R}^2]. \tag{4.49}$$

To that end we prove the following theorem.

**Theorem 4.3:** Let $f\in L^2(\mathbb{R})$ and consider the following statements: (a) for all $\delta<\gamma$ we have

$$f(q)=O(e^{-\pi\delta q^2})\quad (q\in\mathbb{R}),$$
$$(\mathscr{F}f)(p)=O(e^{-\pi\delta p^2})\quad (p\in\mathbb{R});\tag{4.50}$$

(b) for all $\delta<\gamma$ we have

$$(f,\psi_n)=O\left[\left(\frac{1-\delta}{1+\delta}\right)^{n/4}\right]\quad (n=0,1,...);\tag{4.51}$$

and (c) for all $\epsilon<\gamma/(1+\sqrt{1-\gamma^2})$ we have

$$W_f(q,p)=O(\exp[-2\pi\epsilon(q^2+p^2)])\quad [(q,p)\in\mathbb{R}^2].\tag{4.52}$$

Then (a)⟹(b), (b)⟺(c).

*Proof:* The implication (a)⟹(b) follows from Lemma 4.2 (c); in fact the result proved there is slightly more precise. We shall now show that (b)⟹(c). To that end we assume that (b) holds and we let $0<\delta<\gamma$. We can write $f=N_\alpha g$, where $\alpha=\frac{1}{4}\log(1+\delta)(1-\delta)^{-1}$ and where the Hermite coefficients of $g$ equal

$$(g,\psi_n)=\left(\frac{1+\delta}{1-\delta}\right)^{n/4+1/8}(f,\psi_n)\quad (n=0,1,...).\tag{4.53}$$

Hence $g\in S$. Now, by (2.9) and (2.10),

$$W_f(q,p)=W_{N_\alpha g}(q,p)$$
$$=\sqrt{2}(N_{\alpha,2}V_g)(q\sqrt{2},p\sqrt{2})\quad [(q,p)\in\mathbb{R}^2].\tag{4.54}$$

The kernel $K_{\alpha,2}$ of the smoothing operator $N_{\alpha,2}$ can be written as

$$K_{\alpha,2}(q,p;x,y)=\frac{1}{\sinh\alpha}\exp[-\pi(q^2+p^2)\tanh\alpha]$$
$$\times\exp(-\pi[(q-x/\cosh\alpha)^2$$
$$+(p-y/\cosh\alpha)^2]\coth\alpha)$$
$$[(q,p;x,y)\in\mathbb{R}^2\times\mathbb{R}^2].\tag{4.55}$$

Since $V_g\in S^2$ we easily obtain that

$$W_f(q,p)=O(\exp[-2\pi(q^2+p^2)\tanh\alpha])[(q,p)\in\mathbb{R}^2].\tag{4.56}$$

And as

$$\tanh\alpha=\frac{e^{2\alpha}-1}{e^{2\alpha}+1}=\frac{\delta}{1+\sqrt{1-\delta^2}},\tag{4.57}$$

the proof of (b)⟹(c) is complete.

We next show the converse (c)⟹(b), and therefore we assume that (c) holds. It follows that for $0<\epsilon<\gamma/(1+\sqrt{1+\gamma^2})$, the integral

$$\int\int\exp[2\pi\epsilon(q^2+p^2)]\,W_f(q,p)dq\,dp\tag{4.58}$$

converges absolutely. Now let, for $A\geqslant 0$,

$$K(r)=e^{\epsilon r}\quad (r\geqslant 0),$$
$$K_A(r)=\max(K(r),A)\quad (r\geqslant 0).\tag{4.59}$$

Then we have by (2.15) (see Ref. 46), for $A\geqslant 0$,

$$\int\int K_A[2\pi(q^2+p^2)]\,W_f(q,p)dq\,dp$$
$$=\sum_{n=0}^{\infty}(-1)^n|(f,\psi_n)|^2\int_0^\infty e^{-r}K_A(r)L_n(2r)dr.\tag{4.60}$$

Since $K_A$ is nondecreasing we can apply Theorem 2.1, and we find that

$$c_n(A):=(-1)^n\int_0^\infty e^{-r}K_A(r)L_n(2r)dr\geqslant 0$$
$$(A\geqslant 0,n=0,1,...).\tag{4.61}$$

Also, by the generating function of the Laguerre polynomials,

$$\lim_{A\to\infty}c_n(A)=(-1)^n\int_0^\infty e^{-r}K(r)L_n(2r)dr$$
$$=(1-\epsilon)^{n-1}/(1+\epsilon)^n\quad (n=0,1,...).\tag{4.62}$$

Since the left-hand side of (4.60) tends to the finite number in (4.58) as $A\to\infty$, we easily conclude that

$$\sum_{n=0}^{\infty}\frac{(1-\epsilon)^{n-1}}{(1+\epsilon)^n}|(f,\psi_n)|^2<\infty.\tag{4.63}$$

The proof is completed by noting that $(1-\epsilon)^{1/2}\times(1+\epsilon)^{-1/2}=(1-\delta)^{1/4}(1+\delta)^{-1/4}$ when $\epsilon=\delta/(1+\sqrt{1-\delta^2})$.

*Note:* Assume that $f$ satisfies (c). Then it follows from (1.7) that (a) is satisfied with $\gamma$ replaced by $\gamma/(1+\sqrt{1-\gamma^2})$. The implication (a)⟹(b) cannot be strengthened [see (4.43)–(4.45)].

We conclude this paper with the following theorem.

**Theorem 4.4:** Let $G$ be as in (4.46). Then we have

$$G(q,p)=O(\exp[-2\pi\epsilon(q^2+p^2)])\quad [(q,p)\in\mathbb{R}^2]\tag{4.64}$$

for all $\epsilon<\gamma/(1+\sqrt{1-\gamma^2})$.

*Proof:* The proof follows rather closely the proof of the statements (a)⟹(b), (b)⟹(c) in Theorem 4.3. Therefore we shall omit details.

Let $\widetilde{G}(q,p):=(1/\sqrt{2})G(q/\sqrt{2},p/\sqrt{2})$, and define

$$W_{f,g}(q,p)$$
$$=\int e^{-2\pi ipt}f\left(q+\frac{1}{2}t\right)\overline{g\left(q-\frac{1}{2}t\right)}dt\quad [(q,p)\in\mathbb{R}^2],\tag{4.65}$$

$$V_{f,g}(q,p)=\frac{1}{\sqrt{2}}W_{f,g}\left(\frac{q}{\sqrt{2}},\frac{p}{\sqrt{2}}\right)\quad [(q,p)\in\mathbb{R}^2]\tag{4.66}$$

for $f\in S$, $g\in S$. Then we have $(G,W_{f,g})=(\widetilde{G},V_{f,g})$ for all $f\in S$, $g\in S$.

We shall estimate the Hermite coefficients of $\widetilde{G}$. We have

$$(\widetilde{G},\psi_k\otimes\psi_l)=\sum_{i,j}(\widetilde{G},V_{\psi_i,\psi_j})\gamma_{ij,kl},\tag{4.67}$$

with

$$\gamma_{ij,kl}=(V_{\psi_i,\psi_j},\psi_k\otimes\psi_l).\tag{4.68}$$

This follows from completeness and orthonormality of $(V_{\psi_i,\psi_j})_{i,j}$ in $L^2(\mathbb{R}^2)$ (see also the proof of Lemma 4.1). According to Ref. 8, 27.26.1, $\gamma_{ij,kl}$ equals $a_i^{-1}a_j^{-1}a_ka_l$ times the coefficient of $w^iz^j$ in $[(w+z)/\sqrt{2}]^k[(w-z)/i\sqrt{2}]^l$; here $a_n=(n!)^{-1/2}2^{-1/4}(4\pi)^{n/2}$. It is important to observe that $\gamma_{ij,kl}=0$, when $k+l\neq i+j$.

It is easy to see that, for all $i,j$,

$$|(\widetilde{G},V_{\psi_i,\psi_j})|^2 = |(G,W_{\psi_i,\psi_j})|^2 \leqslant (G,W_{\psi_i})(G,W_{\psi_j}), \quad (4.69)$$

whence, as $\Sigma_{i,j}|\gamma_{ij,kl}|^2 = \|\psi_k \otimes \psi_l\|^2 = 1$,

$$|(\widetilde{G},\psi_k \otimes \psi_l)|^2 \leqslant \sum_{i+j=k+l} (G,W_{\psi_i})(G,W_{\psi_j}), \quad (4.70)$$

by the Cauchy–Schwarz inequality.

To estimate $(G,W_{\psi_i})$, we consider $\Sigma_{k=0}^{\infty} w^k (G,W_{\psi_k}) =: F(w)$ for $|w|<1$. We have, as in the proof of Lemma 4.1, for $|w|<1$,

$$F(w) = \left(\frac{2}{1+w^2}\right)^{1/2} \int\int H(q,p) \times \exp\left(-\pi(q^2+p^2)\frac{1-w^2}{1+w^2} - \frac{4\pi i q p w}{1+w^2}\right) dq\, dp, \quad (4.71)$$

with

$$H(q,p) = \sum_n c_n f_n(q)\,\overline{(\mathscr{F} f_n)(p)} \quad [(q,p)\in\mathbb{R}^2]. \quad (4.72)$$

It follows easily from the Cauchy–Schwarz inequality and (4.43) and (4.44) that

$$|H(q,p)| \leqslant \left(\frac{1}{2\gamma}\right)^{1/2} \exp[-\pi\gamma(q^2+p^2)] \quad [(q,p)\in\mathbb{R}^2]. \quad (4.73)$$

As in the proof of Lemma 4.1(c) we conclude that

$$\sum_{k=0}^{N} r^k (G,W_{\psi_k}) = O(N) \quad (N=0,1,...), \quad (4.74)$$

where $r=(1+\gamma)^{1/2}(1-\gamma)^{-1/2}$. Hence $(G,W_{\psi_k}) = O([(1-\delta)/(1+\delta)]^{k/2})$ for all $\delta<\gamma$, and we obtain by (4.70), for all $\delta<\gamma$,

$$(\widetilde{G},\psi_k \otimes \psi_l) = O\left[\left(\frac{1-\delta}{1+\delta}\right)^{(k+l)/4}\right] \quad (k,l=0,1,...). (4.75)$$

This shows that for any $\alpha < \frac{1}{4}\log[(1+\gamma)/(1-\gamma)]$ there is an $F\in S^2$ such that $\widetilde{G} = N_{\alpha,2}F$. As in the proof of the statement (b)⇒(c) in Theorem 4.3 we conclude that, for any $\alpha<\frac{1}{4}\log[(1+\gamma)/(1-\gamma)]$,

$$\widetilde{G}(q,p) = O(\exp[-\pi(q^2+p^2)\tanh\alpha]) \quad [(q,p)\in\mathbb{R}^2], \quad (4.76)$$

and the proof is easily completed now.


## ACKNOWLEDGMENTS

The author thanks T. A. C. M. Claasen and W. F. G. Mecklenbräuker, who anticipated some of the results of this paper, for asking the right questions.


## APPENDIX A: SMOOTH POSITIVE DEFINITE FUNCTIONS OF TWO VARIABLES

In the proof of Theorem 4.2 the following theorem was required.

**Theorem A.1:** Let $K\in S^2$ be positive definite, i.e., $(K,f\otimes\bar{f})\geqslant 0$ for all $f\in L^2(\mathbb{R})$. There are non-negative numbers $c_n$ and orthonormal $f_n\in S$ such that

$$K(q,p) = \sum_n c_n f_n(q)\,\overline{f_n(p)} \quad [(q,p)\in\mathbb{R}^2], \quad (A1)$$

with convergence in the $S^2$ sense. Moreover, when the $c_n$'s are ordered decreasingly we have $c_n = O(e^{-n\epsilon})$ for some $\epsilon>0$.

The proof of this theorem relies on the following lemma.

*Lemma A.1:* Let $K_n\in S^2 (n=0,1,...)$. Then $K_n\to 0$ in the $S^2$ sense if and only if $(K_n,F\otimes\bar{F})\to 0$ for every $F\in S^*$.

*Proof:* It is known[47] that $K_n\to 0$ in the $S^2$ sense if and only if $(K_n,H)\to 0$ for all $H\in S^{2*}$. Hence we only have to show that $(K_n,F\otimes\bar{F})\to 0$ for every $F\in S^*$ implies that $(K_n,H)\to 0$ for every $H\in S^{2*}$.

By polarization we can assume that $(K_n,F\otimes G)\to 0$ for every $F\in S^*$, $G\in S^*$. Let $F\in S^*$. The space $S^*$ is a Fréchet space[48]; as a countable system of norms on $S^*$ we can take, for $m=1,2,...,$

$$\|G\|_m = \left(\sum_{k=0}^{\infty} |(G,\psi_k)|^2 e^{-2k/m}\right)^{1/2} \quad (G\in S^*). \quad (A2)$$

Therefore we can find, by boundedness of $(K_n,F\otimes G)$ $(n=0,1,...)$ for every $G\in S^*$, an $m=1,2,...$ and an $M>0$ such that

$$|(K_n,F\otimes G)| \leqslant M \quad (n=0,1,...) \quad (A3)$$

for all $G\in S^*$ with $\|G\|_m\leqslant 1$. Hence, $S^* = \cup_{l=1}^{\infty} B_l$, where

$$B_l = \{F\in S^* \mid \|G\|_l\leqslant 1 \Rightarrow |(K_n,F\otimes G)|\leqslant l\ (n=0,1,...)\}, \quad (A4)$$

for $l=1,2,...$. Again using that $S^*$ is a Fréchet space we conclude that there is an $l_0=1,2,...$ and an open set in $S^*$ in which $B_{l_0}$ is dense. From this we infer the existence of $M>0$, $k_0=1,2,...$ with

$$|(K_n,F\otimes G)| \leqslant M \quad (n=0,1,...), \quad (A5)$$

for all $F\in S^*$, $G\in S^*$ with $\|F\|_{k_0}\leqslant 1$, $\|G\|_{l_0}\leqslant 1$. If we take $F=\exp(k/k_0)\psi_k$, $G=\exp(l/l_0)\psi_l$, we get

$$|(K_n,\psi_k\otimes\psi_l)| \leqslant M\exp(-k/k_0 - l/l_0) \quad (n,k,l=0,1,...). \quad (A6)$$

It is now easy to show [as $(K_n,\psi_k\otimes\psi_l)\to 0$ for all $k,l$] that $(K_n,H) = \Sigma_{k,l}(K_n,\psi_k\otimes\psi_l)(\psi_k\otimes\psi_l,H)\to 0$ for every $H\in S^{2*}$.

*Corollary:* With an entirely similar proof one can show that if $K_n\in S^2$ and $\lim (K_n,F\otimes\bar{F})$ exists for all $F\in S^*$ then there is exactly one $K\in S^2$ with $K_n\to K$ in the $S^2$ sense.

We now prove Theorem A.1. We have the representation[49]

$$K = \sum_n c_n f_n \otimes \bar{f}_n, \quad (A7)$$

where $c_n\geqslant 0$, $\Sigma_n c_n^2<\infty$, $f_n\in L^2(\mathbb{R})$ orthonormal and where the convergence is in the $L^2(\mathbb{R}^2)$ sense. In addition, for every $n$,

$$c_n f_n(u) = \int K(u,v) f_n(v)\, dv \quad (u\in\mathbb{R}), \quad (A8)$$

and from this one readily concludes that $f_n\in S$, e.g., by expanding $K$ in a Hermite series $\Sigma_{k,l} d_{kl}\,\psi_k\otimes\psi_l$ with $d_{kl} = O(\exp[-\epsilon(k+l)])$ for some $\epsilon>0$. We assume here and in the remainder that $c_n>0$.

Now let $F\in S^*$. We shall check that $\Sigma_n c_n |(f_n,F)|^2<\infty$. To that end we take a sequence $F_k$ in $S$ with $F_k\to F$ in the $S^*$ sense if $k\to\infty$. We have, for all $k$,

$$(K,F_k\otimes\bar{F}_k) = \sum_n c_n |(f_n,F_k)|^2, \quad (A9)$$

by (A6). The terms in the right-hand side series are nonnegative for all $k$ and tend to $c_n |(f_n,F)|^2$ when $k\to\infty$. The left-hand side tends to $(K,F\otimes\bar{F})$, when $k\to\infty$. By Fatou's lemma we conclude that $\Sigma_n c_n |(f_n,F)|^2<\infty$. That is, we have shown that $\lim_{N\to\infty} (\Sigma_{n=0}^N c_n f_n \otimes \bar{f}_n, F\otimes\bar{F})$ exists for all $F\in S^*$. The corollary after Lemma A.1 implies that $\Sigma_{n=0}^N c_n f_n \otimes \bar{f}_n$ converges in the $S^2$ sense. Because of (A7) the limit is $K$, whence $K=\Sigma_n c_n f_n\otimes\bar{f}_n$ with convergence in the $S^2$ sense.

We finally show that $c_n = 0(e^{-n\epsilon})$ for some $\epsilon>0$. It is assumed here that $c_n\geqslant c_{n+1}>0$ (all $n$). We have

$$(K,\psi_k\otimes\psi_k)=\sum_n c_n |(f_n,\psi_k)|^2 = O(e^{-2k\epsilon}) \tag{A10}$$

for some $\epsilon>0$. Hence there is an $M>0$ such that, for all $n$,

$$c_n \sum_k |(f_n,\psi_k)|^2 e^{k\epsilon}\leqslant M. \tag{A11}$$

It follows from orthonormality of the $f_n$'s and Parseval's theorem that for any $m=1,2,\ldots$ there is an $n=n(m)=0,1,\ldots,m+1$ such that

$$\sum_{k=m+1}^{\infty} |(f_n,\psi_k)|^2 \geqslant \frac{1}{m+2}. \tag{A12}$$

Therefore, $c_n(m)\leqslant M(m+2)e^{-(m+1)\epsilon}$.

We have assumed that $c_n>0$ for all $n$, and therefore $n(m)\to\infty$ as $m\to\infty$. Now let $n=1,2,\ldots$, and take an $m$ with $n(m)\leqslant n \leqslant n(m+1)$. Then $m\geqslant n-2$, and, by monotonicity of the $c_n$'s,

$$c_n\leqslant c_{n(m)}\leqslant M(m+2)e^{-(m+1)\epsilon}\leqslant Mne^{-(n-1)\epsilon}, \tag{A13}$$

when $n$ is sufficiently large. This completes the proof of Theorem A.1.

## APPENDIX B: SECOND PROOF OF THEOREM 4.2 (b)

We start from the formula $\varphi * K = \frac{1}{2} W_f$ in (4.26), where $\varphi$ satisfies $\varphi * \tilde{\varphi} = \delta\otimes\delta$, $K(q,p)=\exp[-2\pi(q^2+p^2)]$, and $f\in S$, $\|f\|=1$. This formula can also be written as $\varphi * W_g = W_f$, where $g(q)=2^{1/4}\exp(-\pi q^2)$.

We shall use the following result[50]: when $\varphi_0$ and $\psi_0$ are entire functions, then $(z=x+iy)$

$$2\int\int_{\mathbb{C}} |\varphi_0(z)\psi_0(z)|^2 \quad \exp(-2\pi|z|^2)dx\,dy \leqslant \int\int_{\mathbb{C}} |\varphi_0(z)|^2\exp(-\pi|z|^2)dx\,dy \times \int\int_{\mathbb{C}} |\psi_0(z)|^2 \exp(-\pi|z|^2)dx\,dy, \tag{B1}$$

and, if the right-hand side is finite, there is equality in (B1) if and only if $\varphi_0(z)\psi_0(z)$ can be expressed as $C\exp(2\pi\bar{u}z)$ for some $u\in\mathbb{C}$ and some $C\in\mathbb{C}$. We apply this result with $\varphi_0=\psi_0=Bf$ where $Bf$ is the Bargmann transform[51] of $f$, given by

$$(Bf)(z)=e^{(1/2)\pi z^2}(f*g)(z) = 2^{1/4}\int e^{(1/2)\pi z^2-\pi(z-q)^2} f(q)dq \ (z\in\mathbb{C}). \tag{B2}$$

The Bargmann transform provides an isometry between the spaces $L^2(\mathbb{R},dq)$ and $L^2[\mathbb{C},\exp(-\pi|z|^2)dx\,dy)]$. Hence, the right-hand side of (B1) equals 1, as $\|f\|=1$. We shall show that the left-hand side of (B1) equals 1 as well, so that $[(Bf)(z)]^2$ has the special form as indicated above.

According to Ref. 23, Eq. (2.8), we have $(z=x+iy)$

$$(Bf)(z)\exp(-\tfrac{1}{2}\pi|z|^2)=(f,G_1(x,-y)), \tag{B3}$$

where, for $(a,b)\in\mathbb{R}^2$,

$$G_1(a,b)(q) = 2^{1/4}\exp[-\pi(q-a)^2+2\pi ibq-\pi iab] \quad (q\in\mathbb{R}). \tag{B4}$$

Hence, the left-hand side of (B1) can be brought into the form

$$2\int\int |(f,G_1(x,y))|^4\,dx\,dy. \tag{B5}$$

By Moyal's formula we have

$$|(f,G_1(x,y))|^2 = 2\int\int W_f(a,b)\exp[-2\pi(x-a)^2 - 2\pi(y-b)^2]da\,db = (W_f * W_g)(x,y). \tag{B6}$$

Hence, the left-hand side of (B1) can be written as

$$2(W_f*W_g, W_f*W_g). \tag{B7}$$

Now $W_f=\varphi * W_g$, and $(\varphi * H_1,\varphi * H_2) = (\varphi*\tilde{\varphi}*H_1,H_2)=(H_1,H_2)$ for any $H_1\in S^2$, $H_2\in S^2$. Hence, the left-hand side of (B1) equals $2(W_g*W_g,\ W_g*W_g)$. Using that

$$(W_g*W_g)(a,b)=\exp[(-\pi(a^2+b^2)] \quad [(a,b)\in\mathbb{R}^2], \tag{B8}$$

we see that the left-hand side of (B1) equals 1.

This shows that there is equality in (B1), whence $(Bf)(z)$ is of the form $C\exp(2\pi\bar{u}z)$ for some $C\in\mathbb{C}$ and some $u\in\mathbb{C}$. Writing $\bar{u}=a+ib$, we see from Ref. 23, Eq. (2.8), that $f$ is a multiple of $G_1(a,b)$. And since $\|f\|=G_1(a,b)=1$, we get

$$W_f(q,p)=2\exp[-2\pi(q-a)^2-2\pi(p-b)^2] \quad [(a,b)\in\mathbb{R}^2]. \tag{B9}$$

Finally the formula $\varphi * W_f = W_g$ shows that $\varphi(q,p)=\delta(q+a)\delta(p+b)$. This completes the proof.


[1]L. Cohen, J. Math. Phys. **7**, 781 (1966).
[2]E. Wigner, Phys. Rev. **40**, 749 (1932).
[3]T. A. C. M. Claasen and W. F. G. Mecklenbräuker, Philips J. Res. **35**, 372 (1980).
[4]H. Margenau and L. Cohen, in *Probabilities in Quantum Mechanics, Quantum Theory and Reality*, edited by M. Bunge (Springer-Verlag, Berlin, 1967), Chap. 4.
[5]R. P. Boas, *Entire Functions* (Academic, New York, 1954).
[6]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[7]H. J. Groenewold, Physica The Hague **12**, 405 (1946).
[8]N. G. de Bruijn, Nieuw Arch. Wiskunde **21**, 205 (1973).
[9]A. J. E. M. Janssen, Philips J. Res. **37**, 79 (1982).
[10]H. Margenau and R. N. Hill, Progr. Theoret. Phys. (Kyoto) **26**, 722 (1961).
[11]It is not hard to check that $C_f^{(\varphi)}(q,p)\in\mathbb{R}$ for all $(q,p)\in\mathbb{R}^2$ and all states $f$ if and only if the $\varphi$ of (1.6) is real.
[12]E. P. Wigner, in *Quantum-Mechanical Distribution Functions Revisited,*

*Perspectives in Quantum Theory*, edited by W. Yourgrau and A. van der Merwe (Dover, New York, 1979), Chap. 4.
[13]L. Cohen and Y. I. Zaparovanny, J. Math. Phys. **21**, 794 (1980).
[14]K. Husimi, Proc. Phys. Math. Soc. Jpn. **22**, 264 (1940).
[15]N. G. de Bruijn, in *Uncertainty Principles in Fourier Analysis, Inequalities*, edited by O. Shisha (Academic, New York, 1967).
[16]A. J. E. M. Janssen, SIAM J. Math. Anal. **12**, 752 (1981).
[17]R. L. Hudson, Rep. Math. Phys. **6**, 249 (1974).
[18]F. Soto and P. Claverie, Physica A **109**, 193 (1981).
[19]W. M. de Muynck, P. A. E. M. Janssen, and A. Santman, Found Phys. **9**, 71 (1979).
[20]N. D. Cartwright, Physica A **83**, 210 (1976).
[21]A. J. E. M. Janssen, SIAM J. Math. Anal. **15**, 170 (1984).
[22]This condition is satisfied, e.g., when $C_f^{(\Phi)}$ has the finite support properties (1.11) and (1.12).
[23]A. J. E. M. Janssen, J. Math. Phys. **23**, 720 (1982).
[24]See Ref. 8, Sec. 10.
[25]See Ref. 8, Sec. 25.
[26]See Ref. 8, Sec. 27.12.2.
[27]See Ref. 16, Sec. 2.2; see also Ref. 7 and J. Peetre, Matematiche (Catania) **27**, 301 (1972).
[28]See G. Szegö, *Orthogonal Polynomials*, (AMS, Providence, RI, 1975), 4th ed., Vol. 23.
[29]See Ref. 16, Sec. 2.3.
[30]T. M. Apostol, *Mathematical Analysis* (Addison-Wesley, Reading, MA, 1974), 2nd ed., p. 165.
[31]See Ref. 28, 8.4.
[32]See Ref. 8, Sec. 26.
[33]A. J. E. M. Janssen, Proc. K. N. A. W., Ser. A **82**, 283 (1979).
[34]See Ref. 28, Chap. VI, Sec. 8.
[35]See Ref. 23, Eqs. (2.6) and (2.8).
[36]J. C. T. Pool, J. Math. Phys. **7**, 66 (1966).
[37]A. C. Zaanen, *Linear Analysis* (North-Holland, Amsterdam, 1953).
[38]See Ref. 8, Theorem 23.2.
[39]A. J. E. M. Janssen, "Application of the Wigner Distribution to Harmonic Analysis of Generalized Stochastic Processes," MC-tract 114, Mathematisch Centrum, Amsterdam, 1979, Appendix 1, 1.6.
[40]See Ref. 5. By Hartogs' theorem a $\Phi$ satisfying (1.13) and (1.14) is an entire function in both variables, see, e.g., V. S. Vladimirov, in *Methods of the Theory of Functions of Many Complex Variables*, edited by L. Ehrenpreis (M. I. T., Cambridge, MA, 1966).
[41]This condition means that $\varphi \in \mathscr{C}^2$, the class of all elements of $S^{2*}$ for which $\varphi * f \in S^2$ for all $f \in S^2$ (see Ref. 33). For such a $\varphi$ the convolution $\varphi * F$ with $F \in S^{2*}$ makes sense and yields an element of $S^{2*}$.
[42]The convolution mapping $f \in S^2 \to \varphi * f \in S^2$ is continuous. See Ref. 33, Theorem 2.3 (i).
[43]P. M. Woodward, *Probability and Information Theory, with Applications to Radar* (McGraw-Hill, New York, 1953), p. 120.
[44]R. Price and E. M. Hofstetter, IEEE Trans. Inf. Theory **IT-11**, 207 (1965).
[45]E. C. Titchmarsh, *The Theory of Functions* (Oxford U. P., New York, 1939), 2nd ed.
[46]We cannot use (2.15) directly for $K$ since (2.15) was derived in Ref. 16 for $f \in S$ and measurable $K:[0,\infty) \to \mathbf{R}$ with $K(r) = O[\exp(\epsilon r)]$ for all $\epsilon > 0$.
[47]See Ref. 39, Appendix 1, Sec. 3.
[48]See Ref. 39, Appendix 1, Sec. 2.
[49]See Ref. 37.
[50]S. Saitoh, Proc. Am. Math. Soc. **80**, 254 (1980), Theorem 1.1, with $\varphi_0(z) = \varphi(z\sqrt{\pi})$, $\psi_0(z) = \psi(z\sqrt{\pi})$. The inequality given in this theorem is to some extent equivalent to the inequality in Ref. 44 which was used in the first proof of Theorem 4.2(b) of our paper.
[51]See Ref. 23.

# A quantum-mechanical theory of distant correlations

Milan Vujičić and Fedor Herbut
*Department of Physics, Faculty of Science, 11001 Belgrade, Yugoslavia*

(Received 16 November 1983; accepted for publication 9 December 1983)

A composite quantum system consisting of two distant subsystems and described by a correlated state vector $\phi_{12}$ is considered. It was shown in a previous work by the authors [Ann. Phys. **96**, 382 (1976)] that such a system can be equivalently described in terms of the reduced statistical operators $\rho_1$ and $\rho_2$ of $\phi_{12}$ applying to the subsystems and a correlation operator $U_a$ between them. It is argued that this description has a firm physical foundation for the system considered in view of the fact that, on account of the subsystems being distant, one can only measure pairs of subsystem observables $A_1$, $B_2$ in coincidence. The direct measurement of $A_1$ such that $[A_1,\rho_1] = 0$ on the ensemble of first subsystems performs distantly (without interaction) an orthogonal decomposition of the ensemble of second subsystems $\rho_2$, that amounts to the measurement of the twin observable $A_2 (A_2 \equiv U_a A_1 U_a^{-1} Q_2$, $Q_2$ being the range projector of $\rho_2$). A number of coincidence experiments have confirmed this claim, and have disproved all attempts (on the quantum and on the subquantum levels) to view this decomposition of $\rho_2$ as being present also before the measurement of $A_1$. Hence, this decomposition into subensembles comes about in the very measurement of $A_1$, and $U_a$ determines them in a simple way. It is demonstrated that $U_a$ is essential for twin observables and twin symmetry operators. A detailed study of these operators is presented from a unified point of view. Puzzling features of quantum correlations described by $U_a$ show up in composite states when the mentioned distant decompositions of $\rho_2$ into subensembles can be incompatible with one another. A general definition of such $\phi_{12}^{\mathrm{EPR}}$ states (called Einstein–Podolsky–Rosen states) is given in a few equivalent forms, and the nonuniqueness of the Schmidt canonical form of $\phi_{12}^{\mathrm{EPR}}$ is investigated in order to encourage further theoretical and experimental exploration of distant quantum correlations.

PACS numbers: 03.65.Bz, 03.65.Ca

## I. INTRODUCTION

To begin with, we try to give an answer to the question: What is intuitively paradoxical about distant correlations in quantum mechanics?

To this purpose, we are considering a quantum system consisting of two subsystems, that is described by a wave vector $\phi_{12}$. We have shown[1] that, within the framework of quantum mechanics, this system can be equivalently described in terms of the separate states of the two subsystems (the reduced statistical operators $\rho_1$ and $\rho_2$) and the quantum correlations between them (the antiunitary correlation operator $U_a$ mapping the range of $\rho_1$ onto that of $\rho_2$): $\phi_{12} \leftrightarrow \{\rho_1, U_a, \rho_2\}$. Inspired by Schrödinger,[2] we made[1] a systematic investigation of the nature and physical implications of the correlations established by $U_a$ by studying distant measurement of subsystem observables (that are complete and have a purely discrete spectrum).

Let us restrict ourselves, for the sake of an illustration, to the two-photon system used in the Freedman and Clauser experiment.[3] When one finds out by measurement that the first photon is in the state of polarization $\varphi_1$, then the second photon is necessarily in the state of polarization $U_a\varphi_1$. From the quantum-mechanical point of view, $\phi_{12}$ collapses (without any interaction with the second photon) into $\varphi_1 \otimes (U_a\varphi_1)$. This quantum-mechanical prediction was confirmed by direct polarization measurement on the second photon in the Freedman and Clauser experiment. Thus, it is experimentally verified that the correlation operator $U_a$ determines the state of the distant photon (after the measurement on the first one).

More generally, measurement of any observable $A_1$ on the first photon implies[1] the distant measurement of the twin observable $A_2 \equiv U_a A_1 U_a^{-1}$ on the second photon. Moreover, if one considers two incompatible observables $A_1$ and $B_1$ on the first photon (e.g., linear polarizations through two different planes), the corresponding twin observables $A_2$ and $B_2$ on the second photon are also incompatible because the above similarity transformation by $U_a$ preserves commutators. It means that one can distantly, hence without disturbance, measure any of the two incompatible observables on the second photon. Hence, one may conclude that the second photon "knows the answer"[2] to both measurements, suggesting incompleteness of the quantum-mechanical description by $\phi_{12}$. This is the essence of the famous Einstein–Podolsky–Rosen (EPR) paradox.[4]

In a correct statistical language, the direct measurement on the ensemble $\rho_1$ of first photons singles out distantly the subensemble $(U_a\varphi_1)$ from the ensemble $\rho_2$ of second photons present before the measurement. Asking the question what is actually happening with the ensemble of second photons in this change, one takes the position of physical realism. There are two possible answers. Either (a) the change is taking place in reality (under distant influence without interaction of any type that we know today), or (b) the change is only in our knowledge, so that the second photons were in the same quantum-mechanical subensemble $(U_a\varphi_1)$ also before the meausrement on the first photons.

It should be noted that from the point of view of the Copenhagen school of thought, the question of the realistic meaning of the collapse $\phi_{12} \rightarrow \varphi_1 \otimes (U_a\varphi_1)$ is not physical.

 0022-2488/84/072253-07$02.50  

Contrarily, Einstein, Schrödinger, and others did consider this question physical, but they could not accept alternative (a).

As far as alternative (b) is concerned, both Einstein and Schrödinger had their visions of it. Schrödinger's hypothesis[2] was in terms of quantum-mechanical entities: He envisaged that $\phi_{12}$ goes over spontaneously into a mixed state $\rho_{12}$, where the phases in the coherent mixture $\phi_{12}$ disappear when the two particles get sufficiently apart so that they are out of the range of mutual interaction. In this mixed state quasi-classical statistical correlations appear, and this type of correlation is intuitively easy to grasp. The mixed state $\rho_{12}$ gives some predictions that are different from those implied by $\phi_{12}$, hence experiment could decide. The Schrödinger hypothesis was experimentally refuted[5,6] (cf. Ref. 7, p. 1922; also Ref. 8).

In the Bell model[9] (inspired by Einstein) the existence of quasiclassical statistical correlations was assumed on a subquantum level (the so-called model of local hidden variables). This model enables one to view each individual pair of photons as having a definite state of polarization in every plane simultaneously. Bell's theorem[9] revealed a contradiction between this model and quantum mechanics, so it became possible to make an experimental decision,[3,7,10] which disproved the model of local hidden variables.

At present, as far as we know, there is no third way within alternative (b). Thus, the apparent untenability of this alternative is what is intuitively paradoxical about quantum distant correlations[11]: It remains either to reject physical realism independent of the measuring arrangements or to consider seriously alternative (a). One wonders if Einstein were alive today how he would react to this dilemma, to which the new experimental facts have brought us. We believe that alternative (a) deserves systematic investigation. We feel that quantum correlations in the ensemble $\phi_{12}$ are something real, and that the correlation operator $U_a$ plays a key role in their understanding (cf. Sec. VA).

The two basic aims of this article are as follows. (i) To explore quantum correlations in any pure composite state $\phi_{12}$ from the point of view of measurement. In other words, since twin observables are the basic form how the correlation operator $U_a$ shows up, we study *twin observables in general* (i.e., without the restrictions imposed in the previous article[1]). (ii) To study different conditions under which quantum correlations show up in a nontrivial way, i.e., when one has a *general EPR-type state vector*.

For the second aim it will turn out that twin symmetry operators are useful. Therefore, it is desirable to investigate twin observables and twin symmetry operators from a unified point of view, as particular cases of twins of normal operators (cf. Sec. IIB).[12]

## II. MATHEMATICAL INTERMEZZO

### A. Description of correlated subsystems in terms of the polar factors of antilinear operators

If $H_1$ and $H_2$ are the state spaces of the two subsystems of a composite quantum system, then the Hilbert space of antilinear operators $A_a$ mapping $H_1$ into $H_2$ and satisfying $\mathrm{Tr}_1 A_a^\dagger A_a < \infty$ is a realization[13–15] of the tensor product $H_1 \otimes H_2$.

A simple way to see the meaning of $A_a$ that corresponds to a given composite state $\phi_{12}$ is to choose an arbitrary orthonormal basis $\{\varphi_n | n = 1,2,...\}$ in $H_1$ and to expand[1] $\phi_{12}$ in this basis:

$$\phi_{12} = \sum_n \varphi_n \otimes (A_a \varphi_n). \tag{1}$$

The physical interpretation of (1) is as follows: When a first-subsystem measurement results in $\varphi_n$, the second subsystem is by this very fact in the state $A_a\varphi_n / \|A_a\varphi_n\|$. Besides, the square of the norm $\|A_a\varphi_n\|^2$ is the probability of this result.[1]

The advantage of the antilinear-operator realization of $H_1 \otimes H_2$ lies in the fact that $A_a$ connects $H_1$ with $H_2$, and thus it is well suited for the description of the quantum correlations between the two subsystems.

The measurement of a first-subsystem observable that has $\{\varphi_n | n = 1,2,\cdots\}$ as its eigenbasis is not necessarily a measurement on the second subsystem. It is such a distant measurement on the second subsystem if and only if the "relative states"[16] $\{A_a\varphi_n | n = 1,2,\cdots\}$ are orthogonal. This is the case if and only if[1] $\{\varphi_n | n = 1,2,\cdots\}$ is an eigenbasis of the reduced statistical operator $\rho_1 \equiv \mathrm{Tr}_2 |\phi_{12}\rangle\langle\phi_{12}|$ (which means that the measured observable is compatible with $\rho_1$). Then (1) becomes the Schmidt canonical form

$$\phi_{12} = \sum_m r_m^{1/2} \varphi_m \otimes (U_a \varphi_m), \tag{2}$$

where

$$\rho_1 = \sum_m r_m |\varphi_m\rangle\langle\varphi_m|, \tag{3}$$

all $r_m > 0$, and

$$A_a = U_a \rho_1^{1/2} \tag{4}$$

is the polar factorization of $A_a$ (cf. Appendix 4 of Ref. 1).

In the context of distant measurement the two polar factors of $A_a$ have separate physical meanings in statistical terms: $\rho_1$ describes the improper ensemble[17] of first subsystems implied by the proper ensemble of composite systems represented by $\phi_{12}$; $U_a$ is the correlation operator[1] connecting the states $\varphi_m$ obtained in the direct measurement with the states $U_a\varphi_m$ that come about in the distant measurement. Actually, $U_a$ determines the subensemble $(U_a\varphi_m)$ of second subsystems that is singled out in distant measurement (when the direct measurement has selected the subensemble $\varphi_m$).

### B. Normal operators as twins

*Definition 1*: Let $H_1$ and $H_2$ be the state spaces of two subsystems and let $\phi_{12} \in H_1 \otimes H_2$ be a composite state vector. Two normal bounded operators $A_1$ in $H_1$ and $A_2$ in $H_2$ are called *twin operators* with respect to $\phi_{12}$ if they satisfy

$$A_1 \phi_{12} = A_2^\dagger \phi_{12} \tag{5a}$$

and

$$A_1^\dagger \phi_{12} = A_2 \phi_{12}. \tag{5b}$$

**Theorem 1**: Conditions 5(a) and (b) are equivalent to

$$[A_1,\rho_1]=0 \qquad (6a)$$

and

$$A_2Q_2=U_aA_1U_a^{-1}Q_2, \qquad (6b)$$

where $Q_2$ projects onto $R(\rho_2)$, the range of $\rho_2\equiv\mathrm{Tr}_1|\phi_{12}\rangle\langle\phi_{12}|$.

*Proof*: Let us assume the validity of (5a) and (5b). Then, utilizing $A_1\,\mathrm{Tr}_2B_{12}=\mathrm{Tr}_2A_1B_{12}$, $\mathrm{Tr}_2\,B_{12}A_1=(\mathrm{Tr}_2\,B_{12})A_1$, and $\mathrm{Tr}_2\,A_2B_{12}=\mathrm{Tr}_2\,B_{12}A_2$ (which are valid for every bounded linear operator $B_{12}$ in $H_1\otimes H_2$ as can be easily checked), one can write $A_1\rho_1=\mathrm{Tr}_2\,A_1|\phi_{12}\rangle\langle\phi_{12}|=\mathrm{Tr}_2\,A_2^\dagger|\phi_{12}\rangle\langle\phi_{12}|$ $=\mathrm{Tr}_2|\phi_{12}\rangle\langle\phi_{12}|A_2^\dagger=\mathrm{Tr}_2|\phi_{12}\rangle\langle\phi_{12}|A_1=\rho_1A_1$, which proves (6a). Therefore, we can take a common eigenbasis $\{\varphi_m|m=1,2,\cdots\}$ of $\rho_1$ and of $A_1$ (hence also of $A_1^\dagger$) in $R(\rho_1)$, the range of $\rho_1$. Expanding $\phi_{12}$ in this basis, one obtains a Schmidt canonical form (2). Replacing (2) in (5a), one arrives at

$$\sum_m r_m^{1/2}(A_1\varphi_m)\otimes(U_a\varphi_m)=\sum_m r_m^{1/2}\varphi_m\otimes(A_2^\dagger U_a\varphi_m). \qquad (7)$$

Owing to $A_1\varphi_m=a_m\varphi_m$, partial scalar product (cf. Appendix 1 of Ref. 1) of $r_m^{-1/2}\varphi_m$ with (7) gives $A_2^\dagger U_a\varphi_m$ $=a_mU_a\varphi_m$ or $A_2U_a\varphi_m=a_m^*U_a\varphi_m=U_aA_1U_a^{-1}(U_a\varphi_m)$. Since $\{U_a\varphi_m|m=1,2\cdots\}$ spans $R(\rho_2)$, (6b) follows.

If on the other hand, (6a) is valid, then (2) and $A_1\varphi_m=a_m\varphi_m$ follow as above. Further, Eq. (6b) implies $A_2U_a\varphi_m=a_m^*U_a\varphi_m$, and $A_2^\dagger U_a\varphi_m=a_mU_a\varphi_m$. Consequently, (7) and (5a) hold true. One proves (5b) analogously. Q.E.D.

Since $A_1$ and $A_2$ play symmetrical roles in (5a) and (5b), the latter are equivalent to

$$[A_2,\rho_2]=0, \qquad (8a)$$

$$A_1Q_1=U_a^{-1}A_2U_aQ_1, \qquad (8b)$$

where $Q_1$ is the range projector of $\rho_1$.

Furthermore, as (5b) is symmetrical to (5a) with respect to adjoining, one has two more pairs of equations equivalent to (5a) and (5b):

$$[A_1^\dagger,\rho_1]=0, \qquad (9a)$$

$$A_2^\dagger Q_2=U_aA_1^\dagger U_a^{-1}Q_2, \qquad (9b)$$

and

$$[A_2^\dagger,\rho_2]=0, \qquad (10a)$$

$$A_1^\dagger Q_1=U_a^{-1}A_a^\dagger U_aQ_1. \qquad (10b)$$

The reduced statistical operators $\rho_1$ and $\rho_2$, as well as their range projectors $Q_1$ and $Q_2$, are basic examples of twin operators. This follows from[1]

$$\rho_2=U_a\rho_1U_a^{-1}Q_2, \qquad (11)$$

and from

$$Q_1\phi_{12}=\phi_{12}=Q_2\phi_{12}, \qquad (12)$$

that is evident when $\phi_{12}$ is written in a Schmidt canonical form (2), respectively.

### C. Hermitian twins

Let us now discuss the most important class of normal operators—the Hermitian ones. In this case conditions (5a) and (5b) reduce into one equation,

$$A_1\phi_{12}=A_2\phi_{12}. \qquad (13)$$

However, the equivalent conditions (6a) and (6b) are both required when a given pair of operators $A_1$ and $A_2$ are tested, whether they are twins or not.

If, on the other hand, one asks the question which first-subsystem observable $A_1$ has a twin when $\phi_{12}$ is given, then Eq. (13) is of no use. But (6a) by itself gives a complete answer to this question.

An observable $A_1$ has a twin if and only if it is compatible with $\rho_1$. The proof of this statement is obvious if the right-hand side of (6b) is understood as a prescript for the construction of an $A_2$ observable.

### D. Unitary twins

If $U_1$ and $U_2$ are unitary operators in $H_1$ and $H_2$, respectively, then, owing to commutation of any operator from $H_1$ with any one from $H_2$, it follows immediately from Definition 1 that $U_1$ and $U_2$ are twins if and only if

$$U_1U_2\phi_{12}=\phi_{12} \qquad (14)$$

($U_1U_2=U_1\otimes U_2$), or equivalently (according to Theorem 1),

$$[U_1,\rho_1]=0 \qquad (15a)$$

and

$$U_2Q_2=U_aU_1U_a^{-1}Q_2. \qquad (15b)$$

If we consider the maximal symmetry group $G_1$ of $\rho_1$, i.e., all $U_1$ satisfying (15a), and the analogous group $G_2$ of $\rho_2$, then each of these groups is broken up into equivalence classes where equivalent operators are those that reduce into the same operator in the range of the corresponding $\rho$. In other words, equivalent operators differ only in the corresponding null space $N(\rho)$. Thus, the canonical operator in each class is that among the elements of the latter which acts as the identity operator in $N(\rho)$.

Denoting by $I_1$ the identity operator in $H_1$, and by $Q_1^\perp$ the complementary projector $(I_1-Q_1)$ of $Q_1$, the canonical operator equivalent to $U_1$ [satisfying (15a)] is

$$U_1^c=U_1Q_1+Q_1^\perp. \qquad (16)$$

In both $H_1$ and $H_2$ the canonical operators form groups that we denote $G_1^c$ and $G_2^c$, respectively.

The correlation operator $U_a$ gives via (15b) an isomorphism between $G_1^c$ and $G_2^c$, enabling one to single out the subgroup $(G_1^c\times G_2^c)_d$, the so-called diagonal of the direct product $G_1^c\times G_2^c$, consisting of the ordered pairs of the form $(U_1Q_1+Q_1^\perp, U_aU_1U_a^{-1}Q_2+Q_2^\perp)$, $U_1\in G_1$.

Now, we can rephrase (15a) and (15b) as follows: Two unitary operators $U_1$ and $U_2$ are twins if and only if $U_1\in G_1$, $U_2\in G_2$, and $(U_1^c,U_2^c)\in(G_1^c\times G_2^c)_d$. We denote by $G_{12}$ the group of all $U_1U_2$ in $H_1\otimes H_2$, where $U_1$ and $U_2$ are twins.

## III. DISTANT CORRELATIONS IN TERMS OF MEASUREMENT

### A. Detectable part of a subsystem observable

Since the measurement of $A_1$ compatible with $\rho_1$ on $\phi_{12}$ lies at the root of the study of twin observables, we first concentrate on it. As a matter of fact, it can be replaced by the measurement of the *detectable part* of $A_1$: $A_1Q_1$ (on $\phi_{12}$).

Before we elaborate this, we have to derive a suitable spectral form of $A_1Q_1$ from the spectral form of $A_1$.

The operator $A_1Q_1$ has necessarily a purely discrete spectrum whatever is the spectrum of $A_1$. Namely, $A_1$ reduces in each eigensubspace of $\rho_1$, and all the eigensubspaces corresponding to positive eigenvalues of $\rho_1$ [and making up its range $R(\rho_1) = R(Q_1)$] are finite dimensional (because $\rho_1$ has a purely discrete spectrum, see Ref. 18, p. 329; and $\mathrm{Tr}_1 \rho_1 = 1$). Therefore, the entire possible cotinuous part of the spectral form of $A_1$ falls into the null space of $Q_1$.

If $\Sigma_n a_n P_1^{(n)}$ is the discrete part of the spectral form of $A_1$, and if we enumerate by $m$ those values of $n$ for which $P_1^{(n)}Q_1 \neq 0$, then we have

$$A_1Q_1 = \sum_m a_m P_1^{(m)} Q_1.$$

All terms omitted from $\Sigma_n a_n P_1^{(n)}$ (for which $P_1^{(n)}Q_1 = 0$) correspond to undetectable eigenvalues $a_n$ of $A_1$, because the probability to obtain such a value in the measurement of $A_1$ on $\phi_{12}$ is zero:

$$\begin{aligned} p(a_n, A_1, \phi_{12}) &= \langle \phi_{12} | P_1^{(n)} | \phi_{12} \rangle \\ &= \mathrm{Tr}_{12}\, P_1^{(n)} | \phi_{12} \rangle \langle \phi_{12} | = \mathrm{Tr}_1\, P_1^{(n)} \rho_1 \\ &= \mathrm{Tr}_1\, P_1^{(n)} Q_1 \rho_1 = 0. \end{aligned}$$

The remaining eigenvalues $a_m$ are all detectable because $P_1^{(m)}Q_1 \neq 0$ implies $\mathrm{Tr}_1\, P_1^{(m)}\rho_1 > 0$. To see this, we choose a unit vector $|\varphi\rangle$ such that $P_1^{(m)}Q_1|\varphi\rangle = |\varphi\rangle$. Then $\mathrm{Tr}_1\, P_1^{(m)}\rho_1 = \mathrm{Tr}_1 (P_1^{(m)}Q_1)\rho_1 (P_1^{(m)}Q_1)$

$\geqslant \langle \varphi | (P_1^{(m)}Q_1)\rho_1 (P_1^{(m)}Q_1) | \varphi \rangle = \langle \varphi | \rho_1 | \varphi \rangle > 0.$

In this way we have proved:

*Lemma 1*: Whatever the spectral form of $A_1$ that is compatible with $\rho_1$, the spectrum of $A_1Q_1$ is purely discrete, and one can write

$$A_1Q_1 = \sum_m a_m P_1^{(m)} Q_1, \tag{17}$$

where $m$ enumerates the distinct detectable eigenvalues of $A_1$, i.e., those which have a positive probability in the measurement of $A_1$ on $\phi_{12}$. Decomposition (17) is unique under the requirement

$$Q_1 = \sum_m P_1^{(m)} Q_1, \tag{18}$$

and we refer to (17) as *the suitable spectral form* of $A_1Q_1$.

Owing to $[A_1, \rho_1] = 0$, the range of $\rho_1$ is invariant for $A_1$, and the latter reduces there into its relevant part $A_1^r$.[13] In order to avoid domain restrictions, we utilize the detectable part $A_1Q_1$ (defined in the entire first-subsystem state space) instead of $A_1^r$. However, the suitable spectral form of $A_1Q_1$ corresponds in fact to the standard spectral form of $A_1^r$ [in which the eigenvalues are distinct and the eigenprojectors add up into the identity operator in $R(\rho_1)$].

Now we can elaborate the physical relation between $A_1$ and $A_1Q_1$, that makes them indistinguishable on $\phi_{12}$.

*Lemma 2*: (i) The entire continuous spectrum of $A_1$ that is compatible with $\rho_1$ is undetectable on $\phi_{12}$.

(ii) The probability of a detectable eigenvalue $a_m$ of $A_1$ on $\phi_{12}$ is the same as that of $A_1Q_1$ on $\phi_{12}$: $p(a_m, A_1, \phi_{12}) = p(a_m, A_1Q_1, \phi_{12})$.

(iii) Any predictive measurement of either $A_1$ or $A_1Q_1$ on $\phi_{12}$ giving $a_m$ as the result, converts $\phi_{12}$ into the same state $P_1^{(m)}\phi_{12}/\|P_1^{(m)}\phi_{12}\|$.

*Proof*: (i) Let $D$ be an arbitrary domain on the real axis, and let $P_1^{(D)}$ be its spectral projector (or spectral measure) determined by $A_1$. The probability to obtain a result from $D$ in a measurement of $A_1$ on $\phi_{12}$ is $\mathrm{Tr}_1 \rho_1 P_1^{(D)}$. Since $\rho_1 = Q_1\rho_1$, this probability is zero whenever $P_1^{(D)}Q_1 = 0$, i.e., whenever $R(P_1^{(D)})$ is part of the null space of $\rho_1$. This is the case when $D$ is the continuous spectrum of $A_1$.

(ii) $\mathrm{Tr}_1 \rho_1 P_1^{(m)} = \mathrm{Tr}_1 \rho_1 (P_1^{(m)}Q_1)$.

*(iii)* $P_1^{(m)}\phi_{12}/\|P_1^{(m)}\phi_{12}\| = (P_1^{(m)}Q_1)\phi_{12}/\|P_1^{(m)}Q_1\phi_{12}\|$ due to (12).

Q.E.D.

*Corollary*: Two first-subsystem observables compatible with $\rho_1$ are indistinguishable in measurement on $\phi_{12}$ if and only if their detectable parts coincide. This indistinguishability is obviously an equivalence relation in the set of all first-subsystem observables compatible with $\rho_1$. We take for the canonical representative of any equivalence class the Hermitian operator that acts as zero in the null space $N(\rho_1)$. We call such operators canonical (with respect to $\phi_{12}$).

*Remark*: For any given $\phi_{12}$ the canonical operators of the first subsystem form a Lie algebra $L_1^c$ with "$(i/\hbar)$ times the commutator" as the Lie product.

### B. Twin observables

Now we assume that $A_1$ and $A_2$ are twin observables, and we derive the basic mathematical and physical implications of this relation.

*Lemma 3*: If $A_1$ and $A_2$ are twin observables, then:

(i) $A_1Q_1$ and $A_2Q_2$ are also twins, and vice versa.

(ii) The detectable eigenvalues of $A_1$ and those of $A_2$ are the same, i.e., if (17) is the suitable spectral form of $A_1Q_1$, then that of $A_2Q_2$ is

$$A_2Q_2 = \sum_m a_m P_2^{(m)} Q_2, \tag{19}$$

and

$$Q_2 = \sum_m P_2^{(m)} Q_2. \tag{20}$$

(iii) The eigenprojectors $P_1^{(m)}Q_1$ and $P_2^{(m)}Q_2$ corresponding to the same detectable eigenvalue $a_m$ are also twins.

*Proof*: (i) $A_1Q_1\phi_{12} = A_2Q_2\phi_{12}$ is equivalent to (13) due to (12).

(ii) and (iii) Applying $U_a \cdots U_a^{-1} Q_2$ to (17), and taking into account $A_2Q_2 = U_a(A_1Q_1)U_a^{-1}Q_2$ [cf. (6b)], one obtains

$$A_2Q_2 = \sum_m a_m \left[ U_a (P_1^{(m)}Q_1) U_a^{-1} \right] Q_2. \tag{21}$$

The antiunitary operator $U_a$ takes by similarity transformation orthogonal projectors decomposing the identity operator in $R(\rho_1)$ into orthogonal projectors which decompose the identity operator in $R(\rho_2)$. Hence, (21) is the suitable spectral form of $A_2Q_2$. Further,

$$\forall m \quad [U_a(P_1^{(m)}Q_1)U_a^{-1}]Q_2 = P_2^{(m)}Q_2,$$

where $P_2^{(m)}$ are the eigenprojectors of $A_2$ corresponding to the detectable eigenvalues.

Q.E.D.

The quantum-mechanical meaning of twin observables can be summarized in the following way.

**Theorem 2**: If $A_1$ and $A_2$ are twin observables with respect to $\phi_{12}$ [see Eq. (13)], then their measurements on $\phi_{12}$ are indistinguishable:

(i) The probability $p(a_m, A_1, \phi_{12})$ to obtain a detectable eigenvalue $a_m$ in a measurement of $A_1$ on $\phi_{12}$ is the same as that of $a_m$ when $A_2$ is measured on $\phi_{12}$, i.e., the same as $p(a_m, A_2, \phi_{12})$.

(ii) If the two measurements mentioned in (i) are predictive, they have the same effect on $\phi_{12}$, i.e., they convert the latter into

$$P_1^{(m)}\phi_{12}/\|P_1^{(m)}\phi_{12}\| = P_2^{(m)}\phi_{12}/\|P_2^{(m)}\phi_{12}\|.$$

*Proof*: (i) $p(a_m, A_1, \phi_{12}) = p(a_m, A_1 Q_1, \phi_{12})$ $= \langle\phi_{12}|P_1^{(m)}Q_1|\phi_{12}\rangle = \langle\phi_{12}|P_2^{(m)}Q_2|\phi_{12}\rangle = p(a_m, A_2Q_2, \phi_{12})$ $= p(a_m, A_2, \phi_{12})$ [cf. Lemma 2(ii) and Lemma 3(iii)].

(ii) Follows immediately from Eq. (12) and Lemma 3(iii).

Q.E.D.

Thus, a direct measurement of $A_1$ is by this very fact a *distant measurement* of $A_2$ and vice versa. The term "distant" refers to the fact that the measurement of a first-subsystem observable $A_1 \equiv A_1 \otimes I_2$ requires lack of interaction between the measuring apparatus and the second subsystem. The concept of distant measurement was introduced in previous work[1] for the special case of complete observables $A_1$. Now we have extended this concept to all first-subsystem observables compatible with $\rho_1$.

In distant-correlation experiments (which were invented to decide for or against local hidden variable theories),[7,10] as a rule one deals with a special kind of twin observables—with twin projectors $P_1$ and $P_2$, having the physical meaning of simultaneous occurrence of events on distant subsystems (e.g., the first photon goes or does not go through an analyzer, and the same happens with the second photon; see Discussion C in Ref. 1). These twin projectors $P_1$ and $P_2$ provide us with an important example of distant measurement: When the event $P_1$ happens in the laboratory, then $P_2$ occurs on the distant subsystem. The coincidence measurements in the above experiments check this quantum-mechanical statement confirming it.

## IV. DISTANT CORRELATIONS IN THE EPR CASE

### A. Criteria

*Definition 2*: A composite state vector $\phi_{12}$ is an EPR-type state vector ( a $\phi_{12}^{\rm EPR}$) if there exist two first-subsystem observables $A_1$ and $B_1$ such that both are compatible with $\rho_1$ and that their detectable parts $A_1Q_1$ and $B_1Q_1$ are *incompatible* with each other. In other words, this condition means that the Lie algebra $L_1^c$ (see Remark) is nonabelian.

Thus, in a $\phi_{12}^{\rm EPR}$ one can measure distantly (i.e., without interaction with the second subsystem) either of the two twin observables $A_2Q_2$ and $B_2Q_2$, which are necessarily [due to 6(b)] incompatible with each other. We believe this is a natural generalization of the original EPR state vector[4] (where $A_1$ was the coordinate and $B_1$ was the linear momentum), as well as of all the other examples studied in the literature since 1935.[7,10]

An obvious necessary and sufficient condition for a $\phi_{12}$ to be a $\phi_{12}^{\rm EPR}$ is that at least one positive eigenvalue of $\rho_1$ (or equivalently of $\rho_2$) be degenerate. Necessity is due to the fact that $[A_1, \rho_1] = 0$ and $[B_1, \rho_1] = 0$ imply that both $A_1$ and $B_1$ reduce in each eigensubspace of $\rho_1$ in $R(\rho_1)$. Unless one of these eigensubspaces is more than one-dimensional, $A_1Q_1$ and $B_1Q_1$ have to commute. Sufficiency is obvious.

A group-theoretical version of this condition is given in the following theorem.

**Theorem 3**: A state vector $\phi_{12}$ is of the EPR type if and only if its symmetry group $G_1^c$ is nonabelian.

*Proof*: The group $G_1^c$ is a Lie group, and its Lie algebra is $L_1^c$. The latter is nonabelian if and only if so is $G_1^c$.

Q.E.D.

### B. Schmidt canonical form

It may not be realized that the Schmidt canonical form of a given $\phi_{12}$ is, in general, nonunique. If $\phi_{12}$ is not of the EPR type, i.e., if all positive eigenvalues of $\rho_1$ are nondegenerate, then the Schmidt canonical form (2) is unique:

$$\phi_{12} = \sum_m r_m^{1/2}\varphi_m \otimes (U_a\varphi_m).$$

Namely, the eigenbasis of $\rho_1$ in $R(\rho_1)$ is unique up to a phase factor $\exp(i\lambda_m)$ for each $m$ independently. But, owing to the antilinear nature of $U_a$, one has $U_a \exp(i\lambda_m)\varphi_m$ $= \exp(-i\lambda_m)U_a\varphi_m$, hence this freedom cancels out, leaving each $\varphi_m \otimes (U_a\varphi_m)$ unchanged.

On the other hand, if $\phi_{12}$ is of the EPR type, then there exists at least one degenerate eigensubspace $V(r_m)$, $r_m > 0$, of $\rho_1$, in which there are orthonormal bases differing from each other more than by a permutation or by phase factors. Since each eigenbasis in $R(\rho_1)$ gives a Schmidt canonical form (2), one thus obtains different forms of this kind, i.e., expansions (2) differing more than by rearrangement of the terms.

**Theorem 4**: The group $G_{12}$ of $\phi_{12}$ is the symmetry group of the Schmidt canonical form of $\phi_{12}$, i.e., for every two canonical forms (2) there exists one element $U_1U_2$ of $G_{12}$ taking one into the other; and vice versa, each element of $G_{12}$, when applied to an expansion (2), gives again such an expansion (which is not necessarily a different one).

*Proof*: Let

$$\sum_m r_m^{1/2}\varphi_m \otimes (U_a\varphi_m) = \phi_{12} = \sum_m r_m^{1/2}\psi_m \otimes (U_a\psi_m)$$

be two canonical forms of $\phi_{12}$. The two eigenbases $\{\varphi_m | m = 1,2,\cdots\}$ and $\{\psi_m | m = 1,2,\cdots\}$ of $\rho_1$ in $R(\rho_1)$ define [nonuniquely in $N(\rho_1)$] an element $U_1 \in G_1$: $\psi_m = U_1\varphi_m$, $m = 1,2,\cdots$, that obviously commutes with $\rho_1$. Let $U_2$ be a twin of $U_1$. Then $\psi_m \otimes (U_a\psi_m) = (U_1\varphi_m) \otimes (U_a U_1\varphi_m)$, and making use of 15(b), one further has

$$\psi_m \otimes (U_a\psi_m) = (U_1\varphi_m) \otimes (U_2U_a\varphi_m).$$

The proof of the converse statement runs along the same lines in the opposite direction.

Q.E.D.

## V. DISCUSSION

### A. On the physical meaning of $\rho_1$, $U_a$, and $\rho_2$

Though $\phi_{12}$ and the pair of operators $\rho_1$, $U_a$ are mathematically equivalent (cf. Theorems 5 and 7 in Ref. 1), physically $\rho_1$ and $U_a$ do not have separate meanings if *all observables* (measurable on the composite system) are taken into account. Restriction to the class of first-subsystem observables $A_1 \otimes I_2$ endows the notion of $\rho_1$ with physical contents, whereas further restriction to the subclass of observables compatible with $\rho_1$ ($[A_1,\rho_1] = 0$) gives physical basis to the concept of the correlation operator $U_a$. [The observables of this subclass are precisely those which have twins $A_2Q_2 = U_aA_1U_a^{-1}Q_2$—cf. (6b)—among the second-subsystem observables.]

Therefore, one cannot disagree with Bohr[19] that the state $\phi_{12}$ of the composite system is actually an unseparable whole, but this does not prevent one from exploring the conditions under which the "parts" (the two subsystems and the correlation between them) have separate physical meaning.

When the subsystems are *distant* (i.e., sufficiently far apart from each other so that they are not interacting), but *correlated* (e.g., have interacted in the past), then the typical experiments are coincidence measurements.[7] These are measurements of composite events of the type $P_1P_2'$, where $P_1$ is some event (projector) in $H_1$ (e.g., a linear polarization analyzer orientated in a certain direction and completed with a detector measuring the event of "passing through" in case of photons), and $P_2'$ is an independently chosen event in $H_2$ (e.g., one measured by a differently orientated polarization-measuring arrangement). We assume that $P_1$ is compatible with $\rho_1$, and we argue as follows.

The probability $p(P_1P_2',\phi_{12}) \equiv p(1,P_1P_2',\phi_{12})$ of the occurrence of $P_1P_2'$ in the state $\phi_{12}$ can be broken down to the conditional probability $p(P_2',\phi_{12}|P_1)$ of $P_2'$ under the condition that $P_1$ took place, and to the probability $p(P_1,\phi_{12})$ of $P_1$:

$$p(P_1P_2',\phi_{12}) = p(P_1,\phi_{12})p(P_2',\phi_{12}|P_1). \tag{22}$$

Evidently,

$$p(P_1,\phi_{12}) = \mathrm{Tr}_1\, P_1\rho_1. \tag{23}$$

Further,

$$p(P_2',\phi_{12}|P_1) = \mathrm{Tr}_2\, P_2'\rho_2(P_1), \tag{24}$$

where $\rho_2(P_1)$ is that subensemble of $\rho_2 \equiv \mathrm{Tr}_1|\phi_{12}\rangle\langle\phi_{12}|$ which corresponds to the subensemble $P_1\rho_1P_1/\mathrm{Tr}_1\, P_1\rho_1$ obtained by the occurrence of $P_1$:

$$\rho_2(P_1) = P_2\rho_2P_2/\mathrm{Tr}_1\, P_1\rho_1, \tag{25}$$

where $P_2$ is the twin event of $P_1$, i.e.,

$$P_2 \equiv U_aP_1U_a^{-1}Q_2, \tag{26}$$

and $\mathrm{Tr}_2\, P_2\rho_2 = \mathrm{Tr}_1\, P_1\rho_1$. Actually, Eq. (25) is a special case of the more familiar general expression

$$\rho_2(P_1) = \mathrm{Tr}_1\, P_1|\phi_{12}\rangle\langle\phi_{12}|P_1/\mathrm{Tr}_1\, P_1\rho_1, \tag{27}$$

obtained from the latter by utilizing (13).

One should note that the above argument reduces any coincidence experiment on distant subsystems to the measurement of $P_2'$ on the *distantly prepared* subensemble $\rho_2(P_1)$. The restriction of the choice of $P_1$ to events compatible with $\rho_1$ means that the distant preparation is, in fact, the distant occurrence of the twin event $P_2$. Thus, coincidence in this case actually reduces to successive measurements of $P_2$ and of $P_2'$ (they need not be compatible with each other).

As seen from Eq. (26), it is the correlation operator $U_a$ that determines which event $P_2$ is the twin of $P_1$. For instance, in the well-known Freedman-Clauser experiment,[3] the two-photon polarization state $\phi_{12}$ implies a $U_a$ such that $P_1$ and $P_2$ correspond to parallel orientations of the analyzers; whereas in another known experiment,[20] $P_1$ and $P_2$ correspond to perpendicular orientations.

The correlation operator $U_a$ is an entity endowed with physical meaning to the extent to which the restriction $[P_1,\rho_1] = 0$ is natural. The weaker restriction to any subsystem events $P_1$ and $P_2'$ is actually not a restriction, because on distant subsystems there is nothing else to be measured. As far as we know, in all experiments performed so far, $[P_1,\rho_1] = 0$ was no restriction either due to $\rho_1 = \frac{1}{2}I_1$. Therefore, in these cases the physical meaning of $U_a$ seems to have been established beyond doubt.

As to a general $\phi_{12}$ describing two distant and correlated suybsystems, the requirement $[P_1,\rho_1] = 0$ is a restriction. It selects out an important class of measurements because this requirement is equivalent to the following: (i) The occurrence of $P_1$ is a no-disturbance direct measurement.[21] (ii) The distantly prepared subensemble $\rho_2(P_1)$ comprises precisely those distant subsystems on which an event $P_2$ occurs. In other words, when $[P_1,\rho_1] \neq 0$, then the nonselective[21] direct measurement of $P_1$ *changes* $\rho_1$ (i.e., $P_1\rho_1P_1 + (I_1 - P_1)\rho_1(I_1 - P_1) \neq \rho_1$), and $\rho_2$ decomposes into the distantly prepared subensemble $\rho_2(P_1)$ [given by (27)] and the remainder, but these two are *not orthogonal* to each other.

To draw a conclusion from the above argument, one should bear in mind that quantum correlations are a kind of entanglement of the predictions of subsystem events, and that there is no other way to disentangle them than to perform subsystem measurements.[2] Therefore, no-disturbance measurements on both subsystems (equivalent to $[P_1,\rho_1] = 0$) seem to be best suited for the study of observable consequences of quantum correlations. On the other hand, this same condition $[P_1,\rho_1] = 0$ makes it possible for the correlation operator $U_a$ to play an important role [determining $\rho_2(P_1)$ via Eqs. (25) and (26)]. Hence, $U_a$ describes basic aspects of quantum correlations in the general state $\phi_{12}$ under the given conditions.

### B. What is paradoxical in distant measurement in the EPR case?

Let us return to this question put in the Introduction. Two possibilities (a) and (b) were given, and it was pointed out that alternative (b) had been disproved experimentally. Now we discuss alternative (a), and we point to two essential aspects of the change taking place as a result of the direct measurement.

(i) When an observable $A_1$, compatible with $\rho_1$, is selected, one has before its direct measurement decomposition (2):

$$\phi_{12}^{\rm EPR} = \sum_m r_m^{1/2} \varphi_m \otimes (U_a \varphi_m),$$

where $\{\varphi_m | m = 1,2,\cdots\}$ is a common eigenbasis of $A_1$ and of $\rho_1$ in $R(\rho_1)$. In the direct measurement of $A_1$, $\phi_{12}^{\rm EPR}$ *collapses* into

$$\rho_{12}(A_1) \equiv \sum_m r_m |\varphi_m\rangle\langle\varphi_m| \otimes (U_a|\varphi_m\rangle\langle\varphi_m|U_a^\dagger). \qquad (28)$$

The entire improper ensemble of second subsystems

$$\rho_2 \equiv {\rm Tr}_1 |\phi_{12}^{\rm EPR}\rangle\langle\phi_{12}^{\rm EPR}| = \sum_m r_m U_a |\varphi_m\rangle\langle\varphi_m| U_a^\dagger \qquad (29)$$

was decomposable, i.e., potentially decomposed, into the subensembles $\{U_a|\varphi_m\rangle\langle\varphi_m|U_a^\dagger | m = 1,2,\cdots\}$ also before the measurement. In the collapse $\phi_{12}^{\rm EPR} \to \rho_{12}(A_1)$ the composite system, containing the distant subsystem, undergoes a physical change that has been checked and proved in coincidence measurements of the $P_1P_2'$ type (cf. Sec. VA). The ensemble $\rho_2$ does not change in the collapse because ${\rm Tr}_1|\phi_{12}^{\rm EPR}\rangle\langle\phi_{12}^{\rm EPR}| = {\rm Tr}_1\,\rho_{12}(A_1)$, but its *potential decomposition* (29) *becomes actual* as given by (28), and this takes place without any interaction with the second subsystem. Namely, the occurrences of $P_1^{(m)} \equiv |\varphi_m\rangle\langle\varphi_m|$ on the first subsystem separate out distantly the subensembles

$$P_2^{(m)}\rho_2 P_2^{(m)}/{\rm Tr}_1\, P_1^{(m)}\rho_1 = U_a|\varphi_m\rangle\langle\varphi_m|U_a^\dagger. \qquad (30)$$

From the point of view of von Neumann's quantum theory of measurement,[18] the direct measurement of $A_1$ on the first subsystem is the second link in a two-link chain, where the first link is the composite state $\phi_{12}^{\rm EPR}$. Von Neumann has shown that the very interaction of the measuring apparatus with the first subsystem gives rise to the collapse $\phi_{12}^{\rm EPR} \to \rho_{12}(A_1)$. (We do not discuss the total collapse of the entire chain, which is the well-known problem of the quantum theory of measurement.)

Thus, the collapse $\phi_{12}^{\rm EPR} \to \rho_{12}(A_1)$ is puzzling by itself. But in the EPR case, there is more to it.

(ii) The nonuniqueness of the Schmidt canonical form (2) (cf. Sec. IVB) allows any of an infinite number of collapsed composite ensembles $\rho_{12}(A_1)$ (but they are not simultaneously realizable if one selects incompatible $A_1$). This has the consequence that $\rho_2$ can be actually decomposed in any of a number of mutually incompatible ways implied by (28) without any interaction with the second subsystem.


[1] F. Herbut and M. Vujičić, Ann. Phys. **96**, 382 (1976).
[2] E. Schrödinger, Proc. Cambridge Philos. Soc. **31**, 555 (1935); **32**, 446 (1936).
[3] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972).
[4] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[5] D. Bohm and Y. Aharonov, Phys. Rev. **108**, 1070 (1957).
[6] C. S. Wu and I. Shaknov, Phys. Rev. **77**, 136 (1950).
[7] J. F. Clauser and A. Shimony, Rep. Prog. Phys. **41**, 1881 (1978).
[8] C. A. Hooker, in *Paradigms and Paradoxes*, edited by R. G. Colodny (Univ. of Pittsburgh, Pittsburgh, PA, 1972), p. 67.
[9] J. S. Bell, Physics **1**, 195 (1965).
[10] A. Aspect, in *Mathematical Problems in Theoretical Physics*, Lecture Notes in Physics 153, edited by R. Schrader, R. Seiler, and D. A. Uhlenbrock (Springer-Verlag, Berlin, 1982), p. 162.
[11] M. Vujičić and F. Herbut, Fizika **12** (3), 74 (1980).
[12] Preliminary results were presented in a poster session at the Sixth International Conference on Mathematical Physics, Berlin, 1981.
[13] F. Herbut and M. Vujičić, in *Foundations of Quantum Mechanics*, edited by B. d'Espagnat (Academic, New York, 1971), p. 316.
[14] J. M. Jauch, *Foundations of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1968).
[15] N. Ja. Vilenkin, *Special Functions and Group Representation Theory* (Nauka, Moscow, 1965) (in Russian), p. 75.
[16] H. Everett III, Rev. Mod. Phys. **29**, 454 (1957).
[17] B. d'Espagnat, *Conceptual Foundations of Quantum Mechanics* (Benjamin, Menlo Park, CA, 1971), Subsection 6.3. In the second edition (Benjamin, London, 1976), this subsection is omitted.
[18] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton Univ., Princeton, NJ, 1955).
[19] N. Bohr, Phys. Rev. **48**, 696 (1935).
[20] H. Langhoff, Z. Phys. **160**, 186 (1960).
[21] F. Herbut, Ann. Phys. **55**, 271 (1969).

# Propensities and the state-property structure of classical and quantum systems[a]

N. Gisin
*Département de Physique Théorique, Université de Genève, 1211 Genève 4, Switzerland*

(Received 23 August 1983; accepted for publication 9 December 1983)

In quantum physics the tests of most properties do not have predetermined outcomes. The latter have nevertheless well-defined probabilities of being realized during a test. Following Popper we interpret these probabilities as physical propensities. A first purpose of the present article is to formalize the propensity interpretation in the framework of state-property structures. Next, Gleason's theorem asserts that in the Hilbert space there exists a unique propensity function (i.e., one probability measure for each state vector); the propensities are thus uniquely determined by the state vector. Conversely, we prove that if the state-property structure admits *one and only one* propensity function, then the set $\mathscr{L}$ of all properties is a complete atomic orthomodular lattice. We point out that according to our assumption the probabilistic aspect of the system is entirely determined by its deterministic aspect. Assuming furthermore that each property can be ideally tested, it follows that $\mathscr{L}$ is isomorphic to the direct union of Hilbertian space lattices. We recover thus the purely classical and purely quantum frameworks as the two extreme cases. The intermediate cases correspond to quantum mechanics with—possibly continuous—superselection variables. Finally, we prove that a system is classical, i.e., all properties are mutually compatible, if and only if the propensity function is dispersion free. In our approach the quantum probabilities appear thus as a generalization of classical determinism rather than a generalization of classical probabilities.

PACS numbers: 03.65.Ca, 02.50. + s

## 1. INTRODUCTION

In Ref. 1 B. d'Espagnat wrote: "most predictions of quantum mechanics are of a statistical nature and therefore make sense only for ensembles." This is probably the root of the discomfort that many people feel about quantum mechanics. Yet, in the late 1950's, Sir Karl R. Popper argued that a different interpretation of probability, called the propensity interpretation, solves the problem of single events, and in turn, the problem of the interpretation of quantum mechanics.[2,3] Indeed, d'Espagnat's statement refers to the frequency interpretation of probabilities, but is in opposition to the propensity interpretation.

We shall come back to the propensity concept in Sec. 3. For the time being, let us just briefly quote Popper[2]: "I propose a new physical hypothesis. The two slits experiment convinced me that probabilities (...) are physical propensities, comparable to Newtonian forces, (...) to realize singular events."

The first purpose of the present article is to formalize Popper's idea in the context of state-property structure.[4]

Another important motivation is the Gleason theorem, which states that there exists one and only one probability measure on the set of closed subspaces of a Hilbert space, with value one on a given ray.[5] We remind that in the usual Hilbert space quantum mechanics the properties are represented by the closed subspaces. A property is then called actual whenever the corresponding subspace contains the state vector. Consequently, any (pure) state is then completely and uniquely determined by the set of all actual properties, and, in turn, any (pure) state completely and uniquely determines the "propensity of any property to realize itself during a measurement." This is a beautiful result. However, it seems to us that the conclusion is physically more natural than the Hilbert space assumption. Accordingly, the second purpose of the present article is to prove a theorem which is in a way the converse of Gleason's one (see Sec. 5).

Our main result is the following: If the state-property structure (see Sec. 2) admits one *and only one* propensity function (see Sec. 3), and if each property can be ideally tested (Sec. 4), then the states are naturally represented by atoms of the property lattice $\mathscr{L}$, and $\mathscr{L}$ is isomorphic to the direct union of Hilbert space lattices (Sec. 5). Hence the system is either purely classical (all Hilbert spaces are of dimension one), or purely quantum (only one Hilbert space), or quantum with—possibly continuous—superselection variables.[6]

In Sec. 6 we characterize compatible properties and classical systems in terms of the propensity function. In the last section we summarize the conclusions.

## 2. THE STATE-PROPERTY STRUCTURE

In this section we first fix the notations, and then recall the concept of a property of a physical system.[6–9]

A state-property structure (S.P.S in short) is a triplet $(\Sigma_\perp, \mathscr{L}, \sigma)$ where $\Sigma$ is a set, whose elements represent all possible (pure) states of the system, and $\perp$ is an orthogonality relation on $\Sigma$: two states $\epsilon$, $\eta$ are orthogonal, $\epsilon \perp \eta$, iff there is an experiment which gives always a certain outcome $\alpha$

[a] Supported in part by the Swiss National Science Foundation.

  

whenever the initial state is $\epsilon$, and a different outcome $\beta \neq \alpha$ whenever the initial state is $\eta$ (see Ref. 7). The set $\mathscr{L}$ of all properties of the system is a complete lattice (see below). It includes the "never actual property" 0. Finally, $\sigma: \Sigma \rightarrow \mathscr{L}$ is the map which maps each state $\epsilon$ onto the strongest (i.e., smallest) property actual in the state $\epsilon$ [hence $\sigma(\epsilon) \neq 0$]. The order relation on $\mathscr{L}$ and the map $\theta$ are related as follows:

$$a < b \Leftrightarrow (\forall \epsilon \in \Sigma, \quad \sigma(\epsilon) < a \Rightarrow \sigma(\epsilon) < b).$$

Consequently, for all $b \in \mathscr{L}$,

$$b = \vee \{\sigma(\epsilon) | \sigma(\epsilon) < b\}, \tag{1}$$

where $\vee$ denotes the lower upper bound.

The orthogonality relation on $\Sigma$ provides $\mathscr{L}$ with the Aerts-orthogonality relation[7]: $\forall\, c, b \in \mathscr{L}$,

$$c \perp b \Leftrightarrow (\forall\, \epsilon, \varphi \in \Sigma, \quad \sigma(\epsilon) < c \text{ and } \sigma(\varphi) < b \Rightarrow \epsilon \perp \varphi).$$

(We use the same notation for the orthogonality relations on $\Sigma$ and $\mathscr{L}$.) The interpretation of $c \perp b$ will become clear after Theorem III. For the time being let us anticipate that whenever $c$ is actual, a test of $b$ cannot give the positive answer. The orthogonality relation on $\mathscr{L}$ is characterized by ($\forall\, a,b,c \in \mathscr{L}$),

(1) $a \perp b \Rightarrow b \perp a$,

(2) $a < b$ and $b \perp c \Rightarrow a \perp c$,

(3) $a \perp a \Rightarrow a = 0$ (or, equivalently, $a \perp b \Rightarrow a \wedge b = 0$).

In the remaining part of this section we remind the concept of property. A property is something which the system can have in act or not and which can be tested by a yes-no experiment. If the system has the property in act, one says that the property is actual. In that case, whenever a test is carried out, the positive result is certain to be secured, i.e., the positive result is predetermined. Hence, an actual property is nothing but what Einstein called an "element of reality."[10] A typical property of a particle is, for instance, the property of being localized in some space region A. The property is actual whenever the particle is in a state such that a counter outside A can never detect the particle. (In that example, the positive result is secured whenever the counter does not detect the particle.)

Clearly, a property can be actual for some states of the system, but nonactual for other states. If $\{b_i\}_{i\in I}$ is a collection of properties, $\wedge\, b_i$ denotes the property which is actual if and only if all the $b_i$'s are actual. Any test of a $b_i$ is also a test of $\wedge\, b_i$. If the $b_i$'s are never simultaneously actual, then $\wedge\, b_i = 0$ (we identify properties which are always simultaneously actual). The order relation on the set $\mathscr{L}$ of all the properties is defined as follows:

$$a < b \;\Leftrightarrow\; a \wedge b = a.$$

It is straightforward to verify that $\mathscr{L}$ is a complete lattice, with $\wedge\, b_i$ the greatest lower bound.[6–8]

Let us emphasize that whenever one tests a nonactual property, both results, in general, are possible.

Several authors use the word proposition instead of property. But this sounds too much as a logical concept rather than a physical one. We consider the concept of property as a primitive one, but different authors define a property as a set of equivalent yes-no experiments,[6–8] or as an "ideal" yes-no experiment (i.e., a kind of limit of actual yes-no experiments).[4,9]

## 3. PROPENSITIES

In the late 1950's Popper proposed the propensity interpretation of probabilities. He "gave up the frequency interpretation" because of "the problems of interpretating quantum mechanics and the probability of single events" (see Ref. 2). This has raised many interesting discussions (see, e.g., Refs. 11–16). The intuitive idea of propensity can be presented as follows. Assume that the system under consideration is a silver atom which just enters a Stern–Gerlach magnet, and assume that there are two counters after the magnet. It is a well-known empirical fact that the atom has a well-defined probability (depending on its initial state and on the Stern–Gerlach magnet) to localize itself in the "upper" or "lower" counter. There are several possible objective interpretations of this probability.[17] First, the epistemic one, which claims that the atom is always localized at some point, but that it is objectively impossible to know where, as for a classical Brownian particle. The de Broglie–Bohm model of quantum mechanics adopts this interpretation.[18,19] Next, the frequency interpretation claims that the probability refers to ensembles of atoms. The statistical interpretation of quantum mechanics refers to this viewpoint.[20,21] Finally, the propensity interpretation, as we understand it, claims that each single atom is spread in both beams simultaneously, and that the interaction with the counters is such that the atom has a physical propensity of localizing itself in one counter or the other.

In order to measure this physical propensity one makes statistics over many silver atoms in the same initial state, i.e., one measures a frequency. But the distinction between the frequency and the propensity interpretation is sharp: in the former the probability is a characteristic of an ensemble of atoms, whereas in the latter the probability is a characteristic of the interaction of a single atom and the counters. Only the last interpretation takes seriously the fact that certain experiments do not have a predetermined outcome.

Now, the counters could be replaced by different ones, working on different physical principles. Experimentally, the propensity of an atom does not depend on the measuring apparatus. Hence the *propensity is a modality of the properties* and not of the way one tests them.

The above idea is formalized below and in the next section. Bohr insisted that one should never speak of a system without specifying the measurement apparatus. In our framework this means that the propensities of properties which cannot be simultaneously tested, do not necessarily satisfy the law of classical probability.[22] We propose thus the following definition.

*Definition*: Let $(\Sigma_\perp, \mathscr{L}, \sigma)$ be a S.P.S. and $w$: $\Sigma \times \mathscr{L} \rightarrow [0,1]$. $w$ is a *propensity* function iff it satisfies the following conditions:

(1) $w(\epsilon, a) = 1 \Leftrightarrow \sigma(\epsilon) < a \quad \forall \epsilon \in \Sigma, a \in \mathscr{L}$,

(2) $w(\epsilon, \sigma(\eta)) = 0 \Leftrightarrow \epsilon \perp \eta \quad \forall\, \epsilon, \eta \in \Sigma$,

(3) $a < b \Rightarrow \forall \epsilon \in \Sigma, \quad w(\epsilon, a) < w(\epsilon, b)$,

(4) $b_i \perp b_j \quad \forall i \neq j = 1,2,3,\cdots \Rightarrow \forall \epsilon \in \Sigma$,

$$w(\epsilon, \vee b_i) = \sum_i w(\epsilon, b_i),$$

(5) $w(\epsilon, b_i) = 0 \quad \forall i \in I \Rightarrow w(\epsilon, \bigvee_I b_i) = 0$.

The two first conditions follow from the structure of $(\Sigma_\perp, \mathscr{L}, \sigma)$. Condition (3) is obvious. Condition (4) stems from the idea that mutually orthogonal properties can be tested simultaneously. Accordingly, the propensity function $w(\epsilon, \cdot)$ restricted to such a set $\{b_i\}$ must satisfy the usual conditions of a probability function. Condition (5) is imposed for symmetry reasons.

Two examples of S.P.S. with propensity functions are given by classical and quantum mechanics. In the latter example $\Sigma$ is the set of rays of a complex separable Hilbert space $\mathscr{H}$, with the usual orthogonality relation, $\mathscr{L}$ is the lattice of closed subspaces of $\mathscr{H}$, and $\sigma$ is the inclusion. For this example Gleason's theorem asserts that there exists one and only one propensity function.[5] In classical physics $\mathscr{L}$ is the power set of the set of states: $\mathscr{L} = P(\Sigma)$, the orthogonality relation on $\Sigma$ is the trivial one: $\epsilon \perp \eta \Leftrightarrow \epsilon \neq \eta$ and $\sigma(\epsilon) = \{\epsilon\}$.[7] Accordingly, it follows from Conditions (1) and (2) that there exists one and only one propensity function:

$$w(\epsilon, a) = \begin{bmatrix} 1 & \text{if } \epsilon \in a \\ 0 & \text{if } \epsilon \notin a \end{bmatrix}.$$

We now come to a crucial remark. The fact that in the classical case only the propensities "one" and "zero" occur means nothing but the well-known fact that classical (i.e., Newtonian) mechanics is deterministic (or predeterministic, since every experiment has a predetermined outcome). An important consequence of this remark is that propensities are generalizations of classical determinism, rather than generalizations of classical probabilities.

Let us make clear that we do not consider statistical mechanics here. Statistical mixture would be introduced with the help of measure theory applied to the state space $\Sigma$.

## 4. THE HYPOTHESES

In this section we formulate our basic assumptions.

*Axioms*: The S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ is such that

(1) $\sigma$ is one-to-one.

(2) There exists one *and only one* propensity function $w$.

(3) For all $\epsilon \in \Sigma$, $b \in \mathscr{L}$, there is a state $\eta \in \Sigma$ such that $\sigma(\eta) < b$ and $w(\epsilon, b) = w(\epsilon, \sigma(\eta))$.

The central remark for motivating Axioms (1) and (2) is that a statement about a property of an individual system can be falsified if and only if the property is actual. Hence we conclude that the state of a system at time $t_0$ must be completely and uniquely determined by the set of properties actual at that time $t_0$ [Axiom (1)]. This is the Jauch–Piron characterization of states.[23] However, we go further by assuming that, in turn, each state determines completely and uniquely the propensities of all the properties [Axiom (2)]. In other words, Axioms (1) and (2) state that the set of Einstein's elements of reality[10] characterize the state of the system and the propensity of each property.

We now interpret Axiom (3). A test of property $b$ is called ideal iff the state $\eta$ after the test has been carried out and the positive result has been secured, depends only on the initial state $\epsilon$ and on the property $b$. Moreover the test is of the first kind iff $\sigma(\eta) < b$.[24] This implies that an ideal test of the first kind of the property $b$ is also a test of $\sigma(\eta)$. Axiom (3) is thus physically motivated.

## 5. THE MAIN RESULTS

The following theorem is the main result of the present article:

**Theorem I**: If the S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ satisfies Axioms (1) and (2), then

(a) The property lattice $\mathscr{L}$ is atomic, canonically orthocomplemented (i.e., $a \perp b \Leftrightarrow a < b'$) and weakly modular.

(b) $\sigma$ is a bijection between $\Sigma$ and the atoms of $\mathscr{L}$.

(c) If furthermore Axiom (3) holds, and $\Sigma$ contains at least four mutually orthogonal states, then $\mathscr{L}$ is isomorphic to the direct union[6] over a set $\Gamma$ of Hilbertian space[25] lattices:

$$\mathscr{L} \cong \bigvee_{\alpha \in \Gamma} P(\mathscr{H}_\alpha).$$

Let us recall that a Hilbertian space is almost, but not precisely, a Hilbert space.[25–28] In fact, if the field over which the Hilbertian space is defined is a finite extention of the real numbers, then the Hilbertian spaces in Theorem I can be replaced by Hilbert spaces.[29]

Except for the above remark, Theorem I states that there are essentially only two S.P.S. satisfying Axioms (1)–(3), namely, the purely classical one (where all $\mathscr{H}_\alpha$ are of dimension one) and the purely quantum one (where $\Gamma$ contains only one point). The intermediary cases correspond to quantum mechanics with—possible continuous—superselection variables.

The proof of Theorem I is done in several steps.

**Theorem II**: If the S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ satisfies Axioms (1) and (2), then $\mathscr{L}$ is atomistic (i.e., atomic and $\forall b \in \mathscr{L}$, $b = \vee \{p | p \text{ is an atom and } p < b\}$[4]) and $\sigma$ is a bijection between $\Sigma$ and the set of atoms of $\mathscr{L}$.

*Proof II*: First we prove that $\forall\ \epsilon \in \Sigma\ \sigma(\epsilon)$ is an atom. The proof proceeds by contradiction. Assume that $\sigma(\epsilon)$ is not an atom for some state $\epsilon \in \Sigma$. Then, $\exists\ b \neq 0$ such that $b < \sigma(\epsilon)$. And $\exists\ \phi \in \Sigma$ such that $\sigma(\phi) < b$. Let $\mu: \Sigma \times \mathscr{L} \to [0,1]$ be defined by:

$$\mu(\eta, a) = \begin{cases} \lambda w(\phi, a) + (1 - \lambda) w(\epsilon, a) & \text{if } \eta = \epsilon \\ w(\eta, a) & \text{if } \eta \neq \epsilon, \end{cases}$$

where $\lambda \in ]0,1[$. It is easy to check that $\mu$ is a propensity function. But $\mu(\epsilon, b) \neq w(\epsilon, b)$ which contradicts Axiom 2. Hence $\sigma(\epsilon)$ is an atom, $\forall\ \epsilon \in \Sigma$, and $\mathscr{L}$ is atomic. It follows from (1) that $\mathscr{L}$ is in fact atomistic.

Finally we prove that $\sigma$ is surjective onto the atoms of $\mathscr{L}$. Let $p \in \mathscr{L}$, then $p$ is actual for some state $\epsilon \in \Sigma$: $\sigma(\epsilon) < p$. But if $p$ is an atom, then $\sigma(\epsilon) = p$. ∎

Henceforth we identify the states $\epsilon, \eta, \phi, \cdots$ with the atoms and write $\epsilon, \eta, \phi, \cdots \in \mathscr{L}$.

**Theorem III**: If the S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ satisfied Axioms (1) and (2), then for all $a, b \in \mathscr{L}$

$a \perp b \Leftrightarrow (\forall \epsilon \in \Sigma,\ w(\epsilon, a) = 1 \Rightarrow w(\epsilon, b) = 0)$.

*Proof III*: First assume that $a \perp b$ and $w(\epsilon, a) = 1$. Then $\epsilon < a$, and $\epsilon \perp \eta\ \ \forall \eta < b$. Hence $w(\epsilon, \eta) = 0\ \forall \eta < b$. And $w(\epsilon, b) = 0$ because $\mathscr{L}$ is atomistic and $w$ satisfies Condition (5) of a propensity function.

Next we prove the converse. Let $\epsilon < a$, $\eta < b$. One has

$$w(\epsilon, a) = 1 \Rightarrow w(\epsilon, b) = 0$$
$$\Rightarrow w(\epsilon, \eta) = 0 \Rightarrow \epsilon \perp \eta.$$

■

*Corollary IV*: Under the same assumption

$w(\epsilon, a) = 0 \Leftrightarrow \epsilon \perp a$.

**Theorem V**: If the S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ satisfies Axioms (1) and (2), then for all $a \in \mathscr{L}$, $a \neq 1$, there is a state $\epsilon \in \Sigma$ such that $\epsilon \perp a$.

*Proof V*: The proof proceeds by contradiction. Let $c \in \mathscr{L}$, $c \neq 1$ be such that $w(\epsilon, c) \neq 0\ \forall\ \epsilon \in \Sigma$. Then $\forall\ b > c$ one has $w(\epsilon, b) \neq 0\ \forall\ \epsilon \in \Sigma$. Hence, Theorem III implies

$$\forall\ b > c, \quad b^\perp \equiv \{a | a \perp b\} = \{0\}.$$

Let

$$\mu(\epsilon, a) = \begin{Bmatrix} \lambda + (1-\lambda) w(\epsilon, a) & \text{if } a > c \\ w(\epsilon, a) & \text{if not} \end{Bmatrix},$$

where $\lambda \in ]0,1[$. It is straightforward to verify that $\mu$ is a propensity function. But $\mu(\epsilon, c) \neq w(\epsilon, c)\ \ \forall\ \epsilon \not< c$, which contradicts Axiom (2).

■

**Theorem VI**: If the S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ satisfies Axioms (1) and (2), then $\mathscr{L}$ is orthocomplemented and weakly modular, and for all $a, b \in \mathscr{L}$, $a \perp b \Leftrightarrow a < b'$ (where $b'$ is the orthocomplement of $b$).

*Proof VI*: First we prove that $\mathscr{L}$ is orthocomplemented. Put

$$a' = \vee \{b \mid b \perp a\}.$$

By Theorem III and Condition (5) of the propensity function $w$, one gets $a' \perp a$. By Theorem V one has $a \vee a' = 1$. Indeed, if not, there would be a state $\epsilon \in \Sigma$ such that $\epsilon \perp a \vee a'$, hence

$$\epsilon \perp a \Rightarrow \epsilon < a' \Rightarrow \epsilon < a \vee a',$$

which is a contradiction. Accordingly one has

$$w(\epsilon, a') = 1 - w(\epsilon, a) \quad \forall\ \epsilon \in \Sigma,\ a \in \mathscr{L},$$

and the map $': a \rightarrow a'$ is an orthocomplementation.

Next, let $a < b'$. One has

$$w(\epsilon, a) = 1 \Rightarrow w(\epsilon, b') = 1 \Rightarrow w(\epsilon, b) = 0.$$

Hence $a \perp b$. The converse is immediate.

Finally $\mathscr{L}$ is weakly modular. Indeed, it is known that every orthocomplemented lattice which admits a propensity function is weakly modular.[30] For completeness we repeat the proof: Let $b < c$, we want to prove that $c \wedge (c' \vee b) < b$. Let $\epsilon < c \wedge (c' \vee b)$, one has

$$b < c \Rightarrow b \perp c'$$
$$\Rightarrow w(\epsilon, b) = w(\epsilon, b \vee c') - w(\epsilon, c') = 1 - 0 = 1$$
$$\Rightarrow \epsilon < b.$$

■

It should be noticed that a property $b$ is nonactual (i.e., potential) iff $w(\epsilon, b) \neq 1$, but its orthocomplement is actual iff $w(\epsilon, b) = 0$. Hence, $b$ nonactual does not imply $b'$ actual.

**Theorem VII**: let $(\Sigma_\perp, \mathscr{L}, \sigma)$ be a S.P.S. satisfying Axioms (1) and (2). $\mathscr{L}$ satisfies the covering law if and only if the third axiom holds.

*Proof VII*: We first prove the "only if" part. For all $\epsilon \in \Sigma,\ a \in \mathscr{L}$ one has:

$$w(\epsilon, a) = 1 - w(\epsilon, a') - w(\epsilon, a \wedge \epsilon')$$
$$= 1 - w(\epsilon, a' \vee (a \wedge \epsilon'))$$
$$= w(\epsilon, a \wedge (\epsilon \vee a')).$$

$a \wedge (\epsilon \vee a')$ is the Sasaki projection,[4] which corresponds to the usual projection postulate in the case of Hilbert space quantum mechanics. It is an atom, hence a state, whenever $\mathscr{L}$ satisfies the covering law.

We now prove the "if" part of the theorem. Let $\epsilon \in \Sigma$, $a \in \mathscr{L}$, $w(\epsilon, a) \neq 0$. And let $\eta \in \Sigma$, $\eta < a$ be such that $w(\epsilon, a) = w(\epsilon, \eta)$. The existence of such a state $\eta$ is the content of Axiom (3). We want to prove that $\eta = a \wedge (\epsilon \vee a')$. Since $\mathscr{L}$ is orthomodular, one has $\eta = a \wedge (\eta \vee a')$, it is thus sufficient to prove that $\eta \vee a' = \epsilon \vee a'$. This is done in three steps:

(a) $\epsilon < \eta \vee a'$. Indeed, $\eta < a$

$$\Rightarrow a = \eta \vee (\eta' \wedge a)$$
$$\Rightarrow w(\epsilon, a) = w(\epsilon, \eta) + w(\epsilon, \eta' \wedge a)$$
$$\Rightarrow w(\epsilon, \eta' \wedge a) = 0$$
$$\Rightarrow \epsilon < \eta \vee a'.$$

(b) $\eta \vee a'$ covers $a'$: Let $b \in \mathscr{L}$ be such that

$$a' \underset{\neq}{<} b < \eta \vee a'.$$

Since $\mathscr{L}$ is orthomodular, there is a $c \in \mathscr{L}$, $c \neq 0$, $a' \perp c$ such that $c \vee a' = b$. Accordingly $c < a$ and $c = a \wedge (a' \vee c) < a \wedge (\eta \vee a') = \eta$. Hence $c = \eta$ and $b = \eta \vee a'$.

(c) $\eta \vee a' = \epsilon \vee a'$. Indeed, one has

$$a' \underset{\neq}{<} \epsilon \vee a' < \eta \vee a'.$$

■

**Theorem VIII**: If the S.P.S. $(\Sigma_\perp, \mathscr{L}, \sigma)$ satisfies Axioms (1) and (2) and if $\mathscr{L}$ is irreducible (i.e., $\mathscr{L}$ is not the direct union of two lattices[6] and $\mathscr{L} \neq \{0,1\}$, then $\mathscr{L}$ contains at least three orthogonal atoms.

*Proof VIII*: Let $\epsilon \in \mathscr{L}$ be an atom. $\mathscr{L} \neq \{0,1\} \Rightarrow \epsilon' \neq 0$. If $\epsilon'$ is not an atom, then $\epsilon'$ contains at least two orthogonal atoms. We thus only have to prove that $\epsilon'$ is not an atom. Let

$$\mu(\eta, a) = w(\eta, a) \quad \text{if } \eta \neq \epsilon$$
$$\mu(\epsilon, a) = \begin{bmatrix} 1 & \text{if } \epsilon < a \\ 0 & \text{if } \epsilon \perp a \\ \lambda w(\epsilon, a) + (1-\lambda) w(\phi, a) & \text{if not} \end{bmatrix},$$

where $\lambda \in ]0,1[$ and $\phi \neq \epsilon$ is a fixed state. If $\epsilon'$ would be an atom, one would have

$$\epsilon < a \Leftrightarrow a' < \epsilon' \Leftrightarrow a = \epsilon \text{ or } a = 1,$$
$$\epsilon \perp a \Leftrightarrow a < \epsilon' \Leftrightarrow a = \epsilon' \text{ or } a = 0,$$

and it would be straightforward to verify that $\mu$ is a propensity function, hence $\mu = w$. In particular $\mu(\epsilon, \phi) = w(\epsilon, \phi)$.

But this is possible only if $\epsilon \perp \phi$, which implies that

$$\phi = \epsilon' \text{ and } \mathscr{L} = \{0, \epsilon, \epsilon', 1\}.$$

But then $\mathscr{L}$ would be reducible. ■

The proof of Theorem I is now a direct consequence of the above theorems and of Piron's representation theorem.[6]

To conclude this section, let us remark that Axiom (3) is used only to prove the covering law. We conjecture that Axiom (3) is not independent of Axioms (1) and (2). Other open problems are the following. Does a nonseparable Hilbert space admit more than one countably additive propensity function?[31,32] Do the Axioms (1)–(3) imply that $w(\epsilon, \eta) = w(\eta, \epsilon)$ for all states $\epsilon, \eta$? And $w(\epsilon, a \wedge b) = w(\epsilon, a) w(a \wedge (\epsilon \vee a'), b)$ for all compatible (see next section) properties $a$ and $b$? Do Axioms (1) and (2) imply that any irreducible $\mathscr{L}$ is necessarily infinite?[33]

The problem of the most general dynamics compatible with our kinematics is considered in Refs. 34 and 35.

## 6. COMPATIBLE PROPERTIES AND CLASSICAL SYSTEMS

In this section we characterize compatible properties and classical systems in terms of the propensity function $w$. In this section $(\Sigma_1, \mathscr{L}, \sigma)$ denotes a S.P.S. satisfying Axioms (1) and (2). First, we recall some definitions.[4,6]

*Definitions*: (1) Let $a,b \in \mathscr{L}$. $a$ and $b$ are compatible properties iff $a = (a \wedge b) \vee (a \wedge b')$. We use the following notation $a \leftrightarrow b$. (2) A property $c$ is classical iff $c \leftrightarrow a$ for all $a \in \mathscr{L}$. (3) $\mathscr{L}$ is classical iff all properties are classical.

It can be shown that this compatibility relation is symmetric (see, e.g., Ref. 6). In the case of Hilbert space lattices compatibility is equivalent with the usual concept of commuting operators. Different lattice characterizations of compatible properties and classical lattices are given, for instance, in Ref. 6. In particular,

$$\text{(i) } a \leftrightarrow b \Leftrightarrow (a \vee b) \wedge b' < a \wedge b', \tag{2}$$

(ii) $\mathscr{L}$ is classical $\Leftrightarrow$ $\mathscr{L}$ is the power set of the set of states: $\mathscr{L} = P(\Sigma)$.

For completeness we recall without proof the following theorem[6]:

**Theorem**: (1) The set $Z$ of all classical properties of $\mathscr{L}$ is a classical atomic orthomodular sublattice of $\mathscr{L}$.

(2) $\mathscr{L}$ is the direct union of irreducible atomic orthomodular lattices $\mathscr{L}_\alpha$:

$$\mathscr{L} = \bigvee_{\alpha \in \Gamma} \mathscr{L}_\alpha,$$

where $\Gamma$ is the set of atoms of $Z$.

(3) $b = \bigvee_{\alpha \in \Gamma} (b \wedge \alpha)$ for all $b \in \mathscr{L}$,

(4) $\epsilon = \epsilon \wedge \alpha$ for a unique $\alpha \in \Gamma$.

*Corollary IX*: For all $\epsilon \in \Sigma$, $b \in \mathscr{L}$ one has $w(\epsilon, b) = w(\epsilon, b \wedge \alpha)$, where $\alpha \in \Gamma$ is the unique classical atom such that $\epsilon \wedge \alpha = \epsilon$. The proof is immediate, since $\alpha \perp \beta \quad \forall \alpha \neq \beta \in \Gamma$.[7]

The following theorems are the main results of this section.

**Theorem X**: For all $a,b \in \mathscr{L}$ one has $a \leftrightarrow b \Leftrightarrow \forall \epsilon \in \Sigma, \quad w(\epsilon, a \wedge b) + w(\epsilon, a \vee b) = w(\epsilon, a) + w(\epsilon, b)$.

**Theorem XI**: $c$ is a classical property $\Leftrightarrow \forall \epsilon \in \Sigma,\ w(\epsilon, c) \in \{0, 1\}$.

*Corollary XII*: $\mathscr{L}$ is classical $\Leftrightarrow$ the propensity function is dispersion free.[4]

*Proof X*: Assume that $a \leftrightarrow b$. One has

$$\begin{aligned} w(\epsilon, a \vee b) + w(\epsilon, a \wedge b) &= w(\epsilon, (a \wedge b') \vee b) + w(\epsilon, a \wedge b) \\ &= w(\epsilon, a \wedge b') + w(\epsilon, b) + w(\epsilon, a \wedge b) \\ &= w(\epsilon, a) + w(\epsilon, b). \end{aligned}$$

Conversely, assume that the right-hand side of Theorem X holds. We want to prove that $(a \vee b) \wedge b' < a \wedge b'$ [See Eq. (2)]. Let $\epsilon < (a \vee b) \wedge b'$, then $w(\epsilon, b) = 0$ and $w(\epsilon, a) = w(\epsilon, a \vee b) + w(\epsilon, a \wedge b) = 1$. Accordingly $\epsilon \perp b$ and $\epsilon < a$, hence $\epsilon < a \wedge b'$. ■

*Proof XI*: Assume that $c$ is a classical property, and let $\epsilon \in \Sigma$. One has $c \leftrightarrow \epsilon$. But $\epsilon$ is an atom, hence $\epsilon < c$ or $\epsilon \perp c$.

Conversely, assume that $w(\epsilon, c) \in \{0, 1\} \ \forall \epsilon \in \Sigma$, and let $b \in \mathscr{L}$. We want to prove that $(c \vee b) \wedge b' < c \wedge b'$ [See Eq. (2)]. Let $\epsilon < (c \vee b) \wedge b'$. If $\epsilon \perp c$, then $\epsilon \perp c \vee b$ which contradicts $\epsilon < c \vee b$. Consequently $\epsilon < c$, and $\epsilon < c \wedge b'$. ■

Corollary XII follows immediately from Theorem XI. Note that the converse part of Corollary XII is the Jauch–Piron impossibility theorem of noncontextual hidden variables.[36,37]

*Corollary XIII*: $\mathscr{L}$ is classical $\Leftrightarrow$ for all $\epsilon \in \Sigma$, $a \in \mathscr{L}$, $\epsilon < a$, one has $w(\epsilon, a \wedge b) = w(\epsilon, b) \ \forall b \in \mathscr{L}$.

The proof is immediate. Notice the similarity between the right-hand side of Corollary XIII and the classical conditional probabilities. Indeed the former states that the propensity of any property $b$ in a state such that the property $a$ is actual, is equal to the propensity of $a \wedge b$.

## 7. CONCLUSION

The hypothesis that, at any time, the state of the system and the propensities of all properties are completely and uniquely determined by the set of properties actual at that time implies that the states are in one-to-one correspondence with the atoms of the property lattice $\mathscr{L}$. Moreover the latter is canonically orthocomplemented and weakly modular. Let us emphasize that the hypothesis assumes that the system is entirely determined by the set of Einstein's elements of reality,[10] or in other words, that the nondeterministic aspect of the system is entirely determined by its deterministic aspect.

Assuming furthermore that for each state, any property can be ideally tested, implies that $\mathscr{L}$ satisfies the covering law, whence $\mathscr{L}$ is isomorphic to the direct union of Hilbertian space lattices. In this way we recover the usual classical and quantum mechanics (possible with superselection variables) in a common framework. Let us note that the "wave packet reduction" is demonstrated to occur for ideal first-kind tests. It turns out that a system is classical iff the propensity function is dispersion free, i.e., iff only the propensity zero and one occur. Accordingly, the quantum propensities enlarge the concept of classical determinism.

Let us emphasize that our approach is fundamentally concerned with individual systems, which we describe similarly in quantum as in classical physics. In this article we did not consider statistical mechanics. Actually, the description of statistical mixtures of states, or of incomplete knowledge of the state, requires the use of classical probability theory (i.e., measure theory) applied to the state space $\Sigma$.


## ACKNOWLEDGMENT

I wish to thank Professor C. Piron for his encouragement and Dr. N. Giovannini for his careful reading of the manuscript.


*Note added in proof*: Since we submitted this article we noticed that the first axiom is unnecessary. Indeed, a proof similar to the ones of Theorems II and V shows that the second axiom implies that for all states $\epsilon, \eta \in \Sigma$, if $\sigma(\epsilon) = \sigma(\eta)$, then $w(\epsilon, a) = w(\eta, a)$ $\forall a \in \mathscr{L}$. Accordingly, all the results concerning the property lattice $\mathscr{L}$ hold also without Axiom (1). We also noticed that nonseparable Hilbert spaces admit exactly one propensity function [combine condition (5) of a propensity function with Ref. 31].


[1]B. d'Espagnat, *Conceptual Foundations of Quantum Mechanics* (Benjamin,Reading, MA, 1976).
[2]K. R. Popper, British J. Philos. Science **10**, 25 (1959).
[3]K. R. Popper, in *Observation and Interpretation*, edited by S. Körner (Butterworths, London, 1957).
[4]E. G. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison–Wesley, Reading, MA, 1981).
[5]A. M. Gleason, J. Math. Mech. **6**, 885 (1957) (reprinted in Ref. 9).
[6]C. Piron, *Foundations of Quantum Physics* (Benjamin, Reading, MA, 1976).
[7]D. Aerts, J. Math. Phys. **24**, 2441 (1983).
[8]D. Foulis, C. Piron, and C. Randall, Found. Phys. **13**, 813 (1983).
[9]C. A. Hooker, editor, *The Logico-Algebraic Approach to Quantum Mechanics* (Reidel, Boston, 1975).
[10]A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
[11]P. K. Lahti and J. Talja, *Interpretation of quantum mechanics as resulting from the interpretation of probability*, Report Series of the Institute of Theoretical Physics Turku-FTL-R 32 (University of Turku, Finland, 1981).
[12]R. Giere, J. Philos. Logic **5**, 321 (1976); Synthese **40**, 429 (1979).
[13]D. A. Gillies, *An Objective Theory of Probability* (Methuen, London, 1973).
[14]D. H. Mellor, *The Matter of Chance* (Cambridge U. P., Cambridge, England, 1971).
[15]M. Bunge, *Foundations of Physics* (Springer-Verlag, Berlin, 1967).
[16]H. E. Kyberg, British J. Philos. Science **25**, 358 (1974), and references therein.
[17]J. Largeault, *Hasards, Probabilités, Inductions* (Universitae de Toulouse, France, 1979).
[18]D. Bohm, Phys. Rev. **85**, 166 (1952).
[19]L. de Broglie, *Nouvelles persepctive en Microphysique* (Albin-Michel, Paris, 1956); C. R. Acad. Sci. Paris B **277**, 71 (1973).
[20]G. Ludwig, *Foundations of Quantum Mechanics* (Springer-Verlag, Berlin, 1983).
[21]G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972).
[22]P. Suppes, Philos. Science **33**, 14 (1966) (reprinted in Ref. 9).
[23]J. M. Jauch and C. Piron, Helv. Phys. Acta **43**, 842 (1969) (reprinted in Ref. 9).
[24]W. Pauli, *Die Allgemeinen Prinzipien der Wellenmechanik*, Handbuch der Physik, Vol. V (Springer-Verlag, Berlin, 1958).
[25]V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, 1968).
[26]W. J. Wilbur, Trans. Amer. Math. Soc. **233**, 265 (1977).
[27]S. P. Gudder and C. Piron, J. Math. Phys. **12**, 1583 (1971).
[28]M. J. Maczynski, J. Math. Phys. **14**, 1469 (1973).
[29]I. Amemiya and H. Araki, Publ. Research Inst. Math. Sci. Kyoto Univ. A **2**, 423 (1967).
[30]J.C.T. Pool, Commun. Math. Phys. **9**, 118 (1968) (reprinted in Ref. 9).
[31]S. P. Gudder, Int. J. Theor. Phys. **6**, 369 (1972).
[32]M. Eilers and E. Horst, Int. J. Theor. Phys. **13**, 419 (1975).
[33]G. T. Rúttimann, J. Math. Phys. **18**, 189 (1977).
[34]W. Daniel, Helv. Phys. Acta **55**, 330 (1982).
[35]N. Gisin, J. Math. Phys. **24**, 1779 (1983).
[36]J. M. Jauch and C. Piron, Helv. Phys. Act **36**, 827 (1963).
[37]F. J. Belinfante, *A Survey of Hidden-Variables Theories* (Pergamon, Oxford, 1973).

# Bound on the $N$th order term of the partition function of the massive Schwinger model

M. P. Fry
*School of Mathematics, Trinity College, Dublin 2, Ireland*

(Received 22 June 1983; accepted for publication 16 December 1983)

An upper bound on the vacuum-to-vacuum amplitude of the Schwinger model with massive fermions is obtained.

PACS numbers: 03.70. + k, 11.15.Bt

## 1. INTRODUCTION

A question of interest since the early 1950's has been the following: To what extent does renormalized perturbation theory exhaust the information content of a relativistic quantum field theory? Stated differently, how much information is lost in the Feynman series? For the $\phi^4$ theory in two[1] and three[2] dimensions ($\phi^4_{2,3}$) and the Yukawa interaction in two dimensions ($Y_2$)[3,4] the answer is that none is lost in the series for their Euclidean Green's functions (or Schwinger functions), and, for $\phi^4_2$, none is lost as well in the series for its physical mass and two-body $S$-matrix.[5] These theories have sufficient analyticity in the coupling constant and sufficiently slow growth in large orders to allow the unique recovery of these quantities by Borel summation.

There are examples where this is not the case. Massless super-renormalizable field theories are known to contain nonanalytic terms in the coupling constant that forbid expansions in its powers.[6] In QCD in four dimensions with massless quarks, 't Hooft[7] has argued that the correlation function $G(p^2)$ of the color-singlet operator $\bar{q}q$ cannot be uniquely summed if it has the usually assumed analyticity properties in the $p^2$-plane with multiparticle singularities extending to infinity along the cut. Field theories that have a nontrivial ultraviolet fixed point may also impose restrictions on their unique summability.[8]

Presumably field theories exist whose associated Feynman series are not even asymptotic. It is straightforward to construct physically reasonable potentials in quantum mechanics whose ground-state energies have associated Rayleigh–Schrödinger series that are not asymptotic, even though each term is well defined.[9] These potentials have the general form

$$V(x) = \sum_{n=0}^{N} g^n V_n(x),$$

where the $V_n$ are polynomials in $x$, and $g$ is the coupling constant. Since this is just how the nonderivative terms in the Lagrangian of a large class of boson field theories would look after Wick ordering and renormalization, it is not unthinkable that some of them have nonasymptotic Feynman series. In this connection we note the preliminary result of Fröhlich[10] that there is no family of $\phi^4$ theories in four dimensions to which renormalized perturbation theory is asymptotic.

It can be generally said that the faster the coefficients of a Feynman series associated with a field theory grow with order, the more analyticity is required about the origin of the complex coupling constant plane to uniquely reconstruct the quantity of interest from the series. Typically, if the expansion coefficients grow like $(n!)^\lambda$, analyticity in a region about the origin with opening angle $\lambda\pi/2$ is required.[3] Therefore, the large-order behavior of a field theory, by itself, can only be an indication of the odds favoring its unique summability. There are simple examples illustrating the folly of inferring anything more than this.[11]

Table I summarizes current knowledge of the large-order growth of several field theories. To facilitate comparison, the Feynman series for the Euclidean vacuum-to-vacuum amplitude $Z$ (hereafter called the partition function) has been singled out in two[1,12–17] and three dimensions[2,13–15]; in four dimensions[18] the Schwinger functions in order $n$, denoted by $S_n$, are the obvious quantities to compare. The quantity $K$ is a sufficiently large $n$-independent constant. The result for two-dimensional quantum electrodynamics with massive electrons (hereafter called $QED_2$) will be derived here. A related model, the massive Thirring–Schwinger model,[19] is also sometimes referred to as $QED_2$. The charge-0 sector of this model and the massive sine-Gordon theory are equivalent. The authors of Ref. 19 showed that the Feynman series in the coupling constant for the Schwinger functions of the latter theory converge for sufficiently large electric charge.

The decreasing rate of growth of the expansion coefficients as the physically relevant field theories are approached in two and three dimensions is striking. For $\phi^4_2$, all graphs in a fixed order have the same relative sign, so that the growth of the $Z_n$'s is due to the growth in the number of graphs. With the addition of fermions, graphs with an even or odd number of fermion loops differ by an overall sign that is presumably responsible for the sharply reduced upper bound on the $Z_n$'s for the $Y_2$ theory. A (non-) Abelian local gauge symmetry will introduce correlations among graphs in a fixed order, and this may contribute to a further slow down in the growth of the $Z_n$'s. This is illustrated by the Schwinger model ($QED_2$ with massless electrons) whose partition function actually has a convergent power-series expansion.[17] It will be indicated below why it is expected that the bound on the $Z_n$'s in $QED_2$ can be improved to $|Z_n| \leqslant K^n$, as for the Schwinger model.

A further indication of the trend toward better behaved power-series expansions with increasing symmetry is given by conformal covariant QED. This is $QED_4$ in a special gauge with massless electrons and no electron loop subgraphs. The conformal electron propagator turns out to be analytic about the origin of the coupling constant plane.[20]

In four dimensions the subtractions due to renormalization may further ameliorate the growth in large orders. The remarkable bounds of de Calan and Rivasseau[18] on the

 0022-2488/84/072266-10$02.50  

TABLE I. Upper bounds on the Euclidean vacuum-to-vacuum amplitude $Z_n$ and the Schwinger functions $S_n$ in order $n$. The bounds on $Z_n$ in the last column are for QED with massless and massive electrons. The bound for $Y_2$ is also claimed in footnote 32 of Ref. 3. The $S_n$'s for $\phi^4_{2,3}$ and $Y_{2,3}$ have the same dominant bounds as the $Z_n$'s.

| Dimension | $\phi^4$ | Yukawa | QCD,SU(2)$_L\times$U(1), QED |
|---|---|---|---|
| 2 | L $\lvert Z_n\rvert \leqslant K^n n!$ (Refs. 1,12–14) | $\lvert Z_n\rvert \leqslant (K\log n)^n$ (Refs. 15, 16) | $\lvert Z_n\rvert \leqslant K^n$, $m=0$ (Ref. 17); $(K\log n)^n$, $m>0$ |
| 3 | $\lvert Z_n\rvert \leqslant K^n n!$ (Refs. 2,13,14) | $\lvert Z_n\rvert \leqslant K^n (n)!^{1/3}$ (Ref. 15) | $\lvert Z_n\rvert \leqslant ?$ |
| 4 | $\lvert S_n\rvert \leqslant K^n n!$ (Ref. 18) | $\lvert S_n\rvert \leqslant ?$ | $\lvert S_n\rvert \leqslant ?$ |

Schwinger functions of $\phi^4_4$ in order $n$ are encouraging.

As Table I indicates, present knowledge of the large-order behavior of (non-) Abelian gauge field theories that include fermions is deficient. For the simplest case, QED, progress in any number of dimensions has been barred by a lack of knowledge of the order of growth of the renormalized fermion determinant, $\det_{\rm ren}(1-eS\not{A})$, obtained by integrating over the fermion degrees of freedom. Here $A_\mu$ is the vector potential, $S$ is the free electron propagator, and $e$ is the coupling constant. In fact, $\det_{\rm ren}$ is just exp(single fermion loops—counter terms). Ideally one would like to prove that $\det_{\rm ren}$ is an entire function of $e$ and, having established this, determine its order and type, assuming that $A_\mu$ is a Gaussian random field. The desirability of this will become evident in Sec. 3. For the Schwinger model, the solution is well known: $\det_{\rm ren}$ is Gaussian in $A_\mu$.[17] This simple result follows from the fact that $\langle 0| j_{\mu 1}(x_1)\dots j_{\mu n}(x_n)|0\rangle = 0$ for $n\geqslant 4$ and zero electron mass.[17,21] For nonzero electron mass this is no longer true, and the growth properties of $\det_{\rm ren}$ have to be reestablished.

Ito[22] has examined this case and has found that $\det_{\rm ren}$ is Gaussian dominated for real $A_\mu \in L_2\cap L_q (q>2)$ in $QED_2$. Since his upper bound is not almost everywhere finite with respect to the functional measure assocated with $A_\mu$, it cannot be used here to study the large-order behavior of $QED_2$. A new bound is obtained in Sec. 3.

For $QED_4$ some results on the order of growth of $\det_{\rm ren}$ that neglect charge renormalization effects are known for special field configurations and massless electrons.[23,24] It should be stated that charge renormalization is absent by definition in the model studied in Ref. 23. For $QCD_2$ and $QCD_4$ it is known that massive fermions are essential for a satisfactory definition of $\det_{\rm ren}$.[25] Nothing is yet known about their orders of growth.

It is apparent from the foregoing that knowledge of the large-order behavior of $QED_2$ would be desirable before attacking other (non-) Abelian gauge field theories. Attention is focused on its gauge-invariant sectors as these are the physically relevant ones, and because the infrared divergences present in its charged sectors are absent. The large-order behavior of the partition function is singled out because it is the simplest gauge-invariant quantity in $QED_2$. On the basis of previous studies cited in Table I, e.g., Ref. 3, the Schwinger functions in the charge-0 sector are expected to have the same dominant large-order behavior.

The final result, (4.45), is

$$|Z_{2n}| \leqslant [C\ln(mn/\mu)]^{2n}, \tag{1.1}$$

where $m$ is the bare electron mass, $\mu(<m)$ is an infrared cutoff, and $C$ is a sufficiently large constant. The presence of $\mu$ in (1.1) is a result of the upper bound on $\det_{\rm ren}$ in terms of trace ideal norms obtained in Sec. 3. Such norms ruin gauge invariance by putting fermion propagators and vertices in the wrong order in closed loops. The possibility remains that a better bound can be obtained that will permit the limit $\mu = 0$ to be taken. Referring to (1.1), it may then happen that when $\ln(m/\mu)$ drops out, so will the ln $n$ term, yielding $|Z_{2n}| \leqslant C^{2n}$ as for the Schwinger model.

## 2. DEFINITION OF THE PARTITION FUNCTION

Our starting point is the following expression for the partition function obtained by formally integrating out the fermions in the vacuum-to-vacuum amplitude:

$$Z(\Lambda) = \int d\mu(A_\mu)\det{}_{\rm ren}(1-\lambda K), \tag{2.1}$$

where $\det_{\rm ren}$ denotes a suitably renormalized Fredholm determinant that will be defined below. The integral operator $K$ is

$$K = (P^2+m^2)^{1/4} S(x-y)\not{A}_\Lambda(y) g(y)(P^2+m^2)^{-1/4}, \tag{2.2}$$

where $iP_\mu = \partial_\mu$,

$$S = \int \frac{d^2p}{(2\pi)^2} e^{ipx}\frac{m-\not{p}}{p^2+m^2} \tag{2.3}$$

is the two-point Schwinger function for the electron with bare mass $m>0$, $g\in C_0^\infty$ is a space-time cutoff, and $\not{A}_\Lambda = \not{A} * h_\Lambda$. For $A_\mu \in \mathscr{S}'$, the space of tempered distributions, then $\not{A}_\Lambda \in C^\infty$ if the ultraviolet cutoff function $h_\Lambda \in C^\infty$. Our choice for $h_\Lambda$ is

$$h_\Lambda(x) = \int \frac{d^2p}{(2\pi)^2} e^{ipx}\hat{h}_\Lambda(p), \tag{2.4}$$

with $\hat{h}_\Lambda(p)\in C_0^\infty$; $\hat{h}_\Lambda(p) = 1$ for $p^2 \leqslant \Lambda^2$; $\hat{h}_\Lambda(p) = 0$ for $p^2 \geqslant (\Lambda+m)^2$ and $\Lambda>0$. The choice of $\Lambda + m$ as the cutoff point is arbitrary.

The Gaussian measure $d\mu$ for $A_\mu$ is chosen to have mean zero and covariance

$$\int d\mu\, A_{\mu,\Lambda}(x) A_{\nu,\Lambda}(y) = D^\Lambda_{\mu\nu}(x-y), \tag{2.5}$$

whose Fourier transform is

$$\hat{D}^\Lambda_{\mu\nu}(k) = \left(\delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2+\mu^2}\right)\frac{\hat{h}^2_\Lambda(k)}{k^2+\mu^2}, \tag{2.6}$$

where $\mu^2>0$ is an infrared cutoff.

The electric charge is denoted by $\lambda\in\mathbb{C}$ to avoid confusion with the exponential function.

Our conventions for the $\gamma$ matrices are

$$\{\gamma_\mu,\gamma_\nu\} = -2\delta_{\mu\nu} \quad (\mu = 0,1),$$
$$\gamma_\mu{}^* = -\gamma_\mu,$$

and, naturally $\not p = p_0\gamma_0 + p_1\gamma_1$.

A word on the choice of $K$ in (2.2): We work on the Hilbert space $L^2(\mathbb{R}^2, d^2x; \mathbb{C}^2)$ of two-component square-integrable functions on $\mathbb{R}^2$. The $K$ in (2.2) differs from $S\not A g$ on $L^2(\mathbb{R}^2, \sqrt{p^2+m^2}\, d^2p, \mathbb{C}^2)$. But the two are equivalent given the natural unitary equivalence of $L^2(\mathbb{R}^2, d^2x)$ and $L^2(\mathbb{R}^2, \sqrt{p^2+m^2}\, d^2p)$. Our choice of Hilbert space and $K$ is motivated with the view of taking the limit $\Lambda = \infty$ at the end of our calculation.

We now turn to the definition of the renormalized determinant, $\det_{\rm ren}$. The operator $K$ is a compact operator in the trace ideal $\mathscr{C}_{2+\epsilon}$, $\epsilon > 0$. This is an easy consequence of a proposition stated by Seiler and Simon.[26] The trace ideal $\mathscr{C}_n (1 \leqslant n < \infty)$ is defined for compact operators $A$ with $\|A\|_n^n \equiv \mathrm{Tr}(A^*A)^{n/2} < \infty$. Then the determinant

$\det_3(1-\lambda K)$, defined by

$$\det_3(1-\lambda K) = \det\left[(1-\lambda K)e^{\lambda K + (1/2)\lambda^2 K^2}\right], \tag{2.7}$$

is an entire function of $\lambda$ of at most order 3:

$$\det_3(1-\lambda K) = \prod_{i=1}^{\infty}\left[(1-\lambda\lambda_i)e^{\lambda\lambda_i + 1/2(\lambda\lambda_i)^2}\right], \tag{2.8}$$

where $\lambda_1, \ldots$ are the eigenvalues of $K \in \mathscr{C}_3$.[27]

The graph in Fig. 1a is not present in the loop expansion of (2.7). It is only conditionally convergent and may or may not contain a current nonconserving piece, depending on how one regulates. Its offspring obtained by integrating over $A_\mu$, Fig. 1b, has an ultraviolet logarithmic divergence that must be subtracted out. Therefore, define the Wick-ordered quantity

$$\mathrm{Tr}{:}K^2{:} \equiv \int d^2x\, d^2y\, g(x)\rho_{\mu\nu}(x-y)g(y) \times \left[A_{\mu,\Lambda}(x)A_{\nu,\Lambda}(y) - D^{\Lambda}_{\mu\nu}(x-y)\right], \tag{2.9}$$

where $\rho_{\mu\nu}$ is the transverse piece of $\mathrm{tr}(S(x-y)\gamma_\mu S(y-x)\gamma_\nu)$, whose Fourier transform is

$$\hat\rho_{\mu\nu}(q) = \frac{1}{\pi}\left(\delta_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right)\left[1 - \frac{4m^2}{q(q^2+4m^2)^{1/2}} \times \operatorname{arctanh}\left(\frac{q}{(q^2+m^2)^{1/2}}\right)\right]. \tag{2.10}$$

Summation is implied over repeated polarization indices.

We can now define

$$\det_{\rm ren}(1-\lambda K) = e^{-(\lambda^2/2)\mathrm{Tr}{:}K^2{:}}\det_3(1-\lambda K), \tag{2.11}$$

which is depicted graphically in Fig. 2. All loops with an odd number of external photon lines vanish ($C$-invariance) except for the tadpole graph in Fig. 1c which we dropped altogether. If $\det_{\rm ren}$ is expanded in a power series in $\lambda$, inserted in

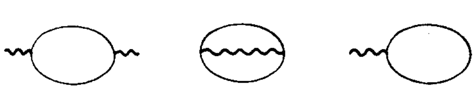


FIG. 1.

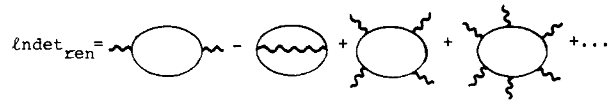


FIG. 2.

(2.1), integrated, and the limit $\Lambda = \infty$ taken term by term in the series

$$Z(\Lambda) = \sum_{n=0}^{\infty} Z_{2n}(\Lambda)\lambda^{2n}, \tag{2.12}$$

the renormalized perturbation expansion for the partition function is obtained. Moreover, if $\delta^2(0)$ is interpreted as a large but finite space-time volume—which will not be done here—and the volume cutoff $g$ is replaced by unity, then the extra powers of momentum in the numerators of graphs obtained by gauge invariance allow the removal of the infrared cutoff $\mu$.

To conclude this section note that

$$|Z_{2n}(\Lambda)| \leqslant \frac{1}{(2n)!}\int d\mu \left|\frac{d^{2n}}{d\lambda^{2n}}\det_{\rm ren}\right|_{\lambda=0} \tag{2.13}$$

## 3. DETERMINANT INEQUALITIES

We proceed to prove the following result:

**Theorem 3.1**:

$$\frac{1}{(2n)!}\left|\frac{d^{2n}}{d\lambda^{2n}}\det_{\rm ren}\right|_{\lambda=0} \leqslant \frac{1}{4}\left(\frac{4e}{n}\right)^n\left(\|L\|_2^{2n} + \alpha^n\|HL\|_1^n \times \beta^n n^{n\epsilon/(2+\epsilon)}\|H\|_{2+\epsilon}^{2n} + \frac{|\mathrm{Tr}{:}K^2{:}|^n}{2^n}\right), \tag{3.1}$$

for $n = 1,2,\ldots, 0<\epsilon\leqslant 1$ and $\alpha,\beta$ sufficiently large. The operator $K$ has been split into low and high momentum parts

$$K = L + H, \tag{3.2}$$

where

$$L = (P^2+m^2)^{1/4}S^{<}\not A_\Lambda g(P^2+m^2)^{-1/4}, \tag{3.3}$$

$$S^{<}(x) = \int_{|p|<\zeta m}\frac{d^2p}{(2\pi)^2}e^{ipx}\frac{m-\not p}{p^2+m^2}, \tag{3.4}$$

and $\zeta \geqslant 0$. As in the case of $K$, a proposition of Seiler and Simon[26] can be used to show that $H \in \mathscr{C}_{2+\epsilon}$, $0<\epsilon<2$. Using the same procedure as Renouard[3] one may easily show that $L \in \mathscr{C}_1$ for $\zeta > 0$. Therefore,[28] $HL \in \mathscr{C}_1$. We will prove Theorem 3.1 by first establishing some relevant lemmas.

*Lemma 3.2*: Let

$$H \in \mathscr{C}_{2+\epsilon},\ \epsilon > 0,\ L \in \mathscr{C}_1.$$

Then

$$\det_3(1+L+H) = \det_3(1+H)\det(1+L) \times \det(1-(1+L)^{-1}(1+H)^{-1}HL) \times \exp(-\mathrm{Tr}\,L + \tfrac{1}{2}\mathrm{Tr}\,L^2 + \mathrm{Tr}\,(HL)). \tag{3.5}$$

*Proof*: It is sufficient to give the proof for $L, H \in \mathscr{C}_1$ since $\det_3(1+A)$ is a continuous function of $A$ on $\mathscr{C}_p$, $1 \leqslant p \leqslant 3$.[29] Then

$$\det_3(1+L+H) = \det_3(1+H)\det(1+L)\det(1+D) \times \exp(-\mathrm{Tr}\,L + \tfrac{1}{2}\mathrm{Tr}\,L^2 + \mathrm{Tr}(HL)),$$

where

$$D = -(1+L)^{-1}(1+H)^{-1}HL, \tag{3.6}$$

since

$$1 + L + H = (1+H)(1+L)(1+D).$$

*Lemma 3.3*: $\operatorname{Tr} L = 0$.

*Proof*: Since $L \in \mathscr{C}_1$, then[30]

$$\operatorname{Tr} L = \int d^2x\, d^2y\, d^2z\, \operatorname{tr}[D_{-1/4}(x-y) S^{<}(y-z) \times (A_\Lambda g)(z) D_{1/4}(z-x)],$$

where

$$D_z(x) = \int \frac{d^2p}{(2\pi)^2} \frac{e^{ipx}}{(p^2+m^2)^z}. \tag{3.7}$$

Hence

$$\operatorname{Tr} L = 2\widehat{A_{\mu,\Lambda} g}(0) \int_{|p| \leqslant \zeta m} \frac{d^2p}{(2\pi)^2} \frac{p^\mu}{p^2+m^2} = 0. \tag{3.8}$$

*Lemma 3.4*: For $H \in \mathscr{C}_{2+\epsilon}$, $\epsilon > 0$, $L \in \mathscr{C}_1$,

$$|\det{}_3(1+H)\det(1+L)\det(1-(1+L)^{-1}(1+H)^{-1}HL)|$$

$$\leqslant \sum_{n=0}^{\infty} \|\det{}_3(1+H)\Lambda^n(1+H)^{-1}\|\, \|\det(1+L)\Lambda^n(1+L)^{-1}\|\, \|HL\|_1^n/n!. \tag{3.9}$$

*Proof*: By the expansion $\det(1+D) = \Sigma_{n=0}^{\infty} \operatorname{Tr}(\Lambda^n(D))$, with $D$ given by (3.6), and the fact that $\Lambda^n(AB) = \Lambda^n(A)\Lambda^n(B)$, we get

$$|\det{}_3(1+H)\det(1+L)\det(1+D)|$$

$$= \left| \sum_{n=0}^{\infty} (-1)^n \det{}_3(1+H)\det(1+L)\operatorname{Tr}(\Lambda^n(1+L)^{-1}\Lambda^n(1+H)^{-1}\Lambda^n(HL)) \right|$$

$$\leqslant \sum_{n=0}^{\infty} \|\det{}_3(1+H)\Lambda^n(1+H)^{-1}\|\, \|\det(1+L)\Lambda^n(1+L)^{-1}\|\, \|\Lambda^n(HL)\|_1,$$

which gives (3.9) using[27] $\|\Lambda^n(HL)\|_1 \leqslant \|HL\|_1^n/n!$.

*Lemma 3.5*: For $L \in \mathscr{C}_1$ and $\operatorname{Tr} L = 0$,

$$\|\det(1+L)\Lambda^n(1+L)^{-1}\|^2 \leqslant e^n e^{\|L\|_2^2}. \tag{3.10}$$

*Proof*: For $L \in \mathscr{C}_1$ we have by a result of Simon,[27]

$$\|\det(1+L)\Lambda^n(1+L)^{-1}\|^2 \leqslant e^n \exp(2\operatorname{Re}(\operatorname{Tr} L) + \|L\|_2^2),$$

from which (3.10) follows with $\operatorname{Tr} L = 0$.

*Lemma 3.6*: For $H \in \mathscr{C}_{2+\epsilon}$, $0 < \epsilon \leqslant 1$,

$$\|\det{}_3(1+H)\Lambda^n(1+H)^{-1}\|^2 \leqslant C^n \exp(\Gamma \|H\|_{2+\epsilon}^{2+\epsilon}), \tag{3.11}$$

for $C$ and $\Gamma$ sufficiently large.

*Proof*: It is sufficient to give the proof for $H \in \mathscr{C}_1$. Then

$$\|\det{}_3(1+H)\Lambda^n(1+H)^{-1}\|^2 = \|\det(1+0_H)\Lambda^n(1+0_H)^{-1}\| \times \exp[-2\operatorname{Re}(\operatorname{Tr} H) + \operatorname{Re}(\operatorname{Tr} H^2)],$$

where $0_H = H + H^* + H^*H$. Let $-1 \leqslant \alpha_1 \leqslant \alpha_2 \leqslant \cdots$ be the eigenvalues of $0_H$ and $\lambda_i$ the eigenvalues of $H$ with the $\beta_i = 2\operatorname{Re}\lambda_i + |\lambda_i|^2$ ordered so that $-1 \leqslant \beta_1 \leqslant \beta_2 \leqslant \cdots$. Using $\det(1+H) = \Pi_{i=1}^{\infty}(1+\lambda_i)$ it follows that

$$\|\det(1+0_H)\Lambda^n(1+0_H)^{-1}\| = \prod_{i=n+1}^{\infty}(1+\alpha_i) = \prod_{i=1}^{n} \frac{1}{1+\alpha_i} \prod_{i=1}^{\infty}(1+\beta_i). \tag{3.12}$$

Since the first equality is finite we conclude that the multiplicities of the eigenvalues $\alpha_i$ with $\alpha_i = -1$ and of the eigenvalues $\lambda_i$ with $\beta_i = -1$ are equal. Let $k \geqslant 0$ denote this multiplicity. The left-hand side of (3.12) is nonvanishing when $n \geqslant k$ and is equal to

$$\prod_{i=k+1}^{n} \frac{1}{1+\alpha_i} \prod_{i=k+1}^{\infty}(1+\beta_i),$$

where $-1 < \alpha_{k+1} \leqslant \alpha_{k+2} \leqslant \cdots$, $-1 < \beta_{k+1} \leqslant \beta_{k+2} \leqslant \cdots$. Choose a constant $C\,(\geqslant 1)$ sufficiently large so that

$$1 + \alpha_i \geqslant (1+\beta_i)/C, \quad i \geqslant k+1.$$

Then

$$\|\det(1+0_H)\Lambda^n(1+0_H)^{-1}\| \leqslant C^{n-k} \prod_{i=n+1}^{\infty}(1+\beta_i),$$

and

$$\|\det{}_3(1+H)\Lambda^n(1+H)^{-1}\|^2 \leqslant C^n \prod_{i=n+1}^{\infty} [(1+2\operatorname{Re}\lambda_i + |\lambda_i|^2) \times \exp(-2\operatorname{Re}\lambda_i + \operatorname{Re}\lambda_i^2)] \times \exp\left[\sum_{i=1}^{n} (\operatorname{Re}\lambda_i^2 - 2\operatorname{Re}\lambda_i)\right]. \tag{3.13}$$

We note that there exists a constant $\Gamma_1$ such that

$$(1 + 2\operatorname{Re}\lambda + |\lambda|^2)\exp(-2\operatorname{Re}\lambda + \operatorname{Re}\lambda^2) \leqslant \exp(\Gamma_1 |\lambda|^{2+\epsilon}),$$

where $0 < \epsilon \leqslant 1$. This is obvious for $|\lambda| > \delta$ for any $\delta$, while for $|\lambda|$ small the left-hand side is $1 + O(|\lambda|^3)$. Then

$$\prod_{i=n+1}^{\infty} [(1+2\operatorname{Re}\lambda_i + |\lambda_i|^2)\exp(-2\operatorname{Re}\lambda_i + \operatorname{Re}\lambda_i^2)] \leqslant \exp\left(\Gamma_1 \sum_{i=n+1}^{\infty} |\lambda_i|^{2+\epsilon}\right). \tag{3.14}$$

Finally, using the inequality

$$\exp(\operatorname{Re}\lambda^2 - 2\operatorname{Re}\lambda) \leqslant 2\exp(\Gamma_2 |\lambda|^{2+\epsilon}), \tag{3.15}$$

for $\epsilon > 0$ and $\Gamma_2$ sufficiently large we get from (3.13)–(3.15)

$$\|\det_3(1+H)\Lambda^n(1+H)^{-1}\|^2$$
$$\leqslant 2^n C^n \exp\Big(\Gamma_1 \sum_{i=n+1}^{\infty} |\lambda_i|^{2+\epsilon} + \Gamma_2 \sum_{i=1}^{n} |\lambda_i|^{2+\epsilon}\Big)$$
$$\leqslant 2^n C^n \exp\Big(\sum_{i=1}^{\infty} |\lambda_i|^{2+\epsilon}\Big)$$
$$\leqslant 2^n C^n \exp(\Gamma \|H\|_{2+\epsilon}^{2+\epsilon}),$$

where $0<\epsilon\leqslant 1$, $\Gamma = \max(\Gamma_1,\Gamma_2)$. For the last inequality we used[27] $\Sigma_{i=1}^{\infty} |\lambda_i(H)|^{2+\epsilon} \leqslant \|H\|_{2+\epsilon}^{2+\epsilon}$, which proves the lemma.

Combining (3.5) and (3.8)–(3.11) we obtain

$$|\det_3(1+L+H)| \leqslant |\exp(\tfrac{1}{2}\operatorname{Tr} L^2 + \operatorname{Tr}(HL))|$$
$$\times \exp((\Gamma/2)\|H\|_{2+\epsilon}^{2+\epsilon} + \tfrac{1}{2}\|L\|_2^2)$$
$$\times \sum_{n=0}^{\infty} \|HL\|_1^n (Ce)^{n/2}/n!$$
$$\leqslant \exp(\|L\|_2^2 + (1+\sqrt{Ce})\|HL\|_1 + (\Gamma/2)\|H\|_{2+\epsilon}^{2+\epsilon}). \tag{3.16}$$

By a Cauchy estimate and the definition (2.11) we get from(3.16)

$$\frac{1}{(2n)!}\left|\frac{d^{2n}}{d\lambda^{2n}}\det_{\text{ren}}\right|_{\lambda=0}$$
$$\leqslant \sup_{\phi} |\det_3(1-\lambda K)e^{-\lambda^2/2\,\operatorname{Tr}:K^2:}|/|\lambda|^{2n}$$
$$\leqslant \exp(a|\lambda|^2 + b|\lambda|^{2+\epsilon} - 2n\ln|\lambda|), \tag{3.17}$$

where

$$\lambda = |\lambda| e^{i\phi},$$
$$a = \|L\|_2^2 + (1+\sqrt{Ce})\|HL\|_1 + \tfrac{1}{2}|\operatorname{Tr}:K^2:| \tag{3.18}$$
$$b = (\Gamma/2)\|H\|_{2+\epsilon}^{2+\epsilon}. \tag{3.19}$$

Let $M$ denote the right-hand side of (3.17). Then for $n>0$,

$$\inf M \leqslant (ae/n + e((2+\epsilon)b/2n)^{2/(2+\epsilon)})^n. \tag{3.20}$$

*Proof*: Since $d^2M/d|\lambda|^2>0$, $M$ has only one minimum at $|\lambda| = r_0$, where $2ar_0^2 + (2+\epsilon)br_0^{2+\epsilon} - 2n = 0$. Thus $M(r_0) \leqslant (e/r_0^2)^n$. Since $r_0^{\epsilon} \leqslant (2n/b(2+\epsilon))^{\epsilon/(2+\epsilon)}$ then

$$\frac{1}{r_0^2} = \frac{a}{n} + \frac{(2+\epsilon)br_0^{\epsilon}}{2n}$$
$$\leqslant \frac{a}{n} + \left(\frac{(2+\epsilon)b}{2n}\right)^{2/(2+\epsilon)},$$

for which (3.20) follows.

Finally, since $a$ in (3.18) and (3.20) is composed of three terms, we apply the inequality

$$(a_1+a_2+a_3+a_4)^n \leqslant 4^{n-1}(a_1^n + a_2^n + a_3^n + a_4^n), \tag{3.21}$$

for $a_1,\ldots,a_4 \geqslant 0$, $n = 1,2,\ldots$to (3.20). From (3.17)–(3.21) we get

$$\frac{1}{(2n)!}\left|\frac{d^{2n}}{d\lambda^{2n}}\det_{\text{ren}}\right|_{\lambda=0}$$
$$\leqslant \frac{1}{4}\Big(\frac{4e}{n}\Big)^n \big[\|L\|_2^{2n} + (1+\sqrt{Ce})^n \|HL\|_1^n$$
$$+ ((2+\epsilon)\Gamma/4)^{2n/(2+\epsilon)} n^{\epsilon n/(2+\epsilon)} \|H\|_{2+\epsilon}^{2n}$$
$$+ |\operatorname{Tr}:K^2:|^n/2^n\big],$$

from which Theorem 3.1 follows.

## 4. BOUNDS

We now proceed to place bounds on the integrals arising from the application of Theorem 3.1 to (2.13). Our main tool in this section is the hypercontractive inequality.[31] It implies that if $Q$ is a polynomial in $A_{\mu,\Lambda}$ of degree $n$ and $p \geqslant 1$ then

$$\int d\mu |Q|^{2p} \leqslant (2p-1)^{np}\Big(\int d\mu |Q|^2\Big)^p. \tag{4.1}$$

### 4.1 $\int\mu\|L\|_2^{2n}$

By the hypercontractive inequality

$$\int d\mu \|L\|_2^{2n} \leqslant (n-1)^n \Big(\int d\mu \|L\|_2^4\Big)^{n/2}, \tag{4.2}$$

for $n \geqslant 2$. Since $L \in \mathscr{C}_2$ we get from (3.3)

$$\|L\|_2^2 = 2\int d^2x\, d^2y (A_{\mu,\Lambda} g)(x) D_{1/2}$$
$$\times (x-y) D_{1/2}^{<}(x-y)(A_{\mu,\Lambda} g)(y), \tag{4.3}$$

where $D_{1/2}$ is given by (3.7) and

$$D_z^{<}(x) = \int_{|p|<m\zeta} \frac{d^2p}{(2\pi)^2} \frac{e^{ipx}}{(p^2+m^2)^z}. \tag{4.4}$$

From (4.3)

$$\int d\mu \|L\|_2^4 = 4\int d^2x_1 \ldots d^2y_2\, g(x_1)\ldots g(y_2) D_{1/2}(x_1-y_1)$$
$$\times D_{1/2}^{<}(x_1-y_1) D_{1/2}(x_2-y_2)$$
$$\times D_{1/2}^{<}(x_2-y_2)[D_{\mu\mu}^{\Lambda}(x_1-y_1) D_{\nu\nu}^{\Lambda}(x_2-y_2)$$
$$+ D_{\mu\nu}^{\Lambda}(x_1-x_2) D_{\mu\nu}^{\Lambda}(y_1-y_2) + (x_2 \leftrightarrow y_2)]. \tag{4.5}$$

The topologies of the Feynman diagrams corresponding to the right-hand side of (4.5) are depicted in Fig. 3. Since these are finite by power counting in the limit $\Lambda = \infty$ we get from (4.2) and (4.5)

$$\lim_{\Lambda\to\infty} \int d\mu \|L\|_2^{2n} \leqslant (n-1)^n (I_1^2 + I_2)^{n/2}, \tag{4.6}$$

where

$$I_1 = \frac{2}{(2\pi)^6} \int \prod_{i=1}^{3} d^2k_i |\hat{g}(k_1)|^2 \hat{D}_{1/2}(k_1+k_2+k_3)$$
$$\times \hat{D}_{1/2}^{<}(k_2) \hat{D}_{\mu\mu}(k_3), \tag{4.7}$$

$$I_2 = \frac{8}{(2\pi)^{12}} \int \prod_{i=1}^{6} d^2k_i \hat{g}(k_1)\hat{g}(k_2)\hat{g}(k_3)\hat{g}(-k_1-k_2-k_3)$$
$$\times \hat{D}_{1/2}^{<}(k_4)\hat{D}_{1/2}^{<}(k_5)\hat{D}_{1/2}(k_1+k_4+k_6)$$
$$\times \hat{D}_{1/2}(k_5+k_6-k_2)\,\hat{D}_{\mu\nu}(k_6)\hat{D}_{\mu\nu}(k_1+k_3+k_6), \tag{4.8}$$

and

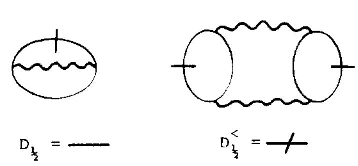

FIG. 3.

$$\hat{D}_{\mu\nu}(k) = \frac{\delta_{\mu\nu} - k_\mu k_\nu/(k^2+\mu^2)}{k^2+\mu^2}. \qquad (4.9)$$

Referring to the Appendix, (A2), (A3), and (A5) give

$$I_1 \leqslant \frac{2^{5/2}}{m(2\pi)^6}\int d^2k\, |\hat{g}(k)|^2 \sqrt{k^2+m^2}\Big\{2\pi^2\Big[\ln\Big(\frac{m\zeta}{\mu}\Big)\Big]^2 + \pi^4 \ln\Big(\frac{2m\zeta}{\mu}+1\Big) + \frac{\pi^4}{6}\Big\}, \qquad (4.10)$$

while (A6), (A7), and (A9) give

$$|I_2| \leqslant \frac{80}{m^2\mu^4(2\pi)^9}\int \prod_{i=1}^{3} d^2k_i\, |\hat{g}(k_1)\hat{g}(k_2)\hat{g}(k_3) \times \hat{g}(k_1+k_2+k_3)|\,[(k_1+k_3)^2+\mu^2](k_1^2+m^2)^{1/2} \times (k_2^2+m^2)^{1/2}\Big[\Big(\ln\Big(4\sqrt{\frac{m^2\zeta^2}{\mu^2}+1}\Big)+\frac{1}{2}\Big)^2+\frac{1}{4}\Big], \qquad (4.11)$$

for $m\zeta \gg \mu$. From (4.6), (4.10), and (4.11) it is clear that for all $m\zeta > \mu$ an $n$- and $\zeta$-independent constant $C_1$ can be found such that

$$\lim_{\Lambda\to\infty}\int d\mu \|L\|_2^{2n} \leqslant [n^{1/2} C_1 \ln(m\zeta/\mu)]^{2n}, \qquad (4.12)$$

for $n \geqslant 2$.

### 4.2 $\int d\mu \|HL\|_1^n$

By the hypercontractive inequality

$$\int d\mu \|HL\|_1^n \leqslant (n-1)^n\Big(\int d\mu \|HL\|_1^2\Big)^{n/2}, \qquad (4.13)$$

where $n \geqslant 2$ and

$$HL = [(P^2+m^2)^{1/4} S^{>} A_\Lambda g (P^2+m^2)^{-1/4-\delta}] \times [(P^2+m^2)^{1/4+\delta} S^{<} A_\Lambda g (P^2+m^2)^{-1/4}] \equiv A_> B_<, \qquad (4.14)$$

with $\delta > 0$ and

$$S^{>}(x) = \int_{|p|>m\zeta} \frac{d^2p}{(2\pi)^2} e^{ipx} \frac{m - \not{p}}{p^2+m^2}. \qquad (4.15)$$

From the definitions of $S^<$ and $S^>$ it is easy to show that $A_>, B_< \in \mathscr{C}_2$. Hence

$$\int d\mu \|HL\|_1^n \leqslant (n-1)^n\Big(\int d\mu \|A_>\|_2^2 \|B_<\|_2^2\Big)^{n/2}, \qquad (4.16)$$

where

$$\|A_>\|_2^2 = 2\int d^2x\, d^2y (A_{\mu,\Lambda} g)(x) D_{1/2+2\delta}(x-y) \times D^{>}_{1/2}(x-y)(A_{\mu,\Lambda} g)(y), \qquad (4.17)$$

$$\|B_<\|_2^2 = 2\int d^2x\, d^2y (A_{\mu,\Lambda})(x) D_{1/2}(x-y) \times D^{<}_{1/2-2\delta}(x-y)(A_{\mu,\Lambda} g)(y), \qquad (4.18)$$

and

$$D^{>}_{z}(x) = \int_{|p|>m\zeta} \frac{d^2p}{(2\pi)^2} \frac{e^{ipx}}{(p^2+m^2)^z}. \qquad (4.19)$$

Thus

$$\int d\mu \|A_>\|_2^2 \|B_>\|_2^2$$

$$= 4\int d^2x_1 \dots d^2y_2\, g(x_1)\dots g(y_2) D_{1/2+2\delta}(x_1-y_1) \times D^{>}_{1/2}(x_1-y_1) D_{1/2}(x_2-y_2) D^{<}_{1/2-2\delta}(x_2-y_2) \times [D^{\Lambda}_{\mu\mu}(x_1-y_1) D^{\Lambda}_{\nu\nu}(x_2-y_2) + D^{\Lambda}_{\mu\nu}(x_1-x_2) D^{\Lambda}_{\mu\nu}(y_1-y_2) + (x_2 \leftrightarrow y_2)]. \qquad (4.20)$$

The topologies of the Feynman diagrams corresponding to the right-hand side of (4.20) are the same as in Fig. 3 with appropriate $D_z$-functions. Again, since these are finite by power counting in the limit $\Lambda = \infty$ we get from (4.16) and (4.20)

$$\lim_{\Lambda\to\infty}\int d\mu \|HL\|_1^n \leqslant (n-1)^n (I_3 I_4 + I_5)^{n/2}, \qquad (4.21)$$

where

$$I_3 = \frac{2}{(2\pi)^6}\int \prod_{i=1}^{3} d^2k_i\, |\hat{g}(k_1)|^2 \hat{D}_{1/2+2\delta}(k_1+k_2+k_3) \times \hat{D}^{>}_{1/2}(k_2)\hat{D}_{\mu\mu}(k_3), \qquad (4.22)$$

$$I_4 = \frac{2}{(2\pi)^6}\int \prod_{i=1}^{3} d^2k_i\, |\hat{g}(k_1)|^2 \hat{D}_{1/2}(k_1+k_2+k_3) \times \hat{D}^{<}_{1/2-2\delta}(k_2)\hat{D}_{\mu\mu}(k_3), \qquad (4.23)$$

and

$$I_5 = \frac{8}{(2\pi)^{12}}\int \prod_{i=1}^{6} d^2k_i \hat{g}(k_1)\hat{g}(k_2)\hat{g}(k_3) \times \hat{g}(-k_1-k_2-k_3)\hat{D}^{>}_{1/2}(k_4)\hat{D}^{<}_{1/2-2\delta}(k_5) \times \hat{D}_{1/2+2\delta}(k_1+k_4+k_6) \times \hat{D}_{1/2}(k_5+k_6-k_2)\hat{D}_{\mu\nu}(k_6)\hat{D}_{\mu\nu}(k_1+k_3+k_6). \qquad (4.24)$$

Referring to the Appendix, (A10), (A11), and (A13) give

$$I_3 \leqslant [\delta(m\zeta)^{4\delta} m^{1+4\delta} 2^{7/2-2\delta}\pi^4]^{-1} \times \int d^2k\, |\hat{g}(k)|^2 (k^2+m^2)^{1/2+2\delta} \times \Big[\ln\Big(\frac{m^2\zeta^2}{\mu^2}+1\Big) + \frac{3}{2\delta(1-4\delta)}\Big], \qquad (4.25)$$

provided $0<\delta<\frac{1}{4}$, while (A14), (A15), and (A17) give, with the same restriction on $\delta$,

$$I_4 \leqslant \frac{(m\zeta)^{4\delta}}{m\delta 2^{9/2}\pi^4}\int d^2k\, |\hat{g}(k)|^2\sqrt{k^2+m^2} \times \Big[\ln\Big(\frac{m^2\zeta^2}{\mu^2}+1\Big)+\pi\Big]. \qquad (4.26)$$

Finally, (A18), (A19), and (A23) give

$$|I_5| \leqslant \frac{4^{\delta}160}{m^{4+4\delta}\mu^2}\int \prod_{i=1}^{3} d^2k_i\, |\hat{g}(k_1)\hat{g}(k_2)\hat{g}(k_3) \times \hat{g}(k_1+k_2+k_3)|\,[(k_1+k_3)^2+\mu^2] \times (k_1^2+m^2)^{1/2+2\delta}(k_2^2+m^2)^{1/2} \times \Big[\frac{4\pi^3 m^2}{\mu^2\delta}\Gamma\Big(\frac{1}{2}-2\delta\Big)\Gamma(4\delta) + \frac{C(\delta)}{\zeta^2}\Big], \qquad (4.27)$$

with $0<\delta<\frac{1}{4}$.

From (4.21) and (4.25)–(4.27) it is possible to find an $n$- and $\zeta$-independent constant $C_2$ such that for all $\zeta>0$ and $n\geqslant 2$

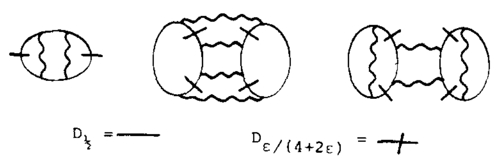


FIG. 4.

$$\lim_{\Lambda\to\infty}\int d\mu\|HL\|_1^n \leqslant \left[nC_2 \ln\left(\frac{m^2\zeta^2}{\mu^2}+1\right)\right]^n. \tag{4.28}$$

**4.3** $\int d\mu\|H\|_{2+\epsilon}^{2n}$

Using the general interpolation theorem for the spaces $\mathscr{C}_p$ as stated by Seiler and Simon[26] we get

$$\|H\|_{2+\epsilon} \leqslant C_3 \|H_{(4+3\epsilon)/(4+2\epsilon)}\|_2^{(2-\epsilon)/(2+\epsilon)} \times \|H_{(6+5\epsilon)/(8+4\epsilon)}\|_4^{2\epsilon/(2+\epsilon)}, \tag{4.29}$$

where $0 < C_3 \leqslant 1$,

$$H_z = (P^2+m^2)^{1/4} S_z^{>} A_\Lambda g (P^2+m^2)^{-1/4}, \tag{4.30}$$

and

$$S_z^{>} = \int_{|P|>m\zeta} \frac{d^2p}{(2\pi)^2} e^{ipx} \frac{m-\not p}{(p^2+m^2)^z}. \tag{4.31}$$

From (4.29) and Hölder's inequality we obtain

$$\int d\mu \|H\|_{2+\epsilon}^{2n} \leqslant C_3^{2n}\left(\int d\mu \|H_{(4+3\epsilon)/(4+2\epsilon)}\|_2^{4n/(2+\epsilon)}\right)^{(2-\epsilon)/2} \times\left(\int d\mu \|H_{(6+5\epsilon)/(8+4\epsilon)}\|_4^{8n/(2+\epsilon)}\right)^{\epsilon/2}, \tag{4.32}$$

where it is recalled that $0<\epsilon\leqslant 1$. Applying the hypercontractive inequality to the two integrals on the right-hand side of (4.32) gives, for $n \geqslant 2+\epsilon$,

$$\int d\mu \|H\|_{2+\epsilon}^{2n} \leqslant C_3^{2n}\left(\frac{2n-2-\epsilon}{2+\epsilon}\right)^n \times\left(\int d\mu \|H_{(4+3\epsilon)/(4+2\epsilon)}\|_2^4\right)^{(n/2)(2-\epsilon)/(2+\epsilon)} \times\left(\int d\mu \|H_{(6+5\epsilon)/(8+4\epsilon)}\|_4^8\right)^{\epsilon n/(4+2\epsilon)}. \tag{4.33}$$

Since for $\operatorname{Re} z > 1$,

$$\|H_z\|_2^2 = \frac{2}{(2\pi)^4}\int_{|p|>m\zeta} d^2q\, d^2p \times \frac{|\widehat{A_{\mu,\Lambda} g}(q)|^2}{[(p+q)^2+m^2]^{1/2}(p^2+m^2)^{2\operatorname{Re} z-3/2}}, \tag{4.34}$$

it follows that

$$\|H_{(4+3\epsilon)/(4+2\epsilon)}\|_2^2 \leqslant 2(2\pi)^{-4}[m^2(1+\zeta^2)]^{-\epsilon/(4+2\epsilon)}\int_{|p|>m\zeta} d^2q\, d^2p \times \frac{|\widehat{A_{\mu,\Lambda} g}(q)|^2}{[(p+q)^2+m^2]^{1/2}(p^2+m^2)^{(1+\epsilon)/(2+\epsilon)}} \leqslant [m^2(1+\zeta^2)]^{-\epsilon/(4+2\epsilon)} \|H^{\zeta=0}_{(8+5\epsilon)/(8+4\epsilon)}\|_2^2. \tag{4.35}$$

The topologies of the Feynman diagrams generated by $\int d\mu \|H^{\zeta=0}_{(8+5\epsilon)/(8+4\epsilon)}\|_2^4$ are those of Fig. 3 with $D^{<}_{1/2}$ replaced with $D^{>}_{(1+\epsilon)/(2+\epsilon)}$. These are finite by power counting in the limit $\Lambda=\infty$. Hence from (4.35)

$$\lim_{\Lambda\to\infty}\int d\mu \|H_{(4+3\epsilon)/(4+2\epsilon)}\|_2^4 \leqslant C_4[m^2(1+\zeta^2)]^{-\epsilon/(2+\epsilon)}, \tag{4.36}$$

where $C_4 = \lim_{\Lambda\to\infty} \int d\mu \|H^{\zeta=0}_{(8+5\epsilon)/(8+4\epsilon)}\|_2^4$.

Next, for $\operatorname{Re} z > 3/4$,

$$\|H_z\|_4^4 = (2\pi)^{-8}\int_{\substack{|k_2|>m\zeta\\|k_4|>m\zeta}} \prod_{i=1}^4 d^2k_i \times \frac{\operatorname{tr}[\widehat{A_\Lambda g}(k_1-k_2)\widehat{A_\Lambda g}(k_2-k_3)\widehat{A_\Lambda g}(k_3-k_4)\widehat{A_\Lambda g}(k_4-k_1)]}{(k_1^2+m^2)^{1/2}(k_2^2+m^2)^{2\operatorname{Re} z-3/2}(k_3^2+m^2)^{1/2}(k_4^2+m^2)^{2\operatorname{Re} z-3/2}}, \tag{4.37}$$

from which one obtains

$$\|H_{(6+5\epsilon)/(8+4\epsilon)}\|_4^4 \leqslant (2\pi)^{-8}[m^2(1+\zeta^2)]^{-\epsilon/(2+\epsilon)} \times \int_{\substack{|k_2|>m\zeta\\|k_4|>m\zeta}} \prod_{i=1}^4 d^2k_i \frac{\operatorname{tr}[\widehat{A_\Lambda g}(k_1-k_2)\ldots\widehat{A_\Lambda g}(k_4-k_1)]}{(k_1^2+m^2)^{1/2}(k_2^2+m^2)^{\epsilon/(4+2\epsilon)}(k_3^2+m^2)^{1/2}(k_4^2+m^2)^{\epsilon/(4+2\epsilon)}} \leqslant [m^2(1+\zeta^2)]^{-\epsilon/(2+\epsilon)} \|H^{\zeta=0}_{(3+2\epsilon)/(4+2\epsilon)}\|_4^4. \tag{4.38}$$

The topologies of the diagrams obtained from $\int d\mu \|H^{\zeta=0}_{(3+2\epsilon)/(4+2\epsilon)}\|_4^8$ are depicted in Fig. 4. All of them, including those obtained by permuting photon lines, are finite by power counting in the limit $\Lambda=\infty$. Therefore, using (4.38) we can state that

$$\lim_{\Lambda\to\infty}\int d\mu \|H_{(6+5\epsilon)/(8+4\epsilon)}\|_4^8 \leqslant C_5[m^2(1+\zeta^2)]^{-2\epsilon/(2+\epsilon)}, \tag{4.39}$$

where

$$C_5 = \lim_{\Lambda\to\infty}\int d\mu \|H^{\zeta=0}_{(3+2\epsilon)/(4+2\epsilon)}\|_4^8 < \infty.$$

Combining (4.33), (4.36), and (4.39) gives, for $n\geqslant 2+\epsilon$,

$$\lim_{\Lambda\to\infty}\int d\mu \|H\|_{2+\epsilon}^{2n} \leqslant n^n C_6^n (1+\zeta^2)^{-n\epsilon/(4+2\epsilon)}, \tag{4.40}$$

where $C_6$ is a $\zeta$-independent constant.

### 4.4 $\int d\mu|\mathrm{Tr}{:}K^2{:}|^n$

Application of the hypercontractive inequality gives

$$\int d\mu|\mathrm{Tr}{:}K^2{:}|^n \leqslant (n-1)^n\Big(\int d\mu(\mathrm{Tr}{:}K^2{:})^2\Big)^{n/2}, \tag{4.41}$$

provided $n\geqslant 2$. From (2.9) it follows that

$$\int d\mu(\mathrm{Tr}{:}K^2{:})^2 = 2\int d^2x_1\dots d^2y_2\, g(x_1)\dots g(y_2)\,\rho_{\mu_1\nu_1}(x_1-y_1)\rho_{\mu_2\nu_2}(x_2-y_2)D^{\Lambda}_{\mu_1\mu_2}(x_1-x_2)D^{\Lambda}_{\nu_1\nu_2}(y_1-y_2). \tag{4.42}$$

Noting from (2.10) that $\rho_{\mu\nu}(q)=O(q^2/m^2)$ for $q^2\to 0$ and $\rho_{\mu\nu}(q)=O(1)$ for $q^2\to\infty$, (4.42) is manifestly finite in the limit $\Lambda=\infty$ by power counting. Hence,

$$\lim_{\Lambda\to\infty}\int d\mu|\mathrm{Tr}{:}K^2{:}|^n \leqslant n^n C_7^n, \tag{4.43}$$

where

$$C_7^2 = \lim_{\Lambda\to\infty}\int d\mu(\mathrm{Tr}{:}K^2{:})^2 < \infty.$$

### 4.5 Bound on $\lim_{\Lambda\to\infty}|Z_{2n}(\Lambda)|$

We now combine (2.13), (3.1), (4.12), (4.28), (4.40), and (4.43) to obtain

$$\lim_{\Lambda\to\infty}|Z_{2n}(\Lambda)|\equiv|Z_{2n}| = \frac{(4e)^n}{4}\left\{\left[C_1\ln\left(\frac{m\zeta}{\mu}\right)\right]^{2n} + \left[\alpha C_2\ln\left(\frac{m^2\zeta^2}{\mu^2}+1\right)\right]^n + (\beta C_6)^n\left(\frac{n^2}{1+\zeta^2}\right)^{n\epsilon/(4+2\epsilon)} + \left(\frac{C_7}{2}\right)^n\right\}, \tag{4.44}$$

provided $n\geqslant 2+\epsilon$ and $m\zeta>\mu$. By setting $\zeta=n$ and $\mu<m$ it is evident that a sufficiently large constant $C$ can be found such that

$$|Z_{2n}|\leqslant[C\ln(mn/\mu)]^{2n}, \tag{4.45}$$

for all $n$.

## APPENDIX: ESTIMATES

### 1. $I_1$

Using[3]

$$[(k_1+k_2+k_3)^2+m^2]^{-1/2}\leqslant(\sqrt{2}/m)(k_1^2+m^2)^{1/2}[(k_2+k_3)^2+m^2]^{-1/2} \tag{A1}$$

for $k_i\in E$, where $E$ denotes a two-dimensional Euclidean space, and letting $k_{2,3}\to m\zeta k_{2,3}$ we get from (4.7)

$$\begin{aligned} I_1 &\leqslant \frac{2^{5/2}}{m(2\pi)^6}\int_{|k_2|<1}\prod_{i=1}^{3}d^2k_i\,\frac{|\hat g(k_1)|^2\sqrt{k_1^2+m^2}}{[(k_2+k_3)^2+1/\zeta^2]^{1/2}(k_2^2+1/\zeta^2)^{1/2}(k_3^2+\mu^2/m^2\zeta^2)} \\ &\leqslant \frac{2^{5/2}}{m(2\pi)^6}\int_{|k_2|<1}\prod_{i=1}^{3}d^2k_i\,\frac{|\hat g(k_1)|^2\sqrt{k_1^2+m^2}}{|k_2|\,|k_2+k_3|(k_3^2+\mu^2/m^2\zeta^2)}. \end{aligned} \tag{A2}$$

Let

$$J_1=\int_{|k_2|<1}\frac{|k_2+k_3|\,d^2k_2\,d^2k_3}{|k_2|(k_2+k_3)^2(k_3^2+\mu^2/m^2\zeta^2)}, \tag{A3}$$

and combine the denominators involving $k_2$ and $k_3$ using

$$\frac{1}{ab}=\int_0^1\frac{dz}{[az+b(1-z)]^2}. \tag{A4}$$

Then

$$J_1\leqslant\int_0^1 dz\int_{|k_2|<1}\frac{[|k_3|+|k_2|(1-z)]d^2k_2\,d^2k_3}{|k_2|[k_3^2+k_2^2z(1-z)+\mu^2(1-z)/m^2\zeta^2]^2}.$$

Letting $k_3^2\to[k_2^2z(1-z)+\mu^2(1-z)/m^2\zeta^2]k_3^2$ it follows that

$$J_1\leqslant\int_0^1\frac{dz}{\sqrt{1-z}}\int_{|k_2|<1}\frac{|k_3|d^2k_2\,d^2k_3}{|k_2|[zk_2^2+\mu^2/m^2\zeta^2]^{1/2}(k_3^2+1)^2}+\int_0^1 dz\int_{|k_2|<1}\frac{d^2k_2\,d^2k_3}{(k_2^2z+\mu^2/m^2\zeta^2)(k_3^2+1)^2}.$$

The remaining estimates are elementary and give

$$J_1\leqslant 2\pi^2\left[\ln\left(\frac{m\zeta}{\mu}\right)\right]^2+\pi^4\ln\left(\frac{2m\zeta}{\mu}+1\right)+\frac{\pi^4}{6}. \tag{A5}$$

Equation (A5) combines with (A2) to give (4.10).

### 2. $I_2$

Using the bound (A1) we get from (4.8)

$$|I_2| \leqslant \frac{160}{m^2\mu^2(2\pi)^{1/2}} \times \int_{\substack{|k_4|<m\zeta \\ |k_5|<m\zeta}} \prod_{i=1}^{6} d^2k_i \, |\hat{g}(k_1)\hat{g}(k_2)\hat{g}(k_3)\hat{g}(k_1+k_2+k_3)| \frac{[(k_1+k_3)^2+\mu^2](k_1^2+m^2)^{1/2}(k_2^2+m^2)^{1/2}}{|k_4|\,|k_4+k_6|\,|k_5|\,|k_5+k_6|(k_6^2+\mu^2)^2}. \tag{A6}$$

Let

$$J_2 = \int \frac{d^2k}{(k^2+\mu^2)^2} \left( \int_{|p|<m\zeta} \frac{d^2p}{|p|\,|p+k|} \right)^2, \tag{A7}$$

and note that

$$\int_{|p|<m\zeta} \frac{d^2p}{|p|\,|p+k|} \leqslant 2\pi \ln\left(4 + \frac{4m\zeta}{|k|}\right). \tag{A8}$$

Setting $x = k^2/(k^2+\mu^2)$ we get

$$J_2 \leqslant \frac{(2\pi)^3}{2\mu^2} \int_0^1 dx \left[ \ln\left(4 + \frac{4m\zeta}{\mu}\sqrt{\frac{1-x}{x}}\right)\right]^2 \leqslant \frac{(2\pi)^3}{2\mu^2} \left\{ \left[ \ln\left(4\sqrt{1+\frac{m^2\zeta^2}{\mu^2}}\right) + \frac{1}{2}\right]^2 + \frac{1}{4} \right\}, \tag{A9}$$

provided $m\zeta \gg \mu$. Insertion of (A9) in (A6) gives the bound (4.11).

### 3. $I_3$

From (A1), (4.22), and the change of scale $k_{2,3} \to m\zeta k_{2,3}$ we obtain

$$I_3 \leqslant \frac{2^{5/2+2\delta}}{(m\zeta)^{4\delta} m^{1+4\delta}(2\pi)^6} \int_{|k_2|>1} \prod_{i=1}^{3} d^2k_i \, \frac{|\hat{g}(k_1)|^2 (k_1^2+m^2)^{1/2+2\delta}}{|k_2|\,|k_2+k_3|^{1+4\delta}(k_3^2+\mu^2/m^2\zeta^2)}. \tag{A10}$$

Let

$$J_3 = \int_{|k_2|>1} \frac{|k_2+k_3|^{1-4\delta} d^2k_2\, d^2k_3}{|k_2|(k_2+k_3)^2(k_3^2+\mu^2/m^2\zeta^2)}. \tag{A11}$$

Combine denominators and rescale $k_3$ as for $I_1$ to obtain

$$J_3 \leqslant \int_0^1 \frac{dz}{(1-z)^{1/2+2\delta}} \int_{|k_2|>1} \frac{|k_3|^{1-4\delta} d^2k_2\, d^2k_3}{|k_2|(zk_2^2+\mu^2/m^2\zeta^2)^{1/2+2\delta}(k_3^2+1)^2} + \int_0^1 \frac{dz}{(1-z)^{4\delta}} \int_{|k_2|>1} \frac{d^2k_2\, d^2k_3}{|k_2|^{4\delta}(zk_2^2+\mu^2/m^2\zeta^2)(k_3^2+1)^2}, \tag{A12}$$

provided $0<\delta<\frac{1}{4}$. After some easy estimates we get

$$J_3 \leqslant \frac{\pi^2}{\delta}\left[ \ln\left(\frac{m^2\zeta^2}{\mu^2}+1\right) + \frac{3}{2\delta(1-4\delta)}\right]. \tag{A13}$$

Equation (A13) combines with (A10) to give (4.25).

### 4. $I_4$

From (4.23) and proceeding exactly as for $I_1$ and $I_3$ one gets

$$I_4 \leqslant \frac{2^{5/2}(m\zeta)^{4\delta}}{m(2\pi)^6} \int_{|k_2|<1} \prod_{i=1}^{3} d^2k_i \, \frac{|\hat{g}(k_1)|^2\sqrt{k_1^2+m^2}}{|k_2|^{1-4\delta}|k_2+k_3|(k_3^2+\mu^2/m^2\zeta^2)}, \tag{A14}$$

provided $0<\delta<\frac{1}{4}$. Let

$$J_4 = \int_{|k_2|<1} \frac{|k_2+k_3|\, d^2k_2 d^2k_3}{|k_2|^{1-4\delta}(k_2+k_3)^2(k_3^2+\mu^2/m^2\zeta^2)}. \tag{A15}$$

Combining denominators and rescaling $k_3$ as for $I_1$ gives

$$J_4 \leqslant \int_0^1 \frac{dz}{\sqrt{1-z}} \int_{|k_2|<1} \frac{|k_3|\, d^2k_2 d^2k_3}{|k_2|^{1-4\delta}(zk_2^2+\mu^2/m^2\zeta^2)^{1/2}(k_3^2+1)^2} + \int_0^1 dz \int_{|k_2|<1} \frac{|k_2|^{4\delta} d^2k_2\, d^2k_3}{(zk_2^2+\mu^2/m^2\zeta^2)(k_3^2+1)^2}, \tag{A16}$$

from which one easily obtains

$$J_4 \leqslant \pi^2\left[\ln\left(\frac{m^2\zeta^2}{\mu^2}+1\right)+\pi\right]\Big/ 2\delta. \tag{A17}$$

Equation (A17) combines with (A14) to give (4.26).

**5. $I_5$**

From (4.24) and repeated use of the bound (A1) we get after the scale change $k_{4,5,6} \to m\zeta k_{4,5,6}$

$$|I_5| \leqslant \frac{4^\delta 160}{(m^{2+2\delta}\mu\zeta)^2} \int_{\substack{|k_4|>1\\|k_5|<1}} \prod_{i=1}^{6} d^2k_i \, |\hat{g}(k_1)\hat{g}(k_2)\hat{g}(k_3)\hat{g}(k_1+k_2+k_3)| \frac{[(k_1+k_3)^2+\mu^2](k_1^2+m^2)^{1/2+2\delta}(k_2^2+m^2)^{1/2}}{|k_4|\,|k_4+k_6|^{1+4\delta}|k_5|^{1-4\delta}|k_5+k_6|(k_6^2+\mu^2/m^2\zeta^2)^2}. \tag{A18}$$

Next we split the $k_6$ integration into a high and low momentum piece. Let

$$J_5^{>} = \int_{\substack{|k_4|>1\\|k_5|<1}} \frac{d^2k_4\, d^2k_5\, d^2k_6\, \theta(|k_6|-1)}{|k_4|\,|k_4+k_6|^{1+4\delta}|k_5|^{1-4\delta}|k_5+k_6|(k_6^2+\mu^2/m^2\zeta^2)^2} \leqslant C(\delta), \tag{A19}$$

where $C$ is a $\zeta$-independent constant that is finite by power counting provided $0<\delta<\frac{1}{4}$. The other contribution from the $k_6$ integration is

$$J_5^{<} = \int_{\substack{|k_4|>1\\|k_5|<1}} \frac{d^2k_4\, d^2k_5\, d^2k_6\, \theta(1-|k_6|)}{|k_4|\,|k_4+k_6|^{1+4\delta}|k_5|^{1-4\delta}|k_5+k_6|(k_6^2+\mu^2/m^2\zeta^2)^2}. \tag{A20}$$

Using the estimates

$$\int_{\substack{|k_4|>1\\|k_6|<1}} \frac{d^2k_4}{|k_4|\,|k_6|^{1+4\delta}} \leqslant 2\pi\Gamma\left(\frac{1}{2}-2\delta\right)\Gamma(4\delta) \tag{A21}$$

$$\int_{\substack{|k_5|<1\\|k_6|<1}} \frac{d^2k_5}{|k_5|^{1-4\delta}|k_5+k_6|} \leqslant \frac{2\pi}{\delta}, \tag{A22}$$

one gets

$$J_5^{<} \leqslant \frac{4\pi^3}{\delta}\Gamma\left(\frac{1}{2}-2\delta\right)\Gamma(4\delta)\left(\frac{m\zeta}{\mu}\right)^2. \tag{A23}$$

Equations (A19) and (A23) combine with (A18) to give (4.27).


[1]J. Eckmann, J. Magnen, and R. Sénéor, Commun. Math. Phys. **39**, 251 (1975); J. Feldman and K. Osterwalder, Ann. Phys. (N.Y.) **97**, 80 (1976).
[2]J. Magnen and R. Sénéor, Commun. Math. Phys. **56**, 237 (1977).
[3]P. Renouard, Ann. Inst. Henri Poincaré A **27**, 237 (1977).
[4]P. Renouard, Ann. Inst. Henri Poincaré A **31**, 235 (1979).
[5]J. Eckmann and H. Epstein, Commun. Math. Phys. **68**, 245 (1979).
[6]M. Bergère and F. David, Ann. Phys. (N.Y.) **142**, 416 (1982) and references therein; R. Jackiw and S. Templeton, Phys. Rev. D **23**, 2291 (1981) and references therein.
[7]G. 't Hooft, in *Orbis Scientiae* (Coral Gables, 1977) (Plenum, New York, 1977); G. 't Hooft, in *the Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1979); N. Khuri, Phys. Rev. D **23**, 2285 (1981).
[8]N. Khuri, Phys. Rev. D **16**, 1754 (1977).
[9]F. Calogero, Lettere Nuovo Cimento **25**, 533 (1979).
[10]J. Fröhlich, Nucl. Phys. B **200** [FS4], 281 (1982); G. Gallavotti and V. Rivasseau, Phys. Lett. B **122**, 268 (1983).
[11]B. Simon, Int. J. Quantum Chem. **21**, 3 (1982).
[12]A. Jaffe, Commun. Math. Phys. **1**, 127 (1965).
[13]G. Parisi, Phys. Lett. B **66**, 167 (1977); B **68**, 361 (1977).
[14]E. Brézén and G. Parisi, J. Stat. Phys. **19**, 269 (1978).
[15]G. Parisi, Phys. Lett. B **66**, 382 (1977).
[16]C. Itzykson, G. Parisi, and J. Zuber, Phys. Rev. D **16**, 996 (1977); M. Fry, Phys. Lett. B **80**, 65 (1978).
[17]J. Schwinger, Phys. Rev. **128**, 2425 (1964).
[18]C. de Calan and V. Rivasseau, Commun. Math. Phys. **82**, 69 (1981).
[19]J. Fröhlich and E. Seiler, Helv. Phys. Acta **49**, 889 (1976).
[20]M. Fry, Nucl. Phys. B **121**, 343 (1977).
[21]H. Georgi and J. Rawls, Phys. Rev. D **3**, 874 (1970).
[22]K. Ito, Phys. Lett. B **91**, 406 (1980).
[23]S. Adler, Phys. Rev. D **10**, 2399 (1974); D **15**, 1803 (E) (1977).
[24]R. Balian, C. Itzykson, G. Parisi, and J. Zuber, Phys. Rev. D **17**, 1041 (1978).
[25]E. Seiler, Phys. Rev. D **22**, 2412 (1980).
[26]E. Seiler and B. Simon, J. Math. Phys. **16**, 2289 (1975).
[27]B. Simon, Adv. Math. **24**, 244 (1977).
[28]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II.
[29]N. Dunford and J. Schwartz, *Linear Operators* (Interscience, New York, 1963), Vol. II.
[30]B. Simon, *Trace Ideals and Their Applications* (Cambridge University, Cambridge, 1979).
[31]E. Nelson, J. Funct. Anal. **12**, 211 (1973); in *Constructive Quantum Field Theory*, edited by G. Velo and A. Wightman (Springer, Berlin, 1973).

# Electromagnetic fields invariant up to a duality rotation under a group of isometries[a)]

Marc Henneaux[b)]
*Center for Theoretical Physics, University of Texas, Austin, Texas 78712*

(Received 1 June 1983; accepted for publication 27 December 1983)

Electromagnetic fields invariant up to a duality rotation under a group $H$ of space–time isometries are analyzed. The symmetry equations $h\,^*F = \cos\alpha(h)F + \sin\alpha(h)F^*$ are integrated by noticing that $\alpha$ defines a homomorphism of $H$ to SO(2). Applications of that concept to Einstein–Maxwell equations are studied. Cosmological models are considered. Special attention is paid to Bianchi universes which are shown to admit nontrivial, spatially homogeneous-up-to-a-duality-rotation, electromagnetic fields of all algebraic types. All L.R.S. type-V solutions to Einstein–Maxwell equations in which the electromagnetic field shares the symmetry of the gravitational field up to a duality transformation are derived. Discrete isometries are also analyzed.

PACS numbers: 04.20.Jb

## I. INTRODUCTION

Let $(M,g)$ be an oriented Riemannian space–time and let $H$ be its group of isometries. An electromagnetic field $F$ defined on $M$ is said to be "invariant up to a duality transformation under the group $H$" if, for all elements $h \in H$, $h * F$ differs from $F$ by a duality rotation,

$$h * F = \cos\alpha(h)\, F + \sin\alpha(h) F^*. \tag{1.1}$$

Here, $h * F$ is the usual pullback of $F$ by $h$, whereas $F^*$ is its dual two-form. The angle $\alpha(h)$ depends on the group element $h$, but is a space–time constant. Unless otherwise stated, the terms "duality transformation" will always mean "constant (in space–time) duality transformation."

When $h$ preserves the orientation, the property (1.1) implies (see Appendix A)

$$h * F^* = -\sin\alpha(h) F + \cos\alpha(h)\, F^* \tag{1.2}$$

and

$$h * F^\dagger = e^{i\alpha(h)} F^\dagger, \tag{1.3}$$

$$h * \widetilde{F} = e^{-i\alpha(h)} \widetilde{F}, \tag{1.4}$$

where $F^\dagger$ and $\widetilde{F}$ are, respectively, the following self-dual and anti-self-dual two-forms:

$$F^\dagger = \tfrac{1}{2}(F - iF^*), \tag{1.5}$$

$$\widetilde{F} = \tfrac{1}{2}(F + iF^*). \tag{1.6}$$

It is well known that if the metric $g$ and the field $F$ obey Einstein–Maxwell equations and if $F$ is nonsingular, then, every symmetry of the metric is a symmetry of the Maxwell field up to a duality transformation. This results from a theorem by Misner and Wheeler that states that the electromagnetic field itself is determined from the metric up to a duality transformation,[1] and motivates our present work. Some examples of Einstein–Maxwell solutions with an electromagnetic field that shares the symmetry of the metric only up to a nontrivial duality rotation have been given in the literature.[2]

As we shall see, the study of the equation (1.1) is somehow similar to the study of gauge fields invariant up to a gauge,[3] of spinor fields invariant up to a phase transformation,[4] and of homothetic motions.[5]

It follows from (1.1), (1.2), and the properties of the pullback of forms that the function $\alpha$: $h \to \alpha(h)$ defines a group homomorphism of $H$ to SO(2),

$$\alpha(hg) = \alpha(h) + \alpha(g). \tag{1.7}$$

When the image of $H$ by this homomorphic mapping is the identity, the relation (1.1) reduces to the strict invariance of $F$. New interesting possibilities appear when the image of $H$ is SO(2) itself or some nontrivial subgroup.

Since SO(2) is abelian, one easily infers from (1.7) that $\alpha(h)$ vanishes for all commutators,

$$\alpha(h_1^{-1} h_2^{-1} h_1 h_2) = 0. \tag{1.8}$$

Accordingly, the derived group $H'$ belongs to the kernel of the homomorphism. When $H$ is abelian, this is obvious, but in the case when $H'$ is equal to $H$ (as for noncommutative simple groups), this imposes $\alpha(H) = \{0\}$.

We shall assume from now on that $H$ is a $n$-dimensional Lie group $(1 \leqslant n \leqslant 10)$ and shall confine our attention on its component connected with the identity. The above formulas can then be rewritten

$$\mathscr{L}_{\xi_A} F = k_A F^*, \quad \mathscr{L}_{\xi_A} F^* = -k_A F \tag{1.9}$$

and

$$\mathscr{L}_{\xi_A} F^\dagger = ik_A F^\dagger, \quad \mathscr{L}_{\xi_A} \widetilde{F} = -ik_A \widetilde{F}, \tag{1.10}$$

where $k_A$ is defined, in our additive notations, by

$$\alpha[\exp t\xi_A] = k_A t, \tag{1.11}$$

and where $\mathscr{L}_{\xi_A}$ are the Lie derivative operators along the Killing vectors $\xi_A$ $(A = 1,\ldots,n)$. Formula (1.8) becomes

$$k_A C^A{}_{BC} = 0, \tag{1.12}$$

where $C^A{}_{BC}$ are the structure constants of the isometry group.

We shall also assume that the group $H$ is transitive on $M$. The discussion is easily extended to the general case of a nontransitive group.

---

[a)] Work supported in part by U.S. National Science Foundation grant number PHY-8216715 to the University of Texas.

[b)] On leave from Département de Physique, Université Libre de Bruxelles, Belgium.

  

## II. HOMOMORPHISMS $H \rightarrow SO(2)$

Let us consider a basis of right-invariant vector fields on $H$, denoted by $\{\xi_A\}$, and its dual basis, $\{\overline{\omega}^A\}$ (no confusion should arise between $\xi_A$, right-invariant vector field on $H$ and $\xi_A$, Killing vector field on $M$). One has

$$[\xi_A, \xi_B] = C^C{}_{AB}\xi_c, \quad d\overline{\omega}^A = -\tfrac{1}{2} C^A{}_{BC}\overline{\omega}^B \wedge \overline{\omega}^c. \quad (2.1)$$

The left-invariant vector fields $X_A$ such that

$$X_A(e) = \xi_A(e) \quad (2.2)$$

($e$ is the identity) obey

$$[X_A, X_B] = -C^C{}_{AB}X_C \quad (2.3)$$

and their dual basis $\{\omega^A\}$ is such that

$$d\omega^A = \tfrac{1}{2} C^A{}_{BC}\omega^B \wedge \omega^C. \quad (2.4)$$

**Theorem**: There is a bijective correspondence between homomorphisms $\alpha$: $H \rightarrow SO(2)$ and functions on $H$, (i) which vanish at the identity, and (ii) the gradients of which are right invariant.

*Proof*: (1) $d\alpha$ is right invariant [$\alpha(e) = 0$ is obvious].

From the homomorphism condition (1.7), one easily derives

$$\alpha[\phi_t(h)] = \alpha[\phi_t(e)] + \alpha(h), \quad (2.5)$$

when $\phi_t$ is the one-parameter group of left translations generated by an arbitrary right-invariant vector field. It thus follows that

$$\mathscr{L}_{\xi_A}\alpha \equiv \partial_{\xi_A}\alpha = k_A, \quad (2.6)$$

where the numbers $k_A$ are the values of $\mathscr{L}_{\xi_A}\alpha$ at the identity. This in turn implies that the gradient of $\alpha$,

$$d\alpha \equiv \partial_{\xi_A}\alpha\, \overline{\omega}^A = k_A\overline{\omega}^A, \quad (2.7)$$

is right invariant.

It is clear that the same argument applied to right translations shows that $d\alpha$ is also left invariant. Moreover, one has

$$\mathscr{L}_{X_A}\alpha = k_A \quad (2.8)$$

(with the same $k_A$), since $X_A = \xi_A$ at the identity. Actually, if the gradient of a function is right (left) invariant, it is automatically left (right) invariant because $X_A$ and $\xi_B$ commute.

The condition (1.12) is equivalent to $d^2\alpha = 0$.

(2) If $df$ is right invariant and if $f(e) = 0$, then $f$ defines a homomorphism of $H$ to SO(2).

Indeed, one finds

$$\begin{aligned} f(g_1g_2) &= f(g_2) + \int_{g_2}^{g_1g_2} df \\ &= f(g_2) + \int_e^{g_1} df \\ &= f(g_2) + f(g_1) - f(e) \\ &= f(g_1) + f(g_2), \end{aligned}$$

where the transformation of the integral is allowed because of the invariance of $df$ (right multiply the path joining $g_2$ to $g_1 g_2$ by $g_2^{-1}$). This proves the theorem.[6]

**Theorem**: Any set of constants $k_A$ obeying (1.12)

$$k_A C^A{}_{BC} = 0$$

defines one and only one local homomorphism of $H$ to SO(2).

Indeed, the right-invariant one-form $\omega = k_A\overline{\omega}^A$ is closed and defines locally one and only one function $\alpha$ such that

(i) $\alpha(e) = 0$

(ii) $d\alpha = \omega$.

Global restrictions arise when $H$ is not simply connected.

## III. SOLUTION TO THE INVARIANCE CONDITIONS—$H$ IS SIMPLY TRANSITIVE ON $M$

In order to derive the solution to the symmetry equations (1.1) for a given $H$, we first consider the case when $H$ is simply transitive: to any pair $(P, P')$ of space–time points, there corresponds one and only one transformation $h \in H$ such that $h(P) = P'$ ($M$ can be identified with $H$; the Killing vector fields and the right-invariant vector fields then coincide; $A = 1,2,3,4$).

Let us choose an arbitrary fiducial point $P_0$ and denote by $h_P$ the unique transformation of $H$ that maps $P$ on $P_0$. Let $\alpha$ be a homomorphism of $H$ to SO(2).

It is clear that $F$ is determined everywhere in $M$ by the symmetry conditions (1.1) whenever $F$ is known at $P_0$, and that these conditions do not restrict $F(P_0)$. The expression

$$F(P) = \cos\alpha(h_P)\mathring{F}(P) - \sin\alpha(h_P)\mathring{F}^*(P) \quad (3.1)$$

with $\mathring{F}(P) = h_P^*\mathring{F}(P_0) = h_P^*F(P_0)$ $(h^*\mathring{F} = \mathring{F}\ \forall\ h \in H)$, is accordingly the general solution to the symmetry equations. $F$ differs from the invariant two-form field $\mathring{F}$ by a space–time dependent duality rotation.

In the invariant basis $\{\omega^A\}$, (3.1) reads

$$F_{AB}(P) = \cos\alpha(h_P)\mathring{F}_{AB} - \sin\alpha(h_P)\mathring{F}^*_{AB}, \quad (3.2)$$

where the components $\mathring{F}_{AB}$ are constant.

**Theorem**: If both $F$ and $\mathring{F}$ obey Maxwell equations ($dF = d\mathring{F} = dF^* = d\mathring{F}^*$), then

(i) either $d\alpha(h_P) \neq 0$ is lightlike, in which case $F$ and $\mathring{F}$ are null ($\mathbf{E}^2 - \mathbf{B}^2 = \mathbf{E}\cdot\mathbf{B} = 0$); (ii) or $\alpha(H) = \{0\}$ and $F$ is strictly invariant.

*Proof*: In terms of the self-dual two-form $F^\dagger$, (3.1) reduces to

$$F^\dagger = e^{-i\alpha(h_P)}\mathring{F}^\dagger. \quad (3.3)$$

This leads, assuming Maxwell equations for both $F$ and $\mathring{F}$, to

$$d\alpha(h_P) \wedge \mathring{F}^\dagger = 0. \quad (3.4)$$

If $\alpha(h_P) \neq 0$ is timelike or spacelike, (3.4) implies $\mathring{F}^\dagger = 0$ (use the self-duality of $\mathring{F}^\dagger$). Accordingly, if the field $F$ is nontrivial, either $\alpha(h_P) = \text{const}(\Rightarrow \alpha(h_P) = \alpha(h_{P_0}) = 0)$, or $d\alpha(h_P)$ is lightlike. In that latter case (3.4) implies that the invariants $\mathbf{E}^2 - \mathbf{B}^2$ and $\mathbf{E}\cdot\mathbf{B}$ both vanish (see Ref. 1).

It results from this theorem that the unphysical invariant form $\mathring{F}$ is, in general, not a solution to Maxwell equations.

## IV. SOLUTION TO THE INVARIANCE CONDITIONS—$H$ IS MULTIPLY TRANSITIVE ON $M$

In that case, $M$ can be identified with the quotient space $H/K$, i.e., with the set of left cosets $hK$ of the stability subgroup at, say, $P_0$. $K$ is isomorphic to the stability subgroups at the other points. Greek indices will refer to $M$, capital Latin indices to $H$, and small Latin indices to the subgroup $K$.

In the mapping $u$: $H \rightarrow M$: $g \rightarrow gK$, the right-invariant vector fields $\xi_A$ are mapped, as is well known, on the Killing vector fields $\xi_A$, whereas the left-invariant vector fields $X_a$ (corresponding to the subgroup $K$) are mapped on 0,

$$u_* \xi_A = \xi_A, \quad u_* X_a = 0. \qquad (4.1)$$

The pullback of any two-form field $\varphi$ on $M$ is a two-form field on $H$ that obeys

$$\mathscr{L}_{X_a} u^* \varphi = 0 \quad X_a \lrcorner u^* \varphi = 0. \qquad (4.2)$$

Moreover, one finds

$$\mathscr{L}_{\xi_A} u^* \varphi = u^* \mathscr{L}_{\xi_A} \varphi. \qquad (4.3)$$

Reciprocally, if a two-form field $\chi$ on $H$ obeys $\mathscr{L}_{X_a} \chi = 0 = X_a \lrcorner \chi$, there is one and only one two-form field on $M$ such that $\chi = u^* \varphi$.

Let $G^\dagger$ be the pullback of $F^\dagger$ ($u^* F^\dagger = G^\dagger$). $G^\dagger$ cannot be self-dual on $H$, since the dimension of $H$ exceeds four. We shall solve the symmetry equations

$$\mathscr{L}_{\xi_A} G^\dagger = i k_A G^\dagger \qquad (4.4)$$

on the group $H$ and then "project" $G^\dagger$ back on space–time (standard trick of differential geometry).

From the analysis of the previous section, it follows that the general solution of (4.4) is given by

$$G^\dagger(h) = \mathring{G}^\dagger{}_{AB} e^{i\alpha(h)} \omega^A \wedge \omega^B, \qquad (4.5)$$

where the $\mathring{G}^\dagger{}_{AB}$'s are constant and where $\alpha$ is a homomorphism of $H$ to SO(2). We must then impose the conditions (4.2), which turn out to be algebraic equations for $G^\dagger_{AB}$. Indeed, the second equation (4.2) becomes

$$\mathring{G}^\dagger{}_{aB} = \mathring{G}^\dagger{}_{Ab} = 0 \qquad (4.6)$$

(only $\mathring{G}^\dagger{}_{\alpha\beta}$ can be different from zero), whereas the first one reads

$$i k_a \mathring{G}^\dagger{}_{AB} + \mathring{G}^\dagger{}_{AF} C^F{}_{aB} - \mathring{G}^\dagger{}_{BF} C^F{}_{aA} = 0 \qquad (4.7)$$

[we have used $\mathscr{L}_{X_a} \omega^A = C^A{}_{aB} \omega^B$, which follows from the identity $\mathscr{L}_X \omega = X \lrcorner d\omega + d(X \lrcorner \omega)$]. These equations can be rewritten as

$$\mathring{G}^\dagger \Lambda_a - (\mathring{G}^\dagger \Lambda_a)^\dagger + i k_a \mathring{G}^\dagger = 0, \qquad (4.8)$$

where the matrix $\Lambda_a$ has components $(\Lambda_a)^C{}_A = C^C{}_{aA}$.

The problem of determining all $H$-invariant two-forms $F$ (up to a duality transformation) is thus reduced to the algebraic problem (4.6)–(4.7) and the demand that $G^\dagger$ induces a self-dual form on $M$.

Note that the two-form $\mathring{G}^\dagger$ is projectable on $M$ if and only if $k_a = 0$, i.e., if the homomorphism of the isotropy subgroup $K$ to SO(2) defined by $\alpha$ is trivial. It is shown in Appendix B that when $k_a \neq 0$, $F$ is necessarily a null two-form.

## V. A CLASS OF HOMOTHETIC MODELS

As a first application, we consider space-times with a four-dimensional transitive group $G_4$(I) of homothetic motions. The group is of type I according to the classification given in Petrov (Ref. 7, p. 63). Its generators are $\xi_0 = \partial_0$, $\xi_1 = \partial_1$, $\xi_2 = -x^1 \partial_0 + \partial_2$, $\xi_3 = (b^2 - 1) x^0 \partial_0 - x^1 \partial_1 + b^2 x^2 \partial_2 + \partial_3 (b \neq 0)$. The metric $g$ is homothetically invariant,

$$\mathscr{L}_{\xi_A} g = 2\sigma_A g \quad (A = 0,1,2,3). \qquad (5.1)$$

This is a generalization of cosmological models homogeneous in space and time.

A basis of invariant forms is given by

$$\omega^0 = e^{(1-b^2)x^3}(dx^0 + x^2\, dx^1),$$
$$\omega^1 = e^{x^3}\, dx^1,$$
$$\omega^2 = e^{-b^2 x^3}\, dx^2, \qquad (5.2)$$
$$\omega^3 = dx^3.$$

Since $\sigma_A$ generates a homomorphism of $G_4$(I) to $R$, $\sigma_A C^A{}_{BC}$ must vanish, which implies that only $\sigma_3$ can be different from zero.

We shall further assume that the metric is diagonal in the basis (5.2) and that $\omega^0$ is timelike. By appropriate normalizations, the coefficients of $(\omega^0)^2$ and $(\omega^1)^2$ can be set equal to $\pm e^{2\sigma_3 x^3}$. The metric reads[5]

$$ds^2 = e^{2\sigma x^3}[-(\omega^0)^2 + (\omega^1)^2 + a^2(\omega^2)^2 + c^2(\omega^3)^2] \qquad (5.3)$$

($\sigma \equiv \sigma_3$).

The Maxwell field must obey

$$\mathscr{L}_{\xi_A} F = \sigma_A F + k_A F^*, \qquad (5.4)$$

which is a natural extension of the equations of the previous section. Again, only $k_3 \neq 0$. This implies

$$F = e^{\sigma x^3}(\mathring{F}_{\lambda\mu} \cos kx^3 + \mathring{F}^*_{\lambda\mu} \sin kx^3) \omega^\lambda \wedge \omega^\mu, \qquad (5.5)$$

where $\mathring{F}_{\lambda\mu}$ are arbitrary constants.

From Maxwell equations, one infers

$$\sigma = -(1 - b^2), \qquad (5.6a)$$
$$\mathring{F}_{01} = \mathring{F}_{02} = \mathring{F}_{13} = \mathring{F}_{23} = 0, \qquad (5.6b)$$
$$\mathring{F}_{03} = 2b\, \mathring{F}_{12}, \qquad (5.6c)$$
$$k = 2b. \qquad (5.6d)$$

Accordingly, the Maxwell field is non-null. One of its principal orthonormal tetrads is just obtained from $\{\omega^\mu\}$ by appropriate rescalings. If one had not allowed for the possibility of a duality rotation in (5.4), one would have been unable to fulfill the Maxwell equations (the field $e^{\sigma x^3}\mathring{F}$ cannot obey these equations) and one would have missed the solutions below. This shows the importance of incorporating the term $k_A F^*$ in (5.4)

Finally, the Einstein equations, which also turn out to be algebraic equations, simply yield

$$c^2 = 4a^2 b^2 \qquad (5.7)$$

and

$$\mathring{F}_{03} = 2b. \qquad (5.8)$$

This completes the resolution of the Einstein–Maxwell equations for the above fields.

The metrics (5.3), (5.6a), (5.7) depend on two parameters. They belong to a class described by Barnes,[8] who found them by algebraic means. When $b^2 = 1$, $\sigma$ vanishes by (5.6a) and the homothetic motions reduce to true isometries (McLenaghan–Taricq–Tupper solutions). Note again that $k$ never vanishes ($b \neq 0$).

## VI. BIANCHI COSMOLOGICAL MODELS WITH AN ELECTROMAGNETIC SOURCE

As a second example, we consider cosmological models of the Bianchi type whose source is an electromagnetic field that shares the symmetry of the metric up to a duality transformation. The isometry groups are three-dimensional and act on spacelike hypersurfaces. Their structure constants can be written as

$$C^a{}_{bc} = \epsilon_{bcd} n^{ad} + \delta^a{}_b a_c - \delta^a{}_c a_b, \tag{6.1}$$

with $n^{ab}a_b = 0$ (see Ref. 9, Chap. 6, for the details). From now on, small Latin indices stand for group indices and run from 1 to 3.

For all types but types VIII and IX (which will be excluded in the sequel), the equations $k_a C^a{}_{bc} = 0$ possess nonzero solutions and allow for the new possibility of electromagnetic fields invariant up to a nontrivial duality rotation. These equations have actually been studied by Eardley in the context of homothetic Bianchi models,[10] and we will not repeat his discussion here [homomorphisms $H \to SO(2)$ and $H \to R$ are locally equivalent].

Let $x^0 = 0$ be a hypersurface of transitivity. It is easy to show that the following equations hold on it as a consequence of the symmetry hypotheses.

$$^{(3)}\mathscr{L}_{\xi_a} g_{km} = 0, \quad ^{(3)}\mathscr{L}_{\xi_a} K_{km} = 0, \tag{6.2}$$

$$^{(3)}\mathscr{L}_{\xi_a} \mathscr{E}^k = -k_a \mathscr{B}^k, \quad ^{(3)}\mathscr{L}_{\xi_a} \mathscr{B}^k = k_a \mathscr{E}^k. \tag{6.3}$$

Here, $g_{km}$ is the metric induced on the hypersurface, $K_{km}$ is its intrinsic curvature whereas $\mathscr{E}^k$ and $\mathscr{B}^k$ are the electric and magnetic components (with respect to the hypersurface) of the electromagnetic field.[11] Moreover, the fields $g_{km}$, $K_{km}$, $\mathscr{E}^k$, and $\mathscr{B}^k$ are constrained on the $x^0 = 0$-hypersurface by the $G_{k\perp} = T_{k\perp}$ equations, as well as by Gauss' law and the div $\mathscr{B} = 0$ equation. These equations are called the constraints, as opposed to the other Einstein–Maxwell equations, which are truly dynamical.

**Theorem**: Let conversely $g_{km}$, $K_{km}$, $\mathscr{E}^k$, and $\mathscr{B}^k$ (i) obey both the conditions (6.2), (6.3) and the constraints on the hypersurface $x^0 = 0$; and (ii) be propagated off that hypersurface by means of the dynamical Einstein–Maxwell equations. Then the group generated by the $\xi_a$'s is an isometry group of the full space–time metric and is such that $\mathscr{L}_{\xi_a} F = k_a F^*$ (and of course, the constraints are preserved in time).

The proof of this theorem, which shows that the assumed symmetry is compatible with the Einstein–Maxwell equations provided it is with the constraints, is standard (see in this context Refs. 10 and 12): take for simplicity a slicing obtained from $x^0 = 0$ by the conditions $\mathscr{L}_{\xi_a} N = 0$, $\mathscr{L}_{\xi_a} N^k = 0$ ($N$ is the lapse, $N^k$ is the shift). Show that the initial conditions (6.2)–(6.3), together with the dynamical equations, imply in that gauge (i) $\partial_0 {}^{(3)}\mathscr{L}_{\xi_a} g_{km}$ $(= {}^{(3)}\mathscr{L}_{\xi_a} \partial_0 g_{km}) = 0 = \partial_0 {}^{(3)}\mathscr{L}_{\xi_a} K_{km}$ $(= {}^{(3)}\mathscr{L}_{\xi_a} \partial_0 K_{km})$ ($T_{km}$ obeys ${}^{(3)}\mathscr{L}_{\xi_a} T_{km} = 0$ because it is duality-invariant) and (ii) $\partial_0({}^{(3)}\mathscr{L}_{\xi_a} \mathscr{E}^k + k_a \mathscr{B}^k)$ $(= {}^{(3)}\mathscr{L}_{\xi_a} \partial_0 \mathscr{E}^k + k_a \partial_0 \mathscr{B}^k) = 0$, $\partial_0({}^{(3)}\mathscr{L}_{\xi_a} \mathscr{B}^k - k_a \mathscr{E}^k)$ $(= {}^{(3)}\mathscr{L}_{\xi_a} \partial_0 \mathscr{B}^k - k_a \partial_0 \mathscr{E}^k) = 0$. Conclude then that (6.2) and (6.3) hold at all times, which easily leads to the desired result.

In the invariant frames $\{dx^0, \omega^a\}$— where $x^0$ is defined by the above gauge conditions—the metric only involves $x^0$. In the same way, the general solution to the symmetry equations (6.3) is

$$\mathscr{E}^a(x^0, \mathbf{x}) = \cos\alpha(\mathbf{x}) \epsilon^a(x^0) - \sin\alpha(\mathbf{x}) \beta^a(x^0),$$
$$\mathscr{B}^a(x^0, \mathbf{x}) = \sin\alpha(\mathbf{x}) \epsilon^a(x^0) + \cos\alpha(\mathbf{x}) \beta^a(x^0), \tag{6.4}$$

where $\epsilon^a$, $\beta^a$ are functions of time only and where $d\alpha = k_a \omega^a$. Without loss of generality, the invariant frame can be taken so that $\alpha = kx^3$ [i.e., $\omega^3 = dx^3$, $k_a = (0,0,k)$].

It results from the above theorem that the dynamical Einstein–Maxwell equations can only restrict the time dependence of $g_{ab}(x^0)$, $\epsilon^a(x^0)$, and $\beta^a(x^0)$, i.e., must be ordinary differential equations for these functions. This is easily checked in the case of the Einstein equations, since $T_{\lambda\mu}[\mathscr{E}^a, \mathscr{B}^b] = T_{\lambda\mu}[\epsilon^a, \beta^b]$ [the spatial dependence (6.4) of $\mathscr{E}^a$, $\mathscr{B}^a$ drops out from the energy-momentum tensor]. As to the dynamical Maxwell equations, they reduce to

$$\dot{Z}^a = \left[((i/2)C^d{}_{bc} Z_d - k_b Z_c)\epsilon^{abc} N/\sqrt{g}\right] + C^a{}_{bc} N^b Z^c + (2a_b + ik_b) N^b Z^a, \tag{6.5}$$

where $Z^a$ are the spatial components of $\mathring{F}^\dagger$,

$$Z^a = \epsilon^a + i\beta^a. \tag{6.6}$$

To completely demonstrate that the application to Bianchi models of Maxwell fields invariant up to a duality roation indeed opens up new nontrivial possibilities, it remains to prove that the constraints do not imply $F = 0$ when $k_a \neq 0$. This can be seen by direct inspection of the constraints, which turn out to be simply algebraic in $g_{ab}$, $K_{ab}$, $\epsilon^a$, and $\beta^a$,

$$(2a_a - ik_a) Z^a = 0, \tag{6.7}$$

$$K_{ab} K^{ab} - K^2 - R + (1/2g)(\epsilon^a \epsilon^b + \beta^a \beta^b) = 0, \tag{6.8a}$$

$$-2K_b{}^c C^b{}_{ac} - 4K_a{}^c a_c = (1/\sqrt{g}) \epsilon_{abc} \epsilon^b \beta^c, \tag{6.8b}$$

where $R\,(g_{ab}, C^c{}_{de})$ is the curvature of the surfaces $x^0 = \text{const}$.

Let us stress that these constraints do not imply that the electromagnetic field is null; all algebraic types are allowed for $F$.

Although the "fictitious field" $(\epsilon^a, \beta^a)$ does not obey the dynamical Maxwell equations because of the $k_a$ term in (6.5), the initial value problem is independent of $k_a$ to a large extent.

**Theorem**: For all class B types, except type III, the constraint (6.7) is equivalent to $a_a Z^a = 0$.

The proof is straightforward since $k_a = ka_a$. The initial value problem is thus obviously independent of $k_a$.

**Theorem:** For types I and II, any solution of the initial value problem with $k_a \neq 0$ is also a solution with $k_a = 0$. Reciprocally, given a solution of the initial value problem with $k_a = 0$, it is possible to find some $k_a \neq 0$ so that Eqs. (6.7)–(6.8) hold.

*Proof*: (i) Type I ($C^a{}_{bc} = 0$)

(6.7) reads $k_a Z^a = 0$. Given $Z^a$, it is always possible to find $k_a \neq 0$ so that $k_a Z^a = 0$.

(ii) Type II ($n^{ab} = \mathrm{diag}(1,0,0)$, $a_a = 0$)

Again, (6.7) reads $k_a Z^a = 0$, but this time, $k_a$ is restricted by $k_a n^{ab} = 0$. Equation (6.8b) implies $\epsilon^2 \beta^3 = \epsilon^3 \beta^2$ so that given a set $(g_{ab}, K_{ab}, \epsilon^a, \beta^a)$ obeying (6.8), one can always find $k_a \neq 0$ solution to $k_1 = 0$, $k_a \epsilon^a = k_a \beta^a = 0$.

We finally note that the cases $k_a \neq 0$ lead, when the electromagnetic field is non-null, to truly new metrics. Indeed the gravitational field determines the electromagnetic field up to a constant duality rotation,[1] whereas the cases $k_a \neq 0$ and $k_a = 0$ differ by a nonconstant duality rotation.[13] Any exhaustive study of electromagnetic Bianchi models must accordingly include the case $k_a \neq 0$.

It is difficult to find exact solutions to the Einstein–Maxwell equations when $k_a \neq 0$ because these models are in general nondiagonal: $k_a$ couples the various components of the electromagnetic field. Noticeable exceptions are models, the diagonality of which results from additional symmetries, as we now pass to discuss.

## VII. L.R.S. BIANCHI MODELS

For definiteness, we consider the L.R.S. type $\mathrm{V/VII}_h$ case,

$$ds^2 = -N^2(x^0)(dx^0)^2 + a^2(x^0)e^{-2x^3}[(dx^1)^2 + (dx^2)^2] + c^2(x^0)(dx^3)^2, \quad (7.1)$$

as it is the only L.R.S. Bianchi model that admits a nontrivial $k_A$. The type V Killing vectors are $\partial_1$, $\partial_2$, and $\partial_3 + x^1\partial_1 + x^2\partial_2$. The generator of the additional isometry is

$$\xi_4 = x^2\partial_1 - x^1\partial_2. \quad (7.2)$$

Taking (6.4) into account, the requirement that the electromagnetic field be invariant under $\xi_4$ up to a duality transformation is equivalent to

$$\begin{aligned} &\epsilon^2 = \bar{k}\beta^1, \quad \beta^2 = -\bar{k}\epsilon^1, \quad -\epsilon^1 = \bar{k}\beta^2, \\ &\beta^1 = \bar{k}\epsilon^2, \quad 0 = \bar{k}\beta^3, \quad 0 = \bar{k}\epsilon^3, \end{aligned} \quad (7.3)$$

where $\bar{k}$ determines a homomorphism of the isotropy subgroup at the origin [generated by $\xi_4$ and isomorphic to SO(2)] to SO(2) and is accordingly restricted to be an integer by global considerations. Actually, it is only when $\bar{k} = 0$ or $\pm 1$ that the equations (7.3) possess a nontrivial solution (let us insist that there is no such restriction on $k$):

$$\begin{aligned} &\bar{k} = 0, \quad \epsilon^1 = \epsilon^2 = \beta^1 = \beta^2 = 0, \\ &\qquad \epsilon^3 \text{ and } \beta^3 \text{ arbitrary;} \end{aligned} \quad (7.4)$$

$$\begin{aligned} &\bar{k} = \epsilon, \quad \epsilon = \pm 1 \quad \epsilon^3 = \beta^3 = 0, \\ &\epsilon^1 = -\epsilon\beta^2, \quad \epsilon^2 = \epsilon\beta^1, \\ &\qquad \beta^1 \text{ and } \beta^2 \text{ arbitrary.} \end{aligned} \quad (7.5)$$

In the first case, the electric and magnetic fields are parallel and point in the third direction. In the second one, the field is null—in agreement with the theorem of Appendix B—and corresponds to a circularly polarized wave propagating along the third axis. Since Gauss' law and the div $\mathscr{B}$-law impose $\epsilon^3 = \beta^3 = 0$, we shall consider from no one that second possibility.

Maxwell equations read, for the field (7.5),

$$\dot{\beta}^2 = N/c(\epsilon\beta^2 - k\beta^1), \quad \dot{\beta}^1 = N/c(\epsilon\beta^1 + k\beta^2). \quad (7.6)$$

In the gauge $N = c$, they can be straightforwardly integrated and yield

$$\begin{aligned} &\beta^1 = E \exp \epsilon x^0 \sin kx^0 = \epsilon\,\epsilon^2, \\ &\beta^2 = E \exp \epsilon x^0 \cos kx^0 = -\epsilon\,\epsilon^1, \end{aligned} \quad (7.7)$$

where $E$ is an integration constant. We have chosen the axes $(x_1, x_2)$ so that $\beta^1 = 0$ and $E > 0$ when $x^0 = 0$. The time scale $x^0$ is related to the proper time $t$ by

$$N\,dx^0 = dt \Leftrightarrow c\,dx^0 = dt. \quad (7.8)$$

When inserted into (6.4), the relation (7.7) leads to

$$\begin{aligned} &-\epsilon\,\mathscr{E}^1 = E \exp \epsilon x^0 \cos k(x^0 - \epsilon x^3) = \mathscr{B}^2, \\ &\epsilon\,\mathscr{E}^2 = E \exp \epsilon x^0 \sin k(x^0 - \epsilon x^3) = \mathscr{B}^1. \end{aligned} \quad (7.9)$$

This represents a wave that propagates in the positive or in the negative $x^3$ direction according to whether $\epsilon$ is equal to $+1$ or $-1$. Its frequency is determined by $|k|$, and its polarization, by the sign of $-k\epsilon$ (positive helicity if $k\epsilon < 0$).

The electromagnetic stress-energy tensor possesses the radiation form and is explicitly given by

$$T_{00} = \frac{E^2}{a^2}\exp 2\epsilon x^0 = T_{33} = -\epsilon T_{03}; \quad (7.10)$$

its other components all vanish.

The nontrivial Einstein equations are equivalent to

$$\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{\dot{a}}{a}\frac{\dot{c}}{c} - 3 = \frac{E^2}{a^2}\exp 2\epsilon x^0, \quad (7.11)$$

$$2\left(\frac{\dot{a}}{a} - \frac{\dot{c}}{c}\right) = -\frac{\epsilon E^2}{a^2}\exp 2\epsilon x^0, \quad (7.12)$$

$$a\ddot{a} + \dot{a}^2 - 2a^2 = 0 \Leftrightarrow \frac{d^2 \ln a}{(dx^0)^2} + 2\left(\frac{d \ln a}{dx^0}\right)^2 - 2 = 0, \quad (7.13)$$

$$\frac{\ddot{c}}{c} - \left(\frac{\dot{c}}{c}\right)^2 + 2\frac{\dot{a}}{a}\frac{\dot{c}}{c} - 2 = \frac{E^2}{a^2}\exp 2\epsilon x^0, \quad (7.14)$$

where we have explicitely used the condition $N = c$. The equation (7.11) is the $G_{00} = T_{00}$ equation, the equation (7.12) is the $R_{03} = T_{03}$ equation, whereas the remaining ones are the $R_{11} = T_{11} = R_{22} = T_{22}$ and the $R_{33} = T_{33}$ equations.

Depending on the sign of $(\dot{a}/a)^2 - 1$, the equation (7.13), which is the same as in vacuum, leads to three possibilities:

$$\begin{aligned} &\text{(ia) } a = A(\sinh 2x^0)^{1/2} \quad (\text{valid for } x^0 > 0) \text{ or} \\ &\qquad \text{(ib) } a = A(-\sinh 2x^0)^{1/2} \quad (x^0 < 0); \end{aligned} \quad (7.15)$$

$$\text{(ii) } a = A(\cosh 2x^0)^{1/2}; \quad (7.16)$$

(iii) $a = Ae^{\epsilon' x^0}, \quad \epsilon' = \pm 1.$ (7.17)

Here, $A$ is a constant of integration which can be set equal to 1 by an appropriate redefinition of $x^1$ and $x^2$ ($x^1 \to Ax^1$, $x^2 \to Ax^2$; this does not modify the structure constants), whereas the origin of $x^0$ has been chosen so that $a(0) = A$ [cases (7.16) and (7.17)] or 0 [case (7.15)].

The constraint equations (7.11) and (7.12) are only compatible with the first and third possibilities and impose (with $A = 1$):

$$\text{Case (7.15):} \quad \text{(ia)} \ \epsilon = -1 \quad E = \sqrt{3},$$

$$\text{(ib)} \ \epsilon = +1 \quad E = \sqrt{3}, \tag{7.18}$$

$$\text{Case (7.17):} \quad \epsilon' = \epsilon, \ \text{no restriction on } E. \tag{7.19}$$

It is then very easy to integrate Eq. (7.12) for $c$. One finds

$$\text{Case (7.15):} \quad \text{(ia)} \ c = Be^{(3/2)x^0}(\sinh 2x^0)^{-1/4}, \tag{7.20}$$

$$\text{(ib)} \ c = Be^{-(3/2)x^0}(-\sinh 2x^0)^{-1/4}, \tag{7.21}$$

$$\text{Case (7.17):} \quad c = B \exp \epsilon(E^2/2 + 1)x^0. \tag{7.22}$$

In both cases, Eq. (7.14) is identically satisfied by the above $a$ and $c$.

## VIII. PROPERTIES OF THE L.R.S. SOLUTIONS

Let us first turn to the solution (7.9), (7.15), and (7.20), with $\epsilon = -1$. The metric reads, explicitly,

$$ds^2 = B^2 e^{3x^0}(\sinh 2x^0)^{-1/2}[-(dx^0)^2 + (dx^3)^2] + (\sinh 2x^0)e^{-2x^3}[(dx^1)^2 + (dx^2)^2] \quad (x^0 > 0). \tag{8.1}$$

It represents an anisotropic universe filled with an electromagnetic wave propagating in the negative $x^3$ direction. This universe expands from an initial singularity located at $x^0 = 0$ (a finite amount of proper time in the past). The singularity is of the "cigar type," with Kasner exponents (2/3, 2/3, $-$1/3). As $x^0 \to \infty$, both $a$ and $c$ increase as $e^{x^0}$ and there is thus "isotropization."

If one takes $\epsilon = +1$, one just gets the time reversed solution, with a singularity in the future. The wave now propagates in the positive $x^3$ direction. We recall that this direction is defined by $a_3 > 0$.

When the electromagnetic wave number vanishes, these solutions reduce to the one described by Ftaclas and Cohen.[14] Note that the stress-energy tensor and hence, the metric, are independent of $k$—actually, the metric is of the "radiation fluid-filled, plane symmetric type." Besides, the solutions with different $k$ (but same metric) can be obtained from one another by a space–time-dependent duality rotation $\beta(x^0 - \epsilon x^3)$, the gradient of which is lightlike and along the direction of propagation of the electromagnetic wave (the field is null).

Although the above metric (8.1) does not possess additional Killing vectors, the solution (7.17), (7.22) is invariant by a seven-dimensional group of motions acting on space–time. Indeed, the change of coordinates

$$e^{\epsilon z x^0} = (z/B)(2uv)^{1/2},$$
$$e^{z x^3} = (z/B)(2u/v)^{1/2}, \tag{8.2}$$
$$x^1 = x^1, \quad x^2 = x^2,$$

with $z \equiv 1 + E^2/2$, brings the metric and the electromagnetic field to the form

$$ds^2 = -2\,du\,dv + v^{2/z}[(dx^1)^2 + (dx^2)^2], \tag{8.3}$$

$$F = \frac{Ev^{1/z-1}}{z}\left[\cos\left(\frac{k}{z}\ln v\right)dv \wedge dx^1 - \epsilon \sin\left(\frac{k}{z}\ln v\right)dv \wedge dx^2\right]. \tag{8.4}$$

The metric (8.3) is conformally flat and represents a Kagan subprojective space (Ref. 7, p. 252), i.e., here, a special type of plane gravitational wave with seven Killing vectors. The electromagnetic field is only strictly invariant under a transitive six-dimensional subgroup ($x\,\partial_y - y\,\partial_x$ never Lie-derives $F$).

The above classes of solutions contain all electromagnetic Bianchi type V universes with local rotational symmetry in which the Maxwell field shares the symmetry of the metric up to a duality transformation.

## IX. BIANCHI MODELS WITH DISCRETE SYMMETRIES

The concept of Maxwell fields invariant up to a duality rotation is also useful for understanding discrete symmetries. Let us consider again the Bianchi type-V case, but this time, without assuming $a = b$,

$$ds^2 = -N^2(x^0)(dx^0)^2 + a^2(x^0)e^{-2x^3}(dx^1)^2 + b^2(x^0)e^{-2x^3}(dx^2)^2 + c^2(x^0)(dx^3)^2. \tag{9.1}$$

The metric possesses the following discrete symmetries [in addition to the $G_3(\text{V})$ group]:

$$\mathscr{F}_1: x^0 \to x^0, \quad x^1 \to -x^1, \quad x^2 \to x^2, \quad x^3 \to x^3, \tag{9.2a}$$

$$\mathscr{F}_2: x^0 \to x^0, \quad x^1 \to x^1, \quad x^2 \to -x^2, \quad x^3 \to x^3, \tag{9.2b}$$

as well as their product

$$\mathscr{R}_3: x^0 \to x^0, \quad x^1 \to -x^1, \quad x^2 \to -x^2, \quad x^3 \to x^3. \tag{9.2c}$$

$\mathscr{R}_3$ preserves the orientation, whereas $\mathscr{F}_1$ and $\mathscr{F}_2$ do not. Conversely, the existence of these discrete symmetries implies the diagonality of the metric. In order to determine the possible electromagnetic, diagonal type-V models, one must thus find all the Maxwell fields invariant up to a duality rotation under the full group $H_3(V) = G_3(V) \cup \{\mathscr{F}_1, \mathscr{F}_2, \mathscr{R}_3\}$ (and their products).

We first turn to the task of determining $\alpha(H_3(V))$. Since both $\mathscr{F}_1$ and $\mathscr{F}_2$ commute with the transformations generated by $\xi_3$, it follows from a property demonstrated in the first appendix that $k_3$ must vanish (together with $k_1$ and $k_2$). In other words, the image of $G_3(V)$ is trivial,

$$\alpha(G_3(V)) = \{0\} \quad (\Leftrightarrow k_a = 0). \tag{9.3}$$

We next note that the product laws $(\mathscr{R}_3)^2 = e$, $\mathscr{F}_1\mathscr{F}_2 = \mathscr{R}_3 = \mathscr{F}_2\mathscr{F}_1$ imply (see Appendix A)

$$2\alpha(\mathscr{R}_3) = 0, \tag{9.4a}$$

$$\alpha(\mathscr{F}_1) - \alpha(\mathscr{F}_2) = \alpha(\mathscr{R}_3). \tag{9.4b}$$

Two cases need to be considered: either $\alpha(\mathscr{R}_3)$ is the identity, or it is half a revolution. To investigate the consequences of the second equation (9.4), we assume that $\alpha(\mathscr{F}_2)$ is the identity, which we can always do by performing an appropriate constant duality rotation $\beta$ on the electromagnetic field ($\alpha(\mathscr{F}_2) \to \alpha(\mathscr{F}_2) - 2\beta$).[15] The relation (9.4b) implies then that $\alpha(\mathscr{F}_1)$ is equal to $\alpha(\mathscr{R}_3)$.

(i) $\alpha(\mathscr{R}_3) = 0, \quad \alpha(\mathscr{F}_1) = 0.$ (9.5a)

The symmetry equations imply

$$\mathscr{E}^1 = \mathscr{B}^1 = \mathscr{E}^2 = \mathscr{B}^2 = 0 = \mathscr{B}^3. \tag{9.5b}$$

Only $\mathscr{E}^3$ can be nonvanishing.

(ii) $\alpha(\mathscr{R}_3) = \alpha(\mathscr{F}_1) = \frac{1}{2}$ (half a revolution). (9.6a)

The symmetry equations imply

$$\mathscr{E}^2 = \mathscr{E}^3 = 0 \quad \mathscr{B}^1 = \mathscr{B}^3 = 0. \tag{9.6b}$$

Only $\mathscr{E}^1$ and $\mathscr{B}^2$ can differ from zero.

Because of the constraint $a_a Z^a = 0$, one must reject the first case. In the second case, that constraint is automatically satisfied. We have thus proved the following theorem:

**Theorem**: In all diagonal type-V Bianchi models filled with a non-null electromagnetic field, the electric and magnetic components $\mathscr{E}^a$ and $\mathscr{B}^a$ are characterized, up to a global duality rotation, by the conditions (9.3) and (9.6).[16]

It is not our purpose here to discuss the integration, in the comoving frame, of the Einstein–Maxwell equations for the above fields. Let us merely mention that solutions do exist, because (9.1) and (9.6) are compatible with the constraints. Moreover, these solutions define Maxwellian involutive structures in the sense of Debever[17]; the two-dimensional abelian group generated by $\partial_1$ and $\partial_2$ is invertible, with $\mathscr{R}_3$ as involution.

The conclusion of this paper is that the concept of Maxwell fields invariant up to a duality rotation is not only mathematically interesting, but also particularly fruitful for understanding some of the properties of solutions to Einstein–Maxwell equations with a group of motions.

## ACKNOWLEDGMENT

The author is grateful to Jacques Demaret for useful conversations.

## APPENDIX A

We consider in this appendix how the formulas of the first section need to be changed when the isomorphism $h$ does not preserve the orientation of space–time.

As is known, $F^*$ is defined in an arbitrary frame $\{\omega^\alpha\}$ as the two-form

$$F^*_{\alpha\beta} = (\epsilon[\omega]/2!\sqrt{-g}) g_{\alpha\lambda} g_{\beta\mu} \epsilon^{\lambda\mu\epsilon\sigma} F_{\epsilon\sigma}. \tag{A1}$$

Here $\epsilon[\omega]$ is $+1$ or $-1$ according to whether the frame $\{\omega^\alpha\}$ has the "right" orientation or not.

From (A1), one infers

$$h^*F^* = \epsilon_h (h^*F)^*, \tag{A2}$$

where $\epsilon_h = +1$ if the isomorphism $h$ preserves the orientation of space-time and $-1$ in the opposite case.

Formula (A2) implies

$$h^*F^* = \epsilon_h(-\sin\alpha(h)F + \cos\alpha(h)F^*) \tag{A3}$$

from which it follows that the composition law reads

$$\alpha(h_1 h_2) = \epsilon_{h_2}\alpha(h_1) + \alpha(h_2). \tag{A4}$$

This leads to

$$\alpha(h^{-1}) = -\epsilon_h \alpha(h) \tag{A5}$$

and

$$\alpha(h_1^{-1} h_2^{-1} h_1 h_2) = \alpha(h_2) + \epsilon_{h_2}\alpha(h_1) - \epsilon_{h_1}\alpha(h_2) - \alpha(h_1). \tag{A6}$$

These relations show that the mapping $\alpha$: $H \to SO(2)$ is in general not a group homomorphism when $H$ possesses elements which do not preserve the orientation.

If $H$ is the direct product of an orientation-preserving, connected, Lie subgroup $G$ with an involutive "reflexion" $s(\epsilon_s = -1, s^2 = e)$, formula (A6) implies that $\alpha(G) = \{0\}$. Indeed, one easily infers from (A6) with $h_2 = h \in G$, $h_1 = s$,

$$0 = 2\alpha(h).$$

Accordingly, $\alpha(h)$ is either the identity or half a revolution. But that second possibility is excluded by the assumption that $G$ is a connected Lie group (and the continuity of $\alpha$).

## APPENDIX B

Let us assume that the isometry group $H$ is multiply transitive on its surfaces of transitivity. In this appendix, $H$ may not be transitive on the space–time manifold. Let $K(P)$ be the isotropy group at $P$, and let $\xi_a$ be the corresponding Killing vectors [we assume that $K(P)$ is at least a one-dimensional Lie group; discrete isotropy subgroups are not considered]. As is well known, the vector fields $\xi_a$ vanish at $P$, but $\xi_a{}^\mu{}_{,\rho}(P) \neq 0$, and the $\xi_a$'s induce a group $K^*$ of transformations of the tangent space at $P$ which is isomorphic to a subgroup of the Lorentz group.

**Theorem**: If one of the $k_a$'s does not vanish, i.e., if $K(P)$ does not belong to the kernel of the homomorphism $\alpha$: $H \to SO(2)$, then, $F$ is a null two-form.

*Proof*: The symmetry equation $\mathscr{L}_{\xi_a} F^\dagger = ik_a F^\dagger$ reads at $P$

$$\Lambda^\rho{}_\alpha[\xi_a] F^\dagger_{\rho\beta} + \Lambda^\rho{}_\beta[\xi_a] F^\dagger_{\alpha\rho} = ik_a F_{\alpha\beta}, \tag{B1}$$

where $\Lambda^\rho{}_\alpha[\xi_a]$ is the infinitesimal generator of the one-parameter subgroup of $K^*$ induced by $\xi_a$. In a suitable orthonormal frame, $\Lambda[\xi_a]$ can be taken to be

$$\Lambda[\xi_a] = \begin{pmatrix} 0 & a & 0 & 0 \\ a & 0 & 0 & -m \\ 0 & 0 & 0 & n \\ 0 & m & -n & 0 \end{pmatrix}. \tag{B2}$$

We can also assume $m = 0$ when $\eta_{\alpha\beta}\Lambda^\beta{}_\rho$ is non-null, or $n = 0$, $|m| = |a|$ when it is null, but, in order to treat both cases simultaneously, we shall not use these simplifications here.

With (B2), formula (B1) becomes

$$ikF^\dagger_{01} - mF^\dagger_{03} = 0,$$
$$ikF^\dagger_{02} - aF^\dagger_{12} + nF^\dagger_{03} = 0,$$
$$ikF^\dagger_{03} - aF^\dagger_{13} + mF^\dagger_{01} - nF^\dagger_{02} = 0, \tag{B3}$$

$$ikF^{\dagger}_{12} - aF^{\dagger}_{02} - mF^{\dagger}_{32} + nF^{\dagger}_{13} = 0,$$
$$ikF^{\dagger}_{13} - aF^{\dagger}_{03} - nF^{\dagger}_{12} = 0,$$
$$ikF^{\dagger}_{23} + mF^{\dagger}_{21} = 0,$$

where we have dropped the index $a$ in $k_a$.

The system (B3) possesses a nonzero solution $F$ only when its determinant, easily evaluated by the Laplace method, vanishes:

$$k^4 + 2k^2(a^2 - m^2 - n^2) + a^4 + m^4 + n^4 - 2m^2a^2 + 2m^2n^2 + 2a^2n^2 = 0. \tag{B4}$$

Since $k^2$ is real, the discriminant of the quadratic (in $k^2$) equation (B4) must be positive.

$$-4a^2n^2 \geqslant 0. \tag{B5}$$

Thus, either $a$ vanishes—in which case (B2) describes a pure rotation and one can also take $n = 0$—or $n$ is equal to zero. But in that latter case, it follows from (B4) that

$$k^2 = m^2 - a^2, \tag{B6}$$

and hence, $|m| > |a| (k \neq 0)$. Thus, by an appropriate Lorentz rotation, one can assume that $a$ vanishes too, and the equations (B3) reduce in both cases to

$$i\epsilon F^{\dagger}_{01} - F^{\dagger}_{03} = 0, \quad F^{\dagger}_{02} = 0,$$
$$i\epsilon F^{\dagger}_{12} - F^{\dagger}_{32} = 0, \quad F^{\dagger}_{13} = 0, \tag{B7}$$

which implies that the electromagnetic field is indeed null.

An alternative derivation of this theorem, somewhat simpler, starts from the equations

$$\mathscr{L}_{\xi_A} I_1 = 2k_A I_2 \quad \mathscr{L}_{\xi_A} I_2 = -2k_A I_1 \tag{B8}$$

for the two invariants $F_{\lambda\mu}F^{\lambda\mu} \equiv I_1$ and $F_{\lambda\mu}F^{*\lambda\mu} \equiv I_2$. These equations clearly show that the electromagnetic field is everywhere null on a surface of transitivity if it is null at one point of that surface. Moreover, since the generators $\xi_a$ of the isotropy group at $P$ vanish at $P$, and since $I_1$ and $I_2$ are scalars, both $\mathscr{L}_{\xi_a} I_1 = \partial_{\xi_a} I_1$ and $\mathscr{L}_{\xi_a} I_2 = \partial_{\xi_a} I_2$ vanish at $P$, which implies

$$k_a I_2 = k_a I_1 = 0. \tag{B9}$$

If $k_a \neq 0$, one infers $I_2 = I_1 = 0$, i.e., the electromagnetic field is null.

As a consequence of this theorem, it follows that all the $k_a$'s associated with isotropy subgroups are zero when the electromagnetic field is everywhere non-null. $F^{\dagger}$ can then be written as

$$F^{\dagger} = e^{i\sigma}\mathring{F}^{\dagger}, \tag{B10}$$

where $\mathring{F}^{\dagger}$ is strictly invariant and where the function $\sigma$ (defined on space–time, not on the group manifold!) obeys

$$\mathscr{L}_{\xi_A}\sigma = k_A. \tag{B11}$$

This is not true when some of the $k_a$'s differ from zero.

[1]C. W. Misner and J. A. Wheeler, "Classical physics as geometry: gravitation, electromagnetism, unquantized charge, and mass as properties of curved empty space," Ann. Phys. (N.Y.) **2**, 525 (1957); M. L. Woolley, "The structure of groups of motions admitted by Einstein–Maxwell space-times," Commun. Math. Phys. **31**, 75 (1973); H. Michalski and J. Wainwright, "Killing vector fields and the Einstein–Maxwell field equations in general relativity," Gen. Relativ. Gravit. **6**, 289 (1975); J. R. Ray and E. L. Thompson, "Space–time symmetries and the complexion of the electromagnetic field," J. Math. Phys. **16**, 345 (1975); C. Hoenselaers, "On the effect of motions on energy-momentum tensors," Prog. Theor. Phys. **59**, 1518 (1978).

[2]R. G. McLenaghan and N. Tariq, "A new solution of the Einstein–Maxwell equations," J. Math. Phys. **16**, 2306–2312 (1975); B. O. J. Tupper, "A class of algebraically general solutions of the Einstein–Maxwell equations for non-null electromagnetic fields, II," Gen. Relativ. Gravit. **7**, 479–486 (1976); see also D. Kramer, H. Stephani, M. Mac Callum, and E. Herlt, *Exact Solutions of Einstein's Field Equations*, Vol. 6 of *Cambridge Monographs on Mathematical Physics* (Cambridge U. P., Cambridge, England, 1980).

[3]P. Forgacs and N. S. Manton, 'Space–time symmetries in gauge theories," Commun. Math. Phys. **72**, 15–35 (1980); M. Henneaux, "Remarks on space–time symmetries and nonabelian gauge fields," J. Math. Phys. **23**, 830–833 (1982).

[4]M. Henneaux, "Gravitational fields, spinor fields and groups of motions," Gen. Relativ. Gravit. **12**, 137–147 (1980); M. Henneaux, "Univers de Bianchi et champs spinoriels," Ann. Inst. Henri Poincaré, **34**, 329–349 (1981); R. T. Jantzen, "Spinor sources in cosmology," J. Math. Phys. **23**, 1137–1146 (1982).

[5]L. Defrise-Carter, "Conformal groups and conformally equivalent isometry groups," Commun. Math. Phys. **40**, 273–282 (1975).

[6]By "function on $H$" is meant a function with values in SO(2) ($\alpha$ and $f$: $H \rightarrow$ SO(2)). $d\alpha$ and $df$ can be viewed as ordinary one-forms since the algebra SO(2) can be identified with $R$. They are not always equal to the gradients of functions $H \rightarrow R$ (it is only locally so). The integral of $df$ denotes the element of SO(2) obtained from the usual integral of $df$ by the standard homomorphism $R \rightarrow R/Z$ ($\equiv$SO(2)).

[7]A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).

[8]A. Barnes, "A class of homogeneous Einstein–Maxwell fields," J. Phys. A: Gen. Phys **11**, 1303–1314 (1978).

[9]M. P. Ryan and L. C. Shepley, "Homogeneous Relativistic Cosmologies," *Princeton Series in Physics* (Princeton U. P., Princeton, N. J., 1975).

[10]D. M. Eardley, "Self-similar spacetimes: geometry and dynamics," Commun. Math. Phys. **37**, 287–309 (1974).

[11]We recall that both $\mathscr{E}^k$ and $\mathscr{B}^k$ behave as $\sqrt{g}$ times a three-dimensional vector under spatial changes of coordinates. We include here the factor $\epsilon(\omega)$ in the definition of $\mathscr{B}^k$ so as to fulfill that condition, where $\epsilon(\omega)$ refers to the orientation of the frame $\{dx^0, dx^k\}$—see formula (A1) of Appendix A (this is somewhat unconventional, but simplifies the present discussion).

[12]K. Kuchař, 'Canonical quantization of cylindrical waves," Phys. Rev. D **4**, 955–986 (1971) (Appendix).

[13]It would thus be wrong to think that one could just get the solutions with different $k_a$ from one another by keeping the metric fixed while "dually rotating" the e.m. field. $k_a$ leaves an indelible imprint on the metric.

[14]C. Ftaclas and J. M. Cohen, "Locally rotationally symmetric cosmological model containing a nonrotationally symmetric electromagnetic field," Phys. Rev. D **18**, 4373–4377 (1978).

[15]More generally, the transformation law of $\alpha(h)$ under a constant duality rotation $\beta$ is: $\alpha(h) \rightarrow \alpha(h) + \epsilon_h\beta - \beta$.

[16]This is not true when the e.m. field is null, as the previous section shows: $\alpha(G_3(V)) \neq \{0\}$ in that case (when $k_3 \neq 0$, the L.R.S. solution is not invariant up to a constant duality rotation under $\mathscr{F}_1, \mathscr{F}_2$).

[17]R. Debever, "Structures pré-maxwelliennes involutives en relativité générale," Bull. Cl. Sci. Acad. R. Belg. **LXII**, 662–677 (1976).

# Noniterative method for constructing many-parameter solutions of the Einstein and Einstein–Maxwell field equations[a)]

Dong-sheng Guo[b)]

*Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616*

(Received 19 July 1983; accepted for publication 9 December 1983)

We present a noniterative method of executing a large class of Kinnersley–Chitre transformations in both the vacuum and the electrovac case. By solving the homogeneous Hilbert problem in the Hauser–Ernst formalism, we generate new many-parameter solutions of the Einstein equations. In the vacuum case, the solution is a natural generalization of the $N$-fold Neugebauer solution, while, in the electrovac case, we have a natural generalization of the $N$-fold Cosgrove solution worked out by Wang, Guo, and Wu.

PACS numbers: 04.20.Jb, 02.70. + d

## I. INTRODUCTION

In recent years many authors have employed Bäcklund transformations[1–3] and Kinnersley–Chitre (K–C) transformations[4–10] in order to generate new solutions of the vacuum and electrovac Einstein field equations. Usually the transformation selected is quite simple and involves only a few parameters, but, by iterating such transformations, solutions with an arbitrary number of parameters can be generated.

In the present paper an alternative approach will be described, in which the K–C transformation selected has an arbitrary number of parameters, and it is applied only once. Starting with Minkowski space as the seed space-time, we first consider the generation of vacuum space-times, and then we turn our attention to the generation of electrovac space-times. In the vacuum case, our new many-parameter solution is a natural generalization of the $N$-fold Neugebauer solution,[2] while in the electrovac case we have a natural generalization of the $N$-fold Cosgrove solution worked out by Wang, Guo, and Wu.[10]

Our method possesses the following features:

(1) The parameters characterizing the transformation are directly related to the coefficients of polynomials in the numerator and denominator of the transformed Ernst potential evaluated on the symmetry axis.

(2) In its simplest vacuum exemplar, our method unifies the Ehlers transformation,[11] Harrison's Bäcklund transformation,[1,3] two types of Hauser transformation,[8,9] and an HKX transformation,[5] while in the electrovac case it unifies the Ehlers transformation, the Cosgrove transformation,[3] and a charged HKX transformation.[6]

(3) By using this method one can more directly obtain a complete symmetry in the parameters characterizing the generated space-time, for one can build it into the characterization of the K–C group element itself. In the iterative method the parameters enter in an ordered way, some with each iteration. The generated space-time does not involve these parameters in a symmetrical fashion, and it is a nontrivial problem to redefine the parameters in such a way as to restore symmetry in the final result.

Our new method, in addition, may provide a way to employ a sequence of exact solutions which in some sense approaches a solution which cannot itself be obtained in closed form because of difficulties in solving the associated homogeneous Hilbert problem (HHP).[7]

## II. VACUUM TRANSFORMATION

In the Hauser–Ernst formalism[6,7] vacuum K–C transformations are represented by $2\times 2$ matrix functions $u(t)$ of a complex parameter $t$, such that

$$\det u(t) = 1, \tag{2.1}$$

$$u^\dagger(t)\epsilon u(t) = \epsilon := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.2}$$

where

$$\begin{pmatrix} 1/t & 0 \\ 0 & 1 \end{pmatrix} u(t) \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix}$$

is holomorphic in an open neighborhood of $t=\infty$. [Note that in Eq. (2.2) $u^\dagger(t)$ stands for the Hermitian conjugate of $u(t^*)$. Because of Eq. (2.1), condition (2.2) may be replaced by the statement that the matrix $u(t)$ is *real* for real values of the parameter $t$. We shall when speaking of $u(t)$ always use the word "real" in this sense.]

Following Cosgrove,[3] we shall introduce a real matrix $\tilde{u}(t)$ such that

$$u(t) = [\det \tilde{u}(t)]^{-1/2}\tilde{u}(t). \tag{2.3}$$

Specifically, we shall choose $\tilde{u}(t)$ of the form

$$\tilde{u}(t) = \begin{pmatrix} \alpha(t) & \beta(t) \\ \gamma(t) & \delta(t) \end{pmatrix}, \tag{2.4}$$

where $\alpha(t)$, $\beta(t)t^{-1}$, $\gamma(t)t$, and $\delta(t)$ are real polynomials in the variable $t^{-1}$. We assume that $\alpha(\infty)\delta(\infty) - \beta(\infty)\gamma(\infty) \neq 0$. Explicitly, we may write

$$\begin{aligned} \alpha(t) &= \alpha_0 + \alpha_{-1}t^{-1} + \cdots + \alpha_{-n}t^{-n}, \\ \beta(t) &= \beta_1 t + \beta_0 + \cdots + \beta_{-n}t^{-n}, \\ \gamma(t) &= \gamma_{-1}t^{-1} + \cdots + \gamma_{-n}t^{-n}, \\ \delta(t) &= \delta_0 + \delta_{-1}t^{-1} + \cdots + \delta_{-n}t^{-n}. \end{aligned} \tag{2.5}$$

Situations in which the four polynomials terminate at different terms will be treated as degenerate cases.

[a)] Research supported in part by National Science Foundation Grant PHY-82-05608.

[b)] The author will submit to the Illinois Institute of Technology a Ph. D. thesis based in part upon material contained in this paper.

 0022-2488/84/072284-09$02.50  

It should be noted that when the seed space-time is Minkowski space the coefficients of the polynomials have a direct interpretation in terms of the new Ernst potential $\mathscr{E}'$ evaluated on the symmetry axis ($z$ axis),[12] where

$$\mathscr{E}' = \frac{(i\alpha_0 - \beta_1) + (i\alpha_{-1} - \beta_0)(2z) + (i\alpha_{-2} - \beta_{-1})(2z)^2 + \cdots}{(\gamma_{-1} + i\delta_0) + (\gamma_{-2} + i\delta_{-1})(2z) + (\gamma_{-3} + i\delta_{-2})(2z)^2 + \cdots}. \tag{2.6}$$

The case $n = 1$, where

$$\tilde{u}(t) = \begin{pmatrix} \alpha_0 + \alpha_{-1}t^{-1} & \beta_1 t + \beta_0 + \beta_{-1}t^{-1} \\ \gamma_{-1}t^{-1} & \delta_0 + \delta_{-1}t^{-1} \end{pmatrix}, \tag{2.7}$$

includes five well-known transformations, the Ehlers transformation,[11]

$$\tilde{u}(t) = \begin{pmatrix} \alpha_0 & \beta_1 t \\ \gamma_{-1}t^{-1} & \delta_0 \end{pmatrix}, \tag{2.8}$$

the Harrison transformation,[1,3]

$$\tilde{u}(t) = \begin{pmatrix} \alpha_0 & \beta_0 \\ \gamma_{-1}t^{-1} & \delta_0 \end{pmatrix}, \tag{2.9}$$

two types of Hauser transformation,[8,9]

$$\tilde{u}(t) = \begin{pmatrix} (\alpha_2 m_1 - m_2\alpha_1) + \frac{1}{2}(\alpha_1 - \alpha_2)t^{-1} & \alpha_1\alpha_2(m_2 - m_1)t \\ (m_1 - m_2)t^{-1} & (\alpha_2 m_2 - \alpha_1 m_1) + \frac{1}{2}(\alpha_1 - \alpha_2)t^{-1} \end{pmatrix}, \tag{2.10}$$

where $\alpha_1$, $\alpha_2$, $m_1$, and $m_2$ are real parameters, and

$$\tilde{u}(t) = i\begin{pmatrix} (\alpha^* m - \alpha m^*) + \frac{1}{2}(\alpha - \alpha^*)t^{-1} & \alpha\alpha^*(m^* - m)t \\ (m - m^*)t^{-1} & (\alpha^* m^* - \alpha m) + \frac{1}{2}(\alpha - \alpha^*)t^{-1} \end{pmatrix}, \tag{2.11}$$

where $\alpha$ and $m$ are complex parameters, and an HKX transformation,[5] which corresponds to the special case when

$$(\alpha_{-1}\delta_0 + \alpha_0\delta_{-1} - \beta_0\gamma_{-1})^2 = 4(\alpha_0\delta_0 - \beta_1\gamma_{-1})(\alpha_{-1}\delta_{-1} - \beta_{-1}\gamma_{-1}) \tag{2.12}$$

is satisfied.

The homogeneous Hilbert problem consists of finding $2\times 2$ matrix potentials $F'(t)$ and $X_-(t)$ satisfying

$$F'(t)\tilde{u}(t)F(t)^{-1} = [\det \tilde{u}(t)]^{1/2}X_-, \tag{2.13}$$

such that regarded as functions of the complex parameter $t$, these matrices possess, respectively, the space-time-dependent singularities of $F(t)$ (the $F$-potential of the seed space-time) and the fixed singularities of $\tilde{u}(t)$. It is further required that $F'(0) = F(0) = i\epsilon$.

Because of the polynomial form assumed for $\tilde{u}(t)$ it can be shown that

$$F'(t)\tilde{u}(t)F(t)^{-1} = A_0 + A_{-1}t^{-1} + \cdots + A_{-n}t^{-n}, \tag{2.14}$$

where the constant matrix coefficients $A_i$ $(i = 1,\ldots,n)$ remain to be determined. Indeed, $A_{-n}$ is easily found to be given by

$$A_{-n} = \lim_{t\to 0} F'(t)\tilde{u}(t)t^n F(t)^{-1} = \begin{pmatrix} \delta_{-n} & -\gamma_{-n} \\ -\beta_{-n} & \alpha_{-n} \end{pmatrix}. \tag{2.15}$$

The new $F$-potential can be obtained from

$$F'(t) = (A_0 + A_{-1}t^{-1} + \cdots + A_{-n}t^{-n}) \times F(t)\begin{pmatrix} \delta(t) & -\beta(t) \\ -\gamma(t) & \alpha(t) \end{pmatrix} \times [\alpha(t)\delta(t) - \beta(t)\gamma(t)]^{-1}. \tag{2.16}$$

The equation

$$\alpha(t)\delta(t) - \beta(t)\gamma(t) = 0 \tag{2.17}$$

has $2n$ roots. We shall denote them by $t = t_1, t_2, \ldots, t_{2n}$, and temporarily we shall assume they are all distinct. None is at $t = \infty$. The condition that $F'(t)$ not have any of the fixed singularities associated with $u(t)$ implies that

$$(A_0 + A_{-1}t_i^{-1} + \cdots + A_{-n}t_i^{-n}) \times F(t_i)\begin{pmatrix} \delta(t_i) & -\beta(t_i) \\ -\gamma(t_i) & \alpha(t_i) \end{pmatrix} = 0 \tag{2.18}$$

for $i = 1,2,\ldots,2n$. By using the relation

$$\det\begin{pmatrix} \delta(t_i) & -\beta(t_i) \\ -\gamma(t_i) & \alpha(t_i) \end{pmatrix} = 0 \quad (i = 1,2,\ldots,2n), \tag{2.19}$$

we can express Eq. (2.18) in the following alternate form:

$$\begin{aligned} &[(A_0)_{33} + (A_{-1})_{33}t_i^{-1} + \cdots + (A_{-(n-1)})_{33}t_i^{-(n-1)}]T_i \\ &\quad + (A_0)_{34} + (A_{-1})_{34}t_i^{-1} + \cdots + (A_{-(n-1)})_{34}t_i^{-(n-1)} \\ &\quad = -\delta_{-n}t_i^{-n}T_i + \gamma_{-n}t_i^{-n}, \\ &[(A_0)_{43} + (A_{-1})_{43}t_i^{-1} + \cdots + (A_{-(n-1)})_{43}t_i^{-(n-1)}]T_i \\ &\quad + (A_0)_{44} + (A_{-1})_{44}t_i^{-1} + \cdots + (A_{-(n-1)})_{44}t_i^{-(n-1)} \\ &\quad = \beta_{-n}t_i^{-n}T_i - \alpha_{-n}t_i^{-n}, \end{aligned} \tag{2.20}$$

where

$$T_i := \frac{F_{33}(t_i)\delta(t_i) - F_{34}(t_i)\gamma(t_i)}{F_{43}(t_i)\delta(t_i) - F_{44}(t_i)\gamma(t_i)} \quad (i = 1,2,\ldots,2n) \tag{2.21}$$

are known quantities.

The solution of Eqs. (2.20) can be expressed in the form

$$(A_{-j})_{33}=\Delta^j_{33}/\Delta,\quad (A_{-j})_{34}=\Delta^j_{34}/\Delta \qquad (j=0,1,\dots,n-1),$$
$$(A_{-j})_{43}=\Delta^j_{43}/\Delta,\quad (A_{-j})_{44}=\Delta^j_{44}/\Delta \tag{2.22}$$

where

$$\Delta^j_{33}=\begin{vmatrix} T_1 & t_1^{-1}T_1 & \cdots & t_1^{-(j-1)}T_1 & -t_1^{-n}(\delta_{-n}T_1-\gamma_{-n}) & \cdots t_1^{-(n-1)}T_1 & 1 & t_1^{-1} & \cdots & t_1^{-(n-1)} \\ & & \cdots & & & & & \cdots & & \\ T_{2n} & t_{2n}^{-1}T_{2n} & \cdots & t_{2n}^{-(j-1)}T_{2n} & -t_{2n}^{-n}(\delta_{-n}T_{2n}-\gamma_{-n}) & \cdots t_{2n}^{-(n-1)}T_{2n} & 1 & t_{2n}^{-1} & \cdots & t_{2n}^{-(n-1)} \end{vmatrix},$$

$$\Delta^j_{34}=\begin{vmatrix} T_1 & t_1^{-1}T_1 & \cdots & t_1^{-(n-1)}T_1 & 1 & t_1^{-1} & \cdots t_1^{-(j-1)} & -t_1^{-n}(\delta_{-n}T_1-\gamma_{-n}) & \cdots & t_1^{-(n-1)} \\ & & \cdots & & & & & \cdots & & \\ T_{2n} & t_{2n}^{-1}T_{2n} & \cdots & t_{2n}^{-(n-1)}T_{2n} & 1 & t_{2n}^{-1} & \cdots t_{2n}^{-(j-1)} & -t_{2n}^{-n}(\delta_{-n}T_{2n}-\gamma_{-n}) & \cdots & t_{2n}^{-(n-1)} \end{vmatrix},$$

$$\Delta^j_{43}=\begin{vmatrix} T_1 & t_1^{-1}T_1 & \cdots & t_1^{-(j-1)}T_1 & t_1^{-n}(\beta_{-n}T_1-\alpha_{-n}) & \cdots t_1^{-(n-1)}T_1 & 1 & t_1^{-1} & \cdots & t_1^{-(n-1)} \\ & & \cdots & & & & & \cdots & & \\ T_{2n} & t_{2n}^{-1}T_{2n} & \cdots & t_{2n}^{-(j-1)}T_{2n} & t_{2n}^{-n}(\beta_{-n}T_{2n}-\alpha_{-n}) & \cdots t_{2n}^{-(n-1)}T_{2n} & 1 & t_{2n}^{-1} & \cdots & t_{2n}^{-(n-1)} \end{vmatrix},$$

$$\Delta^j_{44}=\begin{vmatrix} T_1 & t_1^{-1}T_1 & \cdots & t_1^{-(n-1)}T_1 & 1 & t_1^{-1} & \cdots t_1^{-(j-1)} & t_1^{-n}(\beta_{-n}T_1-\alpha_{-n}) & \cdots & t_1^{-(n-1)} \\ & & \cdots & & & & & \cdots & & \\ T_{2n} & t_{2n}^{-1}T_{2n} & \cdots & t_{2n}^{-(n-1)}T_{2n} & 1 & t_{2n}^{-1} & \cdots t_{2n}^{-(j-1)} & t_{2n}^{-n}(\beta_{-n}T_{2n}-\alpha_{-n}) & \cdots & t_{2n}^{-(n-1)} \end{vmatrix},$$

and

$$\Delta=\begin{vmatrix} T_1 & t_1^{-1}T_1 & \cdots & t_1^{-(n-1)}T_1 & 1 & t_1^{-1} & \cdots & t_1^{-(n-1)} \\ T_{2n} & t_{2n}^{-1}T_{2n} & \cdots & t_{2n}^{-(n-1)}T_{2n} & 1 & t_{2n}^{-1} & \cdots & t_{2n}^{-(n-1)} \end{vmatrix}.$$

From the new $F$-potential we can easily obtain the new $H$-potential using the formula

$$H'=\left.\frac{dF'(t)}{dt}\right|_{t=0}. \tag{2.23}$$

Thus we obtain

$$H'=\left[A_{-n}H+A_{-(n-1)}\Omega-\Omega\begin{pmatrix}\alpha_{-(n-1)} & \beta_{-(n-1)}\\ \gamma_{-(n-1)} & \delta_{-(n-1)}\end{pmatrix}\right]\begin{pmatrix}\alpha_{-n} & \beta_{-n}\\ \gamma_{-n} & \delta_{-n}\end{pmatrix}^{-1}, \tag{2.24}$$

where $H$ is the $H$-potential of the seed space-time, and

$$\Omega:=i\epsilon=\begin{pmatrix}0 & i\\ -i & 0\end{pmatrix}.$$

As an example, we shall work out the case $n=1$ explicitly. In this case the determinants are given by

$$\Delta^0_{33}=\begin{vmatrix}-t_1^{-1}(\delta_{-1}T_1-\gamma_{-1}) & 1\\ -t_2^{-1}(\delta_{-1}T_2-\gamma_{-1}) & 1\end{vmatrix},$$
$$\Delta^0_{34}=\begin{vmatrix}T_1 & -t_1^{-1}(\delta_{-1}T_1-\gamma_{-1})\\ T_2 & -t_2^{-1}(\delta_{-1}T_2-\gamma_{-1})\end{vmatrix},$$
$$\Delta^0_{43}=\begin{vmatrix}t_1^{-1}(\beta_{-1}T_1-\alpha_{-1}) & 1\\ t_2^{-1}(\beta_{-1}T_2-\alpha_{-1}) & 1\end{vmatrix},$$
$$\Delta^0_{44}=\begin{vmatrix}T_1 & t_1^{-1}(\beta_{-1}T_1-\alpha_{-1})\\ T_2 & t_2^{-1}(\beta_{-1}T_2-\alpha_{-1})\end{vmatrix}, \tag{2.25}$$

and $\Delta=T_1-T_2$, where

$$T_i=[F_{33}(t_i)\delta(t_i)-F_{34}(t_i)\gamma(t_i)]\,[F_{43}(t_i)\delta(t_i)-F_{44}(t_i)\gamma(t_i)]^{-1}\quad (i=1,2). \tag{2.26}$$

The $A$ matrices are given by

$$A_0=\frac{1}{T_1-T_2}\begin{pmatrix}\delta_{-1}(t_2^{-1}T_2-t_1^{-1}T_1)+\gamma_{-1}(t_1^{-1}-t_2^{-1}) & T_1T_2\delta_{-1}(t_1^{-1}-t_2^{-1})+\gamma_{-1}(T_1t_2^{-1}-T_2t_1t_1^{-1})\\ -\beta_{-1}(t_2^{-1}T_2-t_1^{-1}T_1)-\alpha_{-1}(t_1^{-1}-t_2^{-1}) & -T_1T_2\beta_{-1}(t_1^{-1}-t_2^{-1})-\alpha_{-1}(T_1t_2^{-1}-T_2t_1t_1^{-1})\end{pmatrix} \tag{2.27}$$

and

$$A_{-1}=\begin{pmatrix}\delta_{-1} & -\gamma_{-1}\\ -\beta_{-1} & \alpha_{-1}\end{pmatrix}. \tag{2.28}$$

From the quadratic equation

$$(\alpha_0\delta_0-\beta_1\gamma_{-1})t^2+(\alpha_{-1}\delta_0+\alpha_0\delta_{-1}-\beta_0\gamma_{-1})t+(\alpha_{-1}\delta_{-1}-\beta_{-1}\gamma_{-1})=0 \tag{2.29}$$

the roots $t_1$ and $t_2$ are easily obtained.

Let us now see how the Harrison transformation can be treated as a degenerate case of the above. In this case $\tilde{u}(t)$ is given by Eq. (2.9). Let $\alpha_{-1}, \beta_{-1}, \beta_1, \gamma_0$, and $\delta_{-1} \to 0$. Then the quadratic equation (2.28) becomes

$$\alpha_0\delta_0 t^2 - \beta_0\gamma_{-1}t = 0. \tag{2.30}$$

Hence the roots are

$$t_1 = \beta_0\gamma_{-1}/\alpha_0\delta_0, \quad t_2 = 0. \tag{2.31}$$

After taking the limits we get

$$A_0 = \begin{pmatrix} \delta_0 - i\mathscr{E}\gamma_{-1} & \gamma_{-1}t_1^{-1} - T_1(\delta_0 - i\mathscr{E}\gamma_{-1}) \\ -\beta_0 & \beta_0 T_1 \end{pmatrix},$$
$$A_{-1} = \begin{pmatrix} 0 & -\gamma_{-1} \\ 0 & 0 \end{pmatrix}, \tag{2.32}$$

where $\mathscr{E}$ is the Ernst potential of the seed space-time. Hence the new $F$-potential is

$$F'(t) = \begin{pmatrix} \delta_0 - i\mathscr{E}\gamma_{-1} & \gamma_{-1}(t_1^{-1} - t^{-1}) - T_1(\delta_0 - i\mathscr{E}\gamma_{-1}) \\ -\beta_0 & \beta_0 T_1 \end{pmatrix}$$
$$\times F(t)\begin{pmatrix} \delta_0 & -\beta_0 \\ -\gamma_{-1}t^{-1} & \alpha_0 \end{pmatrix}[\alpha_0\delta_0 - \beta_0\gamma_{-1}t^{-1}]^{-1}. \tag{2.33}$$

This result is in agreement with the result quoted in the paper of Cosgrove,[3] provided we choose $\alpha_0 = \delta_0 = 1$.

## III. ELECTROVAC TRANSFORMATION

The K–C group element can be defined as a $3\times 3$ matrix function $u(t)$ of the complex variable $t$ subject to the following conditions:

$$\det u(t) = 1,$$
$$u^\dagger(t)\mathfrak{E}(t)u(t) = \mathfrak{E}(t) := \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -it/2 \end{pmatrix}, \tag{3.1}$$

where

$$\begin{pmatrix} 1/t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} u(t) \begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is holomorphic in an open neighborhood of $t = \infty$.

Because of the large number of constraints imposed upon $u(t)$ by Eqs. (3.1), it is not immediately obvious how to generalize the procedure which we employed in the vacuum case. Following Hauser and Ernst,[6,7] we find that it is advantageous to switch from the $t$-plane to the so-called $\tau$-plane representation of the K–C group. We have

$$v^\dagger(\tau)i\mathfrak{E}v(\tau) = i\mathfrak{E}, \quad \det v(\tau) = 1,$$

where

$$i\mathfrak{E} := \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \tag{3.2}$$

and

$$v(\tau) := \begin{pmatrix} 1/t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} u(t) \begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$
$$\tau = 1/2t.$$

$v(\tau)$ must be holomorphic in an open neighborhood of $\tau = 0$.

Now, following Cosgrove,[3] we shall introduce a $\tilde{v}$ such that

$$\tilde{v}(\tau)/[\det \tilde{v}(\tau)]^{1/3} = v(\tau). \tag{3.3}$$

One then automatically satisfies $\det v(\tau) = 1$, while

$$\tilde{v}^\dagger(\tau)i\mathfrak{E}\tilde{v}(\tau) = i\mathfrak{E}[\det \tilde{v}(\tau)]^{1/3}[\det \tilde{v}(\tau^*)]^{1/3*}. \tag{3.4}$$

We shall consider the case in which $\tilde{v}(\tau)$ is a polynomial in $\tau$, i.e.,

$$\tilde{v}(\tau) = v_0 + v_1\tau + v_2\tau^2 + \cdots + v_n\tau^n. \tag{3.5}$$

We shall also stipulate that

$$v_0^\dagger i\mathfrak{E}v_0 = f_0 i\mathfrak{E}, \tag{3.6}$$

where $v_0, v_1, \ldots, v_n$ are $3\times 3$ constant matrices and the real constant $f_0 \neq 0$. Then the holomorphy requirement for $v(\tau)$ at $\tau = 0$ is satisfied automatically.

We can see from the left side of Eq. (3.4) that

$$[\det \tilde{v}(\tau)]^{1/3}[\det \tilde{v}(\tau)]^{1/3*} = f(\tau) \tag{3.7}$$

must be a real polynomial in $\tau$:

$$f(\tau) = f_0 + f_1\tau + f_2\tau^2 + \cdots + f_{2n}\tau^{2n} \quad (f_0 \neq 0). \tag{3.8}$$

Equations (3.7) and (3.8) show that $f(\tau)$ has $2n$ nonvanishing roots, either complex conjugate pairs or doubly repeated real roots.

Combining Eqs. (3.4), (3.7), and (3.8), we obtain the following relations involving $v_0, v_1, \ldots, v_n$:

$$\begin{aligned} &v_0^\dagger i\mathfrak{E}v_0 = f_0 i\mathfrak{E} \quad (f_0 \neq 0), \\ &v_1^\dagger i\mathfrak{E}v_0 + v_0^\dagger i\mathfrak{E}v_1 = f_1 i\mathfrak{E}, \\ &\cdots \\ &\cdots \\ &\sum_{i+j=k} v_i^\dagger i\mathfrak{E}v_j = f_k i\mathfrak{E}, \\ &\cdots \\ &\cdots \\ &v_n^\dagger i\mathfrak{E}v_n = f_{2n} i\mathfrak{E}. \end{aligned} \tag{3.9}$$

These equations are completely equivalent to the single equation

$$\tilde{v}^\dagger(\tau)i\mathfrak{E}\tilde{v}(\tau) = i\mathfrak{E}f(\tau). \tag{3.10}$$

We developed a technique involving projection matrices which can be used to solve the general case. We shall now describe this projection matrix technique.

Let $m$ and $m^*$ be a pair of complex conjugate roots of $f(\tau) = 0$. Then we have

$$\tilde{v}^\dagger(m)i\mathfrak{E}\tilde{v}(m) = 0,$$
$$\tilde{v}^\dagger(m^*)i\mathfrak{E}\tilde{v}(m^*) = 0. \tag{3.11}$$

We can always find projection matrices $P_m$ and $P_{m^*}$, and nonsingular matrices $W_m$ and $W_{m^*}$, such that

$$P_m W_m = \tilde{v}(m), \quad P_m \tilde{v}(m) = \tilde{v}(m),$$
$$P_{m^*} W_{m^*} = \tilde{v}(m^*), \quad P_{m^*}\tilde{v}(m^*) = \tilde{v}(m^*). \tag{3.12}$$

It follows from Eqs. (3.11) and (3.12) that we can find $P_m$ and $P_{m^*}$ satisfying the following equations:

$$P^{\dagger}_{m^*} i\mathfrak{G} P_m = 0,$$
$$P^2_m = P_m, \tag{3.13}$$
$$P^2_{m^*} = P_{m^*}.$$

We are dealing with a three-dimensional linear space. Some solutions of Eqs. (3.13) can be written according to the following types:

$$(1) \quad P_m = h_1 h_1^{\dagger} i\mathfrak{G}, \quad h_1^{\dagger} i\mathfrak{G} h_1 = 1, \quad P_{m^*} = I - h_1 h_1^{\dagger} i\mathfrak{G}; \tag{3.14}$$

$$(2) \quad P_m = h_1 h_2^{\dagger} i\mathfrak{G}, \quad P_{m^*} = h_1 h_2^{\dagger} i\mathfrak{G}; \tag{3.15}$$

$$(3) \quad P_m = h_1 h_2^{\dagger} i\mathfrak{G}, \quad P_{m^*} = h_1 h_2^{\dagger} i\mathfrak{G} + h_3 h_3^{\dagger} i\mathfrak{G}, \quad \text{or} \quad P_{m^*} = h_3 h_3^{+} i\mathfrak{G}; \tag{3.16}$$

$$(4) \quad P_m = h_1 h_1^{\dagger} i\mathfrak{G}, \quad h_1^{\dagger} i\mathfrak{G} h_1 = 1, \quad P_{m^*} = h_2 h_2^{\dagger} i\mathfrak{G}, \quad h_2^{\dagger} i\mathfrak{G} h_2 = 1, \quad h_2^{\dagger} i\mathfrak{G} h_1 = 0. \tag{3.17}$$

$h_1$, $h_2$, and $h_3$ are column matrices. For type 2 and type 3 they satisfy

$$(h_1 h_2 h_3) i\mathfrak{G} \begin{pmatrix} h_2^{\dagger} \\ h_1^{\dagger} \\ h_3^{\dagger} \end{pmatrix} = I.$$

For type 2 one can use the pair of conditions

$$(P_m - I)\tilde{v}(m) = 0,$$
$$(P_m - I)\tilde{v}(m^*) = 0, \tag{3.18}$$

and analogous equations corresponding to other roots, to solve for $v_1, v_2, ..., v_n$. One finds that solutions exist when $m = m^*$. We shall defer the discussion of such repeated real roots until later. We shall at this time concentrate upon the case of type 1 projection matrices with nonreal roots, where

$$P_m P_{m^*} = 0, \quad P_m + P_{m^*} = I, \tag{3.19}$$

and $P_m$ and $P_{m^*}$ are given by Eqs. (3.14).

Equations (3.12) and (3.19) give us

$$P_{m^*}\tilde{v}(m) = 0,$$
$$P_m \tilde{v}(m^*) = 0, \tag{3.20}$$

and, therefore,

$$P_{m^*}\tilde{v}(m) + P_m \tilde{v}(m^*) = 0. \tag{3.21}$$

Let us define a matrix

$$E := (1/m)I + igg^{\dagger} i\mathfrak{G}, \tag{3.22}$$

where

$$g := h_1 [i(m^* - m)/mm^*]^{1/2}. \tag{3.23}$$

It follows from the previously assumed normalization of $h_1$ that

$$g^{\dagger} i\mathfrak{G} g = i(1/m - 1/m^*).$$

Equation (3.21) may be replaced by

$$E^{n-1} v_1 + E^{n-2} v_2 + \cdots + E v_{n-1} + v_n = -E^n v_0, \tag{3.24}$$

where $E^p$ means the $p$th power of $E$.

If we have $n$ pairs of nonreal roots,

$$m_1, m_1^*, m_2, m_2^*, ..., m_n, m_n^*,$$

we can introduce $n$ matrices

$$E_i = (1/m_i)I + i g_i g_i^{\dagger} i\mathfrak{G} \quad (i = 1,2,...,n), \tag{3.25}$$

where $g_i$ satisfies

$$g_i^{\dagger} i\mathfrak{G} g_i = i(1/m_i - 1/m_i^*) \quad (i = 1,2,...,n). \tag{3.26}$$

In this way we obtain

$$\begin{aligned} &E_1^{n-1} v_1 + E_1^{n-2} v_2 + \cdots + v_n = -E_1^n v_0, \\ &\cdots \\ &\cdots \\ &E_n^{n-1} v_1 + E_n^{n-2} v_2 + \cdots + v_n = -E_n^n v_0, \end{aligned} \tag{3.27}$$

or

$$\begin{pmatrix} E_1^{n-1} & E_1^{n-2} & \cdots & E_1 & I \\ E_2^{n-1} & E_2^{n-2} & \cdots & E_2 & I \\ & & \cdots & & \\ E_n^{n-1} & E_n^{n-2} & \cdots & E_n & I \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{pmatrix}, \tag{3.28}$$

where $Q_i = -E_i^n v_0$ $(i = 1,2,..,n)$. By direct calculation we know that

$$E_i^p = (E_i - (1/m_i)I) \frac{m_i^p - m_i^{*p}}{(m_i^* m_i)^{p-1}(m_i - m_i^*)} + \frac{1}{m_i^p} I. \tag{3.29}$$

Equation (3.28) can be solved by several methods, namely, the determinant method, inverse matrix method, and Gaussian elimination.

The solution of Eq. (3.28) can be written in the form

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} E_1^{n-1} & E_1^{n-2} & \cdots & E_1 & I \\ E_2^{n-1} & E_2^{n-2} & \cdots & E_2 & I \\ & & \cdots & & \\ E_n^{n-1} & E_n^{n-2} & \cdots & E_n & I \end{pmatrix}^{-1} \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_n \end{pmatrix}. \tag{3.30}$$

It turns out that the result is even valid in the case of repeated real roots, although the method of proof is different. From Eqs. (3.19) and (3.20) we know that the matrix $\tilde{v}(m) - \tilde{v}(m^*)$ is a rank 3 matrix which has an inverse. We may write

$$\tilde{v}(m) - \tilde{v}(m^*) = (m - m^*)v_1 + (m^2 - m^{*2})v_2 + \cdots + (m^n - m^{*n})v_n. \quad (3.31)$$

Hence, the matrix

$$[\tilde{v}(m) - \tilde{v}(m^*)]/(m - m^*) = v_1 + (m + m^*)v_2 + \cdots$$

also has an inverse. Since $\tilde{v}(\tau)$ is a polynomial, holomorphic at $\tau = m$, $\dot{\tilde{v}}(m) := d\tilde{v}(\tau)/d\tau|_{\tau = m}$ does not depend on the direction of approach as one takes the limit

$$\lim_{(m - m^*)\to 0} \frac{\tilde{v}(m) - \tilde{v}(m^*)}{m - m^*} = \dot{\tilde{v}}(m) \quad (m \text{ real}). \quad (3.32)$$

This shows that, for real $m$, $\dot{\tilde{v}}(m)$ has an inverse.

From Eq. (3.10) we know that for real $m$ the following equations should be satisfied:

$$\tilde{v}(m)^\dagger i\mathfrak{E}\tilde{v}(m) = 0,$$
$$\dot{\tilde{v}}(m)^\dagger i\mathfrak{E}\tilde{v}(m) + \tilde{v}(m)^\dagger i\mathfrak{E}\dot{\tilde{v}}(m) = 0. \quad (3.33)$$

Define a $3\times 3$ matrix

$$r := \tilde{v}(m)\dot{\tilde{v}}(m)^{-1},$$

i.e., (3.34)

$$r\dot{\tilde{v}}(m) = \tilde{v}(m).$$

Then $r$ obeys the following equations:

$$r^\dagger i\mathfrak{E} + i\mathfrak{E}r = 0,$$
$$r^2 = 0. \quad (3.35)$$

The complete solution of Eqs. (3.35) is

$$r = ih'h'^\dagger i\mathfrak{E},$$
$$h'^\dagger i\mathfrak{E}h' = 0, \quad (3.36)$$

where $h'$ is an arbitrary column matrix.

As in the complex case, we define

$$E = (1/m)I + igg^\dagger i\mathfrak{E}.$$

Here $g := h'/m$. Then Eq. (3.34) is equivalent to

$$E^{n-1}v_1 + E^{n-2}v_2 + \cdots + Ev_{n-1} + v_n = -E^n v_0,$$

and $g$ still satisfies the relation

$$g_i^\dagger i\mathfrak{E}g_i = i(1/m_i - 1/m_i^*) \quad (i = 1,2,..,n).$$

In this way we generalize Eqs. (3.25), (3.26), and (3.28) so that the roots may be real or complex (or even infinite, as we shall see later).

Our final result is given by

$$\tilde{v}(\tau) = v_0 - (\tau I, \tau^2 I, \ldots, \tau^n I) \times \begin{pmatrix} E_1^{n-1} & E_1^{n-2} & \cdots & E_1 & I \\ E_2^{n-1} & E_2^{n-2} & \cdots & E_2 & I \\ & & \cdots & & \\ E_n^{n-1} & E_n^{n-2} & \cdots & E_n & I \end{pmatrix}^{-1} \begin{pmatrix} E_1^n \\ E_2^n \\ \vdots \\ E_n^n \end{pmatrix} v_0, \quad (3.37)$$

where $E_i = (1/m_i)I + ig_i g_i^\dagger i\mathfrak{E}, g_i^\dagger i\mathfrak{E}g_i = i(1/m_i - 1/m_i^*)$. As an example, for $n = 1$ we have

$$\tilde{v}(\tau) = [I - \tau((1/m)I + igg^\dagger i\mathfrak{E})]v_0, \quad (3.38)$$

where $g$ satisfies

$$g^\dagger i\mathfrak{E}g = i(1/m - 1/m^*).$$

$m$ can be complex, real, or infinite. When $m$ is complex, we get the Cosgrove transformation.[3] His original form is equivalent to

$$\tilde{v}(\tau) = [-mI + (m - m^*)hh^\dagger i\mathfrak{E}] + I\tau, \quad h^\dagger i\mathfrak{E}h = 1, \quad (3.39)$$

which corresponds to choosing

$$v_0 = -((1/m)I + igg^\dagger i\mathfrak{E})^{-1}.$$

When $m$ is real, we get a charged HKX transformation.[6] When $m$ is infinite, we get a degenerate case of the HKX transformation.

Here we shall show how to treat a simple degenerate case. Let us consider the case of one infinite root, say $m_n = \infty$. Then, we have

$$E_n = ig_n g_n^\dagger i\mathfrak{E}, \; g_n^\dagger i\mathfrak{E}g_n = 0, \; E_n^2 = E_n^3 = \cdots = E_n^n = 0. \quad (3.40)$$

Equations (3.27) reduce to the following:

$$E_1^{n-1}v_1 + \cdots + E_1 v_{n-1} + v_n = -E_1^n v_0,$$
$$\cdots$$
$$E_{n-1}^{n-1}v_1 + \cdots + E_{n-1}v_{n-1} + v_n = -E_{n-1}^n v_0, \quad (3.41)$$
$$E_n v_{n-1} + v_n = 0.$$

In particular, for $n = 2$, one has

$$E_1 v_1 + v_2 = -E_1^2 v_0,$$
$$E_2 v_1 + v_2 = 0. \quad (3.42)$$

The solution of Eqs. (3.42) is given by

$$v_1 = (E_2 - E_1)^{-1}E_1^2 v_0,$$
$$v_2 = E_2(E_1 - E_2)^{-1}E_1^2 v_0. \quad (3.43)$$

Thus we obtain a transformation with nontrivial structure:

$$\tilde{v}(\tau) = [I + (I - E_2\tau)(E_2 - E_1)^{-1}E_1^2\tau]V_0. \quad (3.44)$$

One may check that this result indeed satisfies Eqs. (3.9) and (3.10).

We shall solve the HHP in the $\tau$-plane. We assume that

$$\lim_{\tau\to\infty} \begin{pmatrix} 1/2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} v_n \begin{pmatrix} 2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = v_n'' \quad (3.45)$$

exists. In the event this condition is not satisfied, one can perform a simple Ehlers transformation to make it true. After solving the HHP, we can use the inverse Ehlers transformation to construct the solution for the desired case.

In the nondegenerate case $v_n^{-1}$ is the Ehlers transformation

$$v_n^\dagger i\mathfrak{E}v_n = f_{2n} i\mathfrak{E} \quad (f_{2n} \neq 0), \quad (3.46)$$

$$\tilde{v}(\tau)' = \tilde{v}(\tau)v_n^{-1} = v_0' + v_1'\tau + \cdots + I\tau^n. \quad (3.47)$$

In the degenerate case

$$v_n^\dagger i\mathfrak{E}v_n = 0. \quad (3.48)$$

(i) Suppose $v_n$ can be diagonalized, so that

$$v_n = (h_1 h_2 h_3)\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}(h_1 h_2 h_3)^{-1}. \quad (3.49)$$

Then

$$\begin{pmatrix} h_1^\dagger \\ h_2^\dagger \\ h_3^\dagger \end{pmatrix} v_n^\dagger i\mathfrak{G} v_n (h_1 h_2 h_3)$$

$$= \begin{pmatrix} \lambda_1^* \lambda_1 h_1^\dagger i\mathfrak{G} h_1 & \lambda_1^* \lambda_2 h_1^\dagger i\mathfrak{G} h_2 & 0 \\ \lambda_2^* \lambda_1 h_2^\dagger i\mathfrak{G} h_1 & \lambda_2^* \lambda_2 h_2^\dagger i\mathfrak{G} h_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.50}$$

If $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$, then

$$\begin{pmatrix} h_1^\dagger i\mathfrak{G} h_1 & h_1^\dagger i\mathfrak{G} h_2 \\ h_2^\dagger i\mathfrak{G} h_1 & h_2^\dagger i\mathfrak{G} h_2 \end{pmatrix} = 0.$$

Thus

$$\det \begin{pmatrix} h_1^\dagger \\ h_2^\dagger \\ h_3^\dagger \end{pmatrix} i\mathfrak{G}(h_1 h_2 h_3) = 0,$$

which contradicts the assumption that $(h_1 h_2 h_3)$ is nonsingular. Hence at least one of $\lambda_1$ and $\lambda_2$ must vanish. We can always arrange it so that $\lambda_1 = 0$. If $\lambda_2 \neq 0$, then $h_2^\dagger i\mathfrak{G} h_2 = 0$. In the linear subspace spanned by $h_1$ and $h_3$, we can always choose $h_1'$ and $h_3'$ such that

$$h_1'^\dagger i\mathfrak{G} h_1' = 0, \quad h_3'^\dagger i\mathfrak{G} h_3' = \tfrac{1}{2}.$$

After normalizing, we can always choose a basis $h_1''$, $h_2''$, $h_3''$ such that

$$\begin{pmatrix} h_1''^\dagger \\ h_2''^\dagger \\ h_3''^\dagger \end{pmatrix} i\mathfrak{G}(h_1'' h_2'' h_3'') = i\mathfrak{G}, \tag{3.51}$$

$$v_n h_1'' = 0.$$

Therefore,

$$v_n (h_1'' h_2'' h_3'') = (0 h_2'' h_3'') = v_n' \tag{3.52}$$

and

$$\lim_{\tau \to \infty} \begin{pmatrix} 1/2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} v_n' \begin{pmatrix} 2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = v_n''$$

exists.

(ii) Supposing $v_n$ can be expressed in the canonical form

$$v_n = (h_1 h_2 h_3) \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 0 \end{pmatrix} (h_1 h_2 h_3)^{-1}.$$

Then

$$v_n h_1 = \lambda h_1, \quad v_n h_2 = h_1 + \lambda h_2, \quad v_n h_3 = 0. \tag{3.53}$$

It follows that

$$\begin{pmatrix} h_1^\dagger \\ h_2^\dagger \\ h_3^\dagger \end{pmatrix} v_n^\dagger i\mathfrak{G} v_n (h_1 h_2 h_3) = \begin{pmatrix} M & 0 \\ 0 & 0 \end{pmatrix},$$

where

$$M = \begin{pmatrix} \lambda\lambda^* h_1^\dagger i\mathfrak{G} h_1 & \lambda\lambda^* h_1^\dagger i\mathfrak{G} h_2 + \lambda^* h_1^\dagger i\mathfrak{G} h_1 \\ \lambda\lambda^* h_2^\dagger i\mathfrak{G} h_1 & \lambda\lambda^* h_2^\dagger i\mathfrak{G} h_2 + h_1^\dagger i\mathfrak{G} h_1 \\ + \lambda h_1^\dagger i\mathfrak{G} h_1 & + \lambda^* h_2^\dagger i\mathfrak{G} h_1 + \lambda h_1^\dagger i\mathfrak{G} h_2 \end{pmatrix}.$$

If $\lambda \neq 0$, then

$$h_1^\dagger i\mathfrak{G} h_1 = h_1^\dagger i\mathfrak{G} h_2 = h_2^\dagger i\mathfrak{G} h_1 = h_2^\dagger i\mathfrak{G} h_2 = 0,$$

which contradicts the assumption that $(h_1 h_2 h_3)$ is nonsingular. Therefore, we conclude that $\lambda = 0$ and $h_1^\dagger i\mathfrak{G} h_1 = 0$.

As in the first case, we can construct an $(h_1, h_2, h_3)$ which satisfies

$$\begin{pmatrix} h_1^\dagger \\ h_2^\dagger \\ h_3^\dagger \end{pmatrix} i\mathfrak{G}(h_1 h_2 h_3) = i\mathfrak{G}.$$

The Ehlers transformation $(h_1, h_2, h_3)$ results in

$$v_n (h_1 h_2 h_3) = (0 h_1 0) = v_n'. \tag{3.54}$$

It then follows that

$$\lim_{\tau \to \infty} \begin{pmatrix} 1/2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} v_n' \begin{pmatrix} 2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = v_n''.$$

(iii) Suppose $v_n$ can be expressed in the canonical form

$$v_n = (h_1 h_2 h_3) \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda \end{pmatrix} (h_1 h_2 h_3)^{-1}. \tag{3.55}$$

Then

$$\begin{pmatrix} h_1^\dagger \\ h_2^\dagger \\ h_3^\dagger \end{pmatrix} v_n^\dagger i\mathfrak{G} v_n (h_1 h_2 h_3) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda\lambda^* h_3^\dagger i\mathfrak{G} h_3 \end{pmatrix}.$$

If $\lambda = 0$, we get the same case as in (ii). If $\lambda \neq 0$, then $h_3^\dagger i\mathfrak{G} h_3 = 0$. We can always choose $h_1' = \alpha h_3$, $h_2'$, and $h_3'$ such that

$$\begin{pmatrix} h_1'^\dagger \\ h_2'^\dagger \\ h_3'^\dagger \end{pmatrix} i\mathfrak{G}(h_1' h_2' h_3') = i\mathfrak{G},$$

$$v_n h_1' = 0.$$

Thus

$$v_n (h_1' h_2' h_3') = (0\, v_n h_2'\, v_n h_3') = v_n'$$

and

$$\lim_{\tau \to \infty} \begin{pmatrix} 1/2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} v_n' \begin{pmatrix} 2\tau & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = v_n'' \tag{3.56}$$

exists.

Assuming condition (3.45), we shall attempt to solve the HHP, which has the following form in the $\tau$-plane:

$$P'(\tau)\tilde{v}(\tau)P(\tau)^{-1} = [\det \tilde{v}(\tau)]^{1/3} Y_+(\tau), \tag{3.57}$$

where

$$P(\tau) = F(t) \begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad t = \frac{1}{2\tau}, \tag{3.58}$$

and

$$P'(\tau) = F'(t) \begin{pmatrix} t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.59}$$

are the $P$ potential of the seed space-time and the transformed space-time, respectively. At $\tau=\infty$ the limiting form of the $P$-potential is given by

$$\lim_{\tau\to\infty} P'(\tau)\begin{pmatrix}2\tau & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix} = \Omega := \begin{pmatrix}0 & i & 0\\ -i & 0 & 0\\ 0 & 0 & 1\end{pmatrix}. \quad (3.60)$$

The right-hand side of Eq. (3.57) is holomorphic in a neighborhood of $\tau=0$. Therefore, we may write

$$P'(\tau)\tilde{v}(\tau)P(\tau)^{-1} = C_0 + C_1\tau + C_2\tau^2 + \cdots + C_n\tau^n + \cdots. \quad (3.61)$$

However, when we take the limit

$$\lim_{\tau\to\infty}\frac{P'(\tau)\tilde{v}(\tau)P(\tau)^{-1}}{\tau^n} = \Omega v_n''\Omega \quad (3.62)$$

under the assumption (3.45), we know the expansion (3.61) has to terminate at the $n$th term. Thus,

$$P'(\tau)\tilde{v}(\tau)P(\tau)^{-1} = C_0 + C_1\tau + \cdots + C_n\tau^n, \quad (3.63)$$

where

$$C_n = \Omega v_n''\Omega. \quad (3.64)$$

By using Eq. (3.10), we can express the new $P$ potential in the form

$$P'(\tau) = (C_0 + C_1\tau + \cdots + C_n\tau^n)P(\tau)(i\mathfrak{S})^{-1}\tilde{v}^\dagger(\tau)i\mathfrak{S}/f(\tau). \quad (3.65)$$

The $P$-potential given by Eq. (3.65) should not have poles where the roots of $f(\tau)$ are located. When $m$, $m^*$ are nonreal roots, we have

$$\begin{aligned}&(C_0 + C_1 m + \cdots + C_n m^n)B(m) = 0,\\ &(C_0 + C_1 m^* + \cdots + C_n m^{*n})B(m^*) = 0,\end{aligned} \quad (3.66)$$

where

$$B(\tau) := P(\tau)(i\mathfrak{S})^{-1}\tilde{v}^\dagger(\tau). \quad (3.67)$$

Equivalently, one may write

$$C_0 S_0(m) + C_1 S_1(m) + \cdots + C_{n-1}S_{n-1}(m) = -C_n S_n(m), \quad (3.68)$$

where

$$S_k(m) = m^k B(m) - m^{*k}B(m^*). \quad (3.69)$$

For the case of repeated real roots $m$, we have

$$\begin{aligned}&(C_1 + 2C_2 m + \cdots nC_n m^{n-1})B(m)\\ &\quad + (C_0 + C_1 m + \cdots + C_n m^n)\dot{B}(m) = 0.\end{aligned} \quad (3.70)$$

If for real $m$ we define

$$S_k(m) := km^{k-1}B(m) + m^k\dot{B}(m), \quad (3.71)$$

then Eq. (3.68) again follows.

In summation, for any selected pair of roots $m$, $m^*$, we have

$$\begin{aligned}&C_0S_0(m_1) + C_1S_1(m_1) + \cdots + C_{n-1}S_{n-1}(m_1)\\ &\quad = -C_nS_n(m_1),\\ &\qquad\cdots\\ &C_0S_0(m_n) + C_1S_1(m_n) + \cdots + C_{n-1}S_{n-1}(m_n)\\ &\quad = -C_nS_n(m_n),\end{aligned} \quad (3.72)$$

where

$$S_k(m) = \begin{cases} m^kB(m) - m^{*k}B(m^*) & \text{(when } m \text{ is not real)}\\ km^{k-1}B(m) + m^k\dot{B}(m) & \text{(when } m \text{ is real).}\end{cases} \quad (3.73)$$

Equation (3.72) can also be written in the form

$$(C_0C_1\cdots C_{n-1})\begin{pmatrix} S_0(m_1) & \cdots & S_0(m_n)\\ & \cdots & \\ S_{n-1}(m_1) & \cdots & S_{n-1}(m_n)\end{pmatrix} = (R_1R_2\cdots R_n), \quad (3.74)$$

where

$$R_k = -C_nS_n(m_k) \quad (k = 1,\ldots,n). \quad (3.75)$$

As we did when we identified the group element, we can solve the above linear system in several ways. The element $(C_k)_{pq}$ $(p,q = 3,4,5)$ of the matrix $C_k$ $(k = 1,\ldots,n-1)$ can be written

$$(C_k)_{pq} = D_{kpq}/D, \quad (3.76)$$

$$D = \det\begin{pmatrix} S_0(m_1) & \cdots & S_0(m_n)\\ & \cdots & \\ S_{n-1}(m_1) & \cdots & S_{n-1}(m_n)\\ & & \\ S_0(m_1) & \cdots & S_0(m_n)\end{pmatrix},$$

$$\cdots \quad (3.77)$$

$$D_{kpq} = \det\begin{pmatrix} S_{k-1}(m_1) & \cdots & S_{k-1}(m_n)\\ R_{kpq}(m_1) & \cdots & R_{kpq}(m_n)\\ & \cdots & \\ S_{n-1}(m_1) & \cdots & S_{n-1}(m_n)\end{pmatrix},$$

where $R_{kpq}(m_i)$ is defined as a $3\times 3$ matrix having the same elements as $S_k(m_i)$ except that the $q$th row of $S_k(m_i)$ is replaced by the $p$th row of $R_i$.

We can also express the solution in the form

$$(C_0C_1\cdots C_{n-1}) = -(C_nS_n(m_1)\cdots C_nS_n(m_n)) \times\begin{pmatrix} S_0(m_1) & \cdots & S_0(m_n)\\ & \cdots & \\ S_{n-1}(m_1) & \cdots & S_{n-1}(m_n)\end{pmatrix}^{-1}. \quad (3.78)$$

The final result for the transformed $P$ potential is

$$P'(\tau) = \left[-(C_nS_n(m_1)\cdots C_nS_n(m_n)) \times\begin{pmatrix} S_0(m_1) & \cdots & S_0(m_n)\\ & \cdots & \\ S_{n-1}(m_1) & \cdots & S_{n-1}(m_n)\end{pmatrix}^{-1} \times\begin{pmatrix} I\\ \vdots\\ I\tau^{n-1}\end{pmatrix} + C_n\tau^n\right]P(\tau)\tilde{v}(\tau)^{-1}, \quad (3.79)$$

where

$$S_k(m) = \begin{cases} m^kB(m) - m^{*k}B(m^*) & \text{(when } m \text{ is not real)}\\ km^{k-1}B(m) + m^k\dot{B}(m) & \text{(when } m \text{ is real)}\end{cases},$$

$$B(m) = P(m)(i\mathfrak{S})^{-1}\tilde{v}^\dagger(m).$$

The new $F$ potential can be obtained by using the relation (3.64). In the Hauser–Ernst formalism,[6] the $H$ potential which characterizes the space-time, and the $\varphi$ potential which characterizes the electromagnetic field, are related to the $F$ potential by

$$F^{(1)}:=\frac{dF}{dt}\bigg|_{t=0}=\begin{pmatrix}H & \varphi\\ 2iL & 2iK\end{pmatrix}. \tag{3.80}$$

Thus, for the new space-time, we have

$$F'^{(1)}=\begin{pmatrix}H' & \varphi'\\ 2iL' & 2iK'\end{pmatrix} = [2C_{n-1}\Omega + C_nF^{(1)} - 2^n\Omega u_{-(n-1)}]v_n''^{-1}, \tag{3.81}$$

where $v_n''$ is given by Eq. (3.45), the constant matrix $C_n$ and the $t$-independent matrix $C_{n-1}$ are given by Eqs. (3.64) and (3.76), and $u_{-(n-1)}$ is a constant matrix given by

$$u_{-(n-1)}=\lim_{t\to 0}\frac{d\,[u(t)t^n]}{dt} = \lim_{t\to 0}\frac{d}{dt}\left[\begin{pmatrix}t & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}V\left(\frac{1}{2t}\right)\begin{pmatrix}1/t & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}t^n\right]. \tag{3.82}$$

In this way Eq. (3.81) yields the Ernst potentials corresponding to a new solution of the Einstein equations.

What we have presented in this paper is a general procedure for solving a quite large class of problems. By using this general technique, one should be able to work out explicitly the Ernst potentials, the metric components, and the electromagnetic field quantities for any given case which is of interest.

Although we worked out the $n = 1$ electrovac transformation explicitly, neither we nor Cosgrove have yet discovered a K–C transformation which generates directly the charged Kerr–NUT solution with $a^2 + e^2 < m^2$.

## ACKNOWLEDGMENTS

I should like to thank Professor F. J. Ernst and Professor I. Hauser for their constant encouragement and help in completing this work, including the assistance of Professor Ernst with the English composition. Finally, I should like to thank Y. Chen and W. Gai for useful discussions.

[1]B. K. Harrison, Phys. Rev. Lett. **411**, 1197 (1978).
[2]G. Neugebauer, J. Phys. A: Math. Gen. (Letters) **12**, L67 (1979); **13**, L19 (1980).
[3]C. M. Cosgrove, J. Math. Phys. **22**, 2624 (1981).
[4]W. Kinnersley and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926, 2037 (1978).
[5]C. Hoenselaers, W. Kinnersley, and B. Xanthopoulos, Phys. Rev. Lett. **42**, 481 (1979); J. Math. Phys. **20**, 2530 (1979).
[6]I. Hauser and F. J. Ernst, Phys. Rev. **20**, 362, 1783 (1979).
[7]I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126, 1418 (1980).
[8]I. Hauser, private communication.
[9]D. S. Guo and F. J. Ernst, J. Math. Phys. **23**, 1359 (1982).
[10]S. K. Wang, H. Y. Guo, and K. Wu, preprint, Academia Sinica (Beijing), 1982.
[11]J. Ehlers, *Les theories relativistes de la gravitation* (CNRS, Paris, 1959).
[12]I. Hauser and F. J. Ernst, J. Math. Phys. **22**, 1051 (1981).

# Pure radiation fields admitting nontrivial null symmetries

L. Radhakrishna and N. Ibohal Singh
*Department of Mathematics, Shivaji University, Kolhapur 416004, India*

(Received 8 November 1982; accepted for publication 17 October 1983)

The sixteen types of geometrical symmetries corresponding to the continuous groups of collineations and motions generated by a null vector $n$ are considered. The common propagation vector of a pure electromagnetic radiation field and a pure gravitational radiation field is chosen to be $n$. For such radiation fields all the sixteen symmetries are expressed in terms of the Newman–Penrose (NP) spin coefficients and then it is shown that when $n$ is a gradient field there are only five independent symmetries. The existence of these five nontrivial null symmetries is established by finding exact solutions of Einstein–Maxwell field equations when $n$ satisfies freedom conditions and when $l$ of the NP null tetrad $(l, m, \bar{m}, n)$ is shear-free. Thus a class of space-times of pure radiation fields that admit (i) a Ricci collineation which is not a curvature collineation (CC), (ii) a CC which is not a special curvature collineation (SCC), (iii) a SCC which is not an affine collineation (AC), (iv) an AC which is not a motion, and (v) a motion is determined.

PACS numbers: 04.20.Jb, 04.40. + c

## 1. INTRODUCTION

In the general theory of relativity, all the symmetries of the stress tensor need not be shared by the metric tensor. Hence, a dynamical symmetry need not necessarily be a geometrical symmetry. For instance in a non-null electrovac universe, the electromagnetic field tensor has four symmetries while the metric tensor has only three.[1] In this context Katzin *et al.*[2] have introduced the concept of collineations for a systematic study of the various types of geometrical symmetries admitted by the gravitational fields due to distributions of matter in motion. Out of the sixteen symmetries, which consist of motions and collineations, the curvature tensor representing the permanent gravitational field explicitly enters in collineations. The role of continuous groups of collineations to generate conservation laws of a dynamical system in the general theory of relativity has been described by Davis and his collaborators in a series of papers.[2–4] This work is analogous to Petrov's classification of gravitational field based on the continuous groups of motions.[5]

The sixteen geometrical symmetries[2,3] under investigation are enumerated in Sec. 3. Curvature collineations (CC) in the absence of free gravitational field (conformally flat spaces) have been studied by Levine and Katzin,[4] while CC's in the absence of a matter field (empty spaces) have been investigated by Collinson.[6] Tariq and Tupper[7] have shown that every CC admitted by null source-free Einstein–Maxwell fields ia a conformal motion except when the Weyl tensor is of Petrov type $N$ or $O$. McIntosh[8] has surveyed the work on CC's from the point of view of generating exact solutions of Einstein's field equations and opined that there exist very few space-times compatible with these symmetries since a CC is almost always a conformal motion. Halford *et al.*[9] have investigated Petrov-type $N$ vacuum metrics which admit nontrivial CC's. Pure gravitational-radiation fields amenable for motions and conformal motions in Einstein spaces are considered by Leroy.[10] Lukacs *et al.*[11] have confined themselves to null motions in electrovacuum. Homothetic motions in vacuum and perfect fluid space-times have been analyzed by McIntosh.[12] For a thermodynamical magnetofluid admitting a RC with respect to the flow vector, Asgekar and Date[13] have shown that (a) the stream lines are expansion-free if and only if the heat-flux vector is divergence-free, and (b) the stream lines are geodesic if and only if the heat-flux vector remains invariant along the system of stream lines. Radhakrishna and Rao[14] have established the compatibility of RC with respect to irrotational flow in perfect fluids collapsing by neutrino emission. Hall[15] has shown that a CC is necessarily a homothetic motion in (a) all non-null as well as (b) all null source free electromagnetic fields with Petrov types of gravitational fields except possibly type $N$ or $O$, and (c) all perfect fluids except possibly the stiff matter. "Actually, in practice, it will be difficult to distinguish proper RC, proper CC and proper SCC in given situations where the explicit form of the symmetry vector is not determined."[16]

In this paper we consider the free gravitational field to be the transverse gravitational wave zone which can be identified as Petrov-type $N$ (Ref. 17) or as a self-conjugate gravitational field[18] or as a pure radiation field.[19] Thus we confine our attention to the interaction of the pure electromagnetic-radiation field and the pure gravitational-radiation field with the common propagation vector $n$. For brevity these two interacting radiation felds are referred as the PR fields. Such PR fields have been discussed by McIntosh and Halford[20] and also Hall.[15] However, they do not obtain exact solutions of Einstein field equations and they are concerned with the one symmetry—the curvature collineation. The aim of this paper is to transcribe all the tensor relations characterizing the sixteen symmetries into the "amazingly useful" Newman–Penrose formalism in the case of pure electromagnetic radiation fields with pure gravitaional-radiation fields and to identify the nontrivial ones. The infinitesimal generator of each one of the sixteen symmetries is chosen to be $n$ of the Newman–Penrose (NP) null tetrad $(l^a, m^a, \bar{m}^a, n^a)$.[21]

Section 2 deals with the relations governing the symmetries of the PR fields. The enumeration of commutative relations, NP equations as well as Bianchi identities gives the

 0022-2488/84/072293-08$02.50  

complete mathematical characterization of the interaction of a pure gravitational-radiation field with a pure electromagnetic-radiation field. Section 3 contains the NP spin-coefficient characterization of all the symmetries for the PR fields. It also demonstrates the reduction of the sixteen symmetries to five independent symmetries for the fields in question and thus the nontrivial null symmetries are identified when the symmetric vector is a gradient field. Section 4 determines the space-times corresponding to these five nontrivial null symmetries, under certain conditions. The Newman–Penrose expressions for $n^a{}_{;cb}$ are given in an appendix.

Katzin *et al.*[2] call a RC which does not degenerate to CC as a proper RC, while Halford, *et al.*,[9] call such a RC as a nontrivial RC. In this paper we follow the nomenclature of Halford, *et al.*

## 2. RELATIONS GOVERNING THE PR FIELDS

### The electromagnetic radiation fields

In NP formalism, the Maxwell scalar characterizing the null electromagnetic field with the propagation vector $n$ is

$$\phi_1 = \phi_2 = 0, \phi \equiv \phi_0 \neq 0,$$

and the electromagnetic field tensor[22] is

$$F_{ab} = -\bar{\phi} n_{[a} m_{b]} - \phi n_{[a} \bar{m}_{b]}, \tag{2.1}$$

where $2n_{[a}n_{b]} = n_a m_b - n_b m_a$. In the absence of the charge-current vector ($J^a = 0$), the Maxwell equations for the null electromagnetic field are

$$\nu = \lambda = 0, \tag{2.2a}$$

$$\Delta\phi = (2\gamma - \mu)\phi, \tag{2.2b}$$

$$\bar{\delta}\phi = (2\alpha - \pi)\phi. \tag{2.2c}$$

From (2.2b) we have

$$\Delta(\phi\bar{\phi}) = [2(\gamma + \bar{\gamma}) - (\mu + \bar{\mu})]\phi\bar{\phi}. \tag{2.2d}$$

### The pure gravitational radiation fields

The Weyl scalar $\psi$ characterizing a pure gravitational-radiation field with the propagation vector $n$ is given by

$$\psi_1 = \psi_2 = \psi_3 = \psi_4 = 0, \ \psi \equiv \psi_0 \neq 0.$$

### The Bianchi identities for PR fields

We designate the Bianchi identities as $B_1, B_2, \ldots, B_{11}$. the enumeration follows the sequence of equations given in Flaherty.[23] The nontrivial Bianchi identities $B_3$, $B_9$ and $B_2$, $B_1$ for PR fields yield, respectively,

$$\mu = 0, \tag{2.3a}$$

$$\Delta\psi = 4\gamma\,\psi, \tag{2.3b}$$

$$\bar{\delta}\psi - \chi\delta(\phi\bar{\phi}) = (4\alpha - \pi)\psi + (\bar{\pi} - 2\bar{\alpha} - 2\beta)\chi\phi\bar{\phi}, \tag{2.3c}$$

where $\chi = -8\pi G/c^4$ is a universal constant, and the Ricci scalars for an electromagnetic field are

$$\phi_{AB} = \chi\phi_A\bar{\phi}_B \ (A,B = 0,1,2).$$

*Remarks*: From the definition of the optical scalars for $n$ (after scaling $n$: $\gamma + \bar{\gamma} = 0$), viz.,

divergence: $n^a{}_{;a} = (\mu + \bar{\mu})$,

twist: $i[2n_{[a;b]}n^{a;b}]^{1/2} = -(\mu - \bar{\mu})$,

shear: $\frac{1}{2}[2n_{(a;b)}n^{a;b} - (n^a{}_{;a})^2]^{1/2} = \lambda\bar{\lambda}$,

where

$$\begin{aligned} n_{a;b} = {} & \nu m_a l_b - \lambda m_a m_b - \mu m_a \bar{m}_b + \pi m_a n_b + \bar{\nu}\bar{m}_a l_b \\ & - \bar{\lambda}\bar{m}_a\bar{m}_b - \bar{\mu}\bar{m}_a m_b + \bar{\pi}\bar{m}_a n_b \\ & - (\gamma + \bar{\gamma})n_a l_b + (\alpha + \bar{\beta})n_a m_b \\ & + (\bar{\alpha} + \beta)n_a\bar{m}_b - (\epsilon + \bar{\epsilon})n_a n_b, \end{aligned} \tag{2.4}$$

we infer that all the optical scalars for the PR fields vanish by virtue of (2.2a) and (2.3a), and so we have the reduced expression

$$\begin{aligned} n_{a;b} = {} & \pi m_a n_b + \bar{\pi}\bar{m}_a n_b + (\alpha + \bar{\beta})n_a m_b \\ & + (\bar{\alpha} + \beta)n_a\bar{m}_b - (\epsilon + \bar{\epsilon})n_a n_b. \end{aligned}$$

### Ricci identities for the PR fields

The NP equations which are equivalent to the Ricci identities with the conditions (2.1a) and (2.3a) are

$$\begin{aligned} D\rho - \bar{\delta}\kappa = {} & \rho^2 + \sigma\bar{\sigma} + (\epsilon + \bar{\epsilon})\rho - \bar{\kappa}\tau \\ & - \kappa(3\alpha + \bar{\beta} - \pi) + \chi\phi\bar{\phi}, \end{aligned} \tag{2.5a}$$

$$\begin{aligned} D\sigma - \delta\kappa = {} & (\rho + \bar{\rho})\sigma + (3\epsilon - \bar{\epsilon})\sigma \\ & - (\tau - \bar{\pi} + \bar{\alpha} + 3\beta)\kappa + \psi, \end{aligned} \tag{2.5b}$$

$$\begin{aligned} D\tau - \Delta\kappa = {} & (\tau + \bar{\pi})\rho + (\bar{\tau} + \pi)\sigma \\ & + (\epsilon - \bar{\epsilon})\tau - (3\gamma + \bar{\gamma})\kappa, \end{aligned} \tag{2.5c}$$

$$\begin{aligned} D\alpha - \bar{\delta}\epsilon = {} & (\rho + \bar{\epsilon} - 2\epsilon)\alpha + \beta\bar{\sigma} - \bar{\beta}\epsilon \\ & - \bar{\kappa}\gamma + (\epsilon + \rho)\pi, \end{aligned} \tag{2.5d}$$

$$D\beta - \delta\epsilon = (\alpha + \pi)\sigma + (\bar{\rho} - \bar{\epsilon})\beta - \gamma\kappa - (\bar{\alpha} - \bar{\pi})\epsilon, \tag{2.5e}$$

$$\begin{aligned} D\gamma - \Delta\epsilon = {} & (\tau + \bar{\pi})\alpha + (\bar{\tau} + \pi)\beta - (\epsilon + \bar{\epsilon})\gamma \\ & - (\gamma + \bar{\gamma})\epsilon + \tau\pi, \end{aligned} \tag{2.5f}$$

$$\bar{\delta}\pi = -\pi^2 - (\alpha - \bar{\beta})\pi, \tag{2.5g}$$

$$\delta\pi = -\pi\bar{\pi} + \pi(\bar{\alpha} - \beta), \tag{2.5h}$$

$$\Delta\pi = -(\gamma - \bar{\gamma})\pi, \tag{2.5i}$$

$$\delta\rho - \bar{\delta}\sigma = \rho(\bar{\alpha} + \beta) - \sigma(3\alpha - \bar{\beta}) + (\rho - \bar{\rho})\tau, \tag{2.5j}$$

$$\delta\alpha - \bar{\delta}\beta = \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta + \gamma(\rho - \bar{\rho}), \tag{2.5k}$$

$$\delta\gamma - \Delta\beta = (\tau - \bar{\alpha} - \beta)\gamma - \beta(\gamma - \bar{\gamma}), \tag{2.5l}$$

$$\delta\tau - \Delta\sigma = (\tau + \beta - \bar{\alpha})\tau - (3\gamma - \bar{\gamma})\sigma, \tag{2.5m}$$

$$\Delta\rho - \bar{\delta}\tau = (\bar{\beta} - \alpha - \bar{\tau})\tau + (\gamma + \bar{\gamma})\rho, \tag{2.5n}$$

$$\Delta\alpha - \bar{\delta}\gamma = \bar{\gamma}\alpha + (\bar{\beta} - \bar{\tau})\gamma. \tag{2.5o}$$

The commutation relations are

$$[\Delta, D] = (\gamma + \bar{\gamma})D + (\epsilon + \bar{\epsilon})\Delta - (\tau + \bar{\pi})\bar{\delta} - (\bar{\tau} + \pi)\delta, \tag{2.6a}$$

$$[\delta, D] = (\bar{\alpha} + \beta - \bar{\pi})D + \kappa\Delta - \sigma\bar{\delta} - (\bar{\rho} + \epsilon - \bar{\epsilon})\delta, \tag{2.6b}$$

$$[\delta, \Delta] = (\tau - \bar{\alpha} - \beta)\Delta + (\bar{\gamma} - \gamma)\delta, \tag{2.6c}$$

$$[\bar{\delta}, \delta] = (\bar{\rho} - \rho)\Delta - (\bar{\alpha} - \beta)\bar{\delta} - (\bar{\beta} - \alpha)\delta. \tag{2.6d}$$

The Weyl conformal tensor characterizing the transverse gravitational field is

$$C_{dcb}{}^{a} = -2\,\mathrm{Re}(\psi U_{dc} U_b{}^a), \tag{2.7}$$

where the bivector is

$$U_{dc} = 2\overline{m}_{[d}\, n_{c]}.$$

The Ricci tensor for the source-free null electromagnetic field with the propagation vector $n$ is

$$R_{ab} = -\tfrac{1}{2}\chi\phi\bar{\phi} n_a n_b. \tag{2.8}$$

For the PR fields the curvature tensor is given by

$$R_{dcb}{}^{a} = -2\,\mathrm{Re}(\psi U_{dc} U_b{}^a) - \tfrac{1}{4}\chi\phi\bar{\phi}(\overline{U}_{dc} U_b{}^a + U_{dc}\overline{U}_b{}^a). \tag{2.9}$$

Similarly one can obtain the Weyl projective curvature tensor

$$W_{dcb}{}^{a} = R_{dcb}{}^{a} - \tfrac{1}{3}(g_d{}^a R_{cb} - g_{db} R_c{}^a) \tag{2.10}$$

by using (2.8) and (2.9). We observe that for the PR fields

$$F_{ab} n^a = R_{ab} n^a = 0, \tag{2.11}$$

$$R_{dcb}{}^{a} n^d = C_{dcb}{}^{a} n^d = C_{dcb}{}^{a} n^b = 0.$$

Equation (2.11) implies that $n$ is the common propagation vector for the gravitational-radiation field as well as the electromagnetic-radiation field.

## 3. NULL SYMMETRIES IN TERMS OF SPIN COEFFICIENTS

### (i) Ricci Collineation (RC)

A space-time is said to admit RC if there exists a vector field $\xi^a$, such that

$$\mathscr{L}_\xi R_{ab} = 0, \tag{3.1}$$

where $\mathscr{L}_\xi$ denotes the Lie derivative with respect to $\xi^a$. Singh, Radhakrishna, and Sharan[24] have studied these relations (3.1) for cylindrically-symmetric space-times permeated by a source-free non-null electromagnetic field and have shown that these are all purely electric. For studying the PR fields we choose the symmetry vector $\xi^a = n^a$. Using the expression $n_{a;b}$ (2.4) and (2.2a), (2.3a), we get from (2.8)

$$\mathscr{L}_n R_{ab} = -\tfrac{1}{2}\chi[\Delta(\phi\bar{\phi}) - 2(\gamma+\bar{\gamma})\phi\bar{\phi}]n_a n_b. \tag{3.2}$$

Now by virtue of the Maxwell equation (2.2d) we have

$$\mathscr{L}_n R_{ab} = 0 \tag{3.3}$$

identically. Thus we infer

*Theorem 1*: The PR fields always admit a RC with respect to $n$.

### (ii) Curvature Collineation (CC)

The CC with respect to the vector field $n$ is defined by the condition

$$\mathscr{L}_n R_{dcb}{}^{a} = 0.$$

The curvature tensor $R_{dcb}{}^{a}$ by definition consists of two parts viz., the free-gravitational part

$$\underset{(F)}{R}{}_{dcb}{}^{a} = C_{dcb}{}^{a},$$

and the matter part

$$\underset{(M)}{R}{}_{dcb}{}^{a} = -\tfrac{1}{2}(g_{db}R_c{}^a - g_d{}^a R_{cb} + g_c{}^a R_{db} - g_{cb} R_d{}^a).$$

For the PR fields in question, we obtain

$$\mathscr{L}_n \underset{(F)}{R}{}_{dcb}{}^{a} = -2[\psi\{(\gamma+\bar{\gamma})\overline{m}_b n^a - (\pi+\alpha+\bar{\beta})n_b n^a\}\overline{m}_{[d}n_{c]} + \bar{\psi}\{(\gamma+\bar{\gamma})m_b n^a - (\bar{\pi}+\bar{\alpha}+\beta)n_b n^a\}m_{[d}n_{c]}] \tag{3.4a}$$

$$\begin{aligned}\mathscr{L}_n \underset{(M)}{R}{}_{dcb}{}^{a} = -\tfrac{1}{2}\chi\phi\bar{\phi}\,[&(\gamma+\bar{\gamma})m_{[d}n_{c]}\overline{m}_b n^a - (\bar{\pi}+\bar{\alpha}+\beta)\overline{m}_{[d}n_{c]}n_b n^a \\ &+ (\gamma+\bar{\gamma})\overline{m}_{[d}n_{c]}m_b n^a - (\pi+\alpha+\bar{\beta})m_{[d}n_{c]}n_b n^a]\end{aligned} \tag{3.4b}$$

by using (2.2a), (2.3a), and (2.3b), (2.6), and (2.8). Then we get from (3.4a) and (3.4b)

$$\begin{aligned}\mathscr{L}_n R_{dcb}{}^{a} &= \mathscr{L}_n \underset{(F)}{R}{}_{dcb}{}^{a} + \mathscr{L}_n \underset{(M)}{R}{}_{dcb}{}^{a}, \\ &= -2[(\gamma+\bar{\gamma})\{\psi\overline{m}_{[d}n_{c]} + \tfrac{1}{4}\chi\phi\bar{\phi}m_{[d}n_{c]}\}\overline{m}_b n^a - \{(\pi+\alpha+\bar{\beta})\psi + \tfrac{1}{4}\chi(\bar{\pi}+\bar{\alpha}+\beta)\phi\bar{\phi}\}\overline{m}_{[d}n_{c]}n_b n^a] \\ &\quad - [\mathrm{c.c.}].\end{aligned} \tag{3.5}$$

Here the symbol [c.c.] denotes the complex conjugate of the terms of the preceding bracket. Thus

$$\mathscr{L}_n R_{dcb}{}^{a} = 0$$

if and only if

$$\gamma + \bar{\gamma} = 0, \tag{3.6a}$$

$$(\pi+\alpha+\bar{\beta})\psi + \tfrac{1}{4}\chi(\bar{\pi}+\bar{\alpha}+\beta)\phi\bar{\phi} = 0. \tag{3.6b}$$

### Gradient field $n$

The evaluation and analysis of SCC and AC is very cumbersome, as is evident from the expression given in Appendices I, and II, even in the inevitable case $\pi = 0$. We henceforth impose the condition that $n$ is a gradient field, i.e.,

$$\nu = \mu + \bar{\mu} = 0, \tag{3.7a}$$

$$\gamma + \bar{\gamma} = \pi - (\alpha + \bar{\beta}) = 0. \tag{3.7b}$$

We note that (3.7a) is already taken care of by (2.2a), and (2.3a).

Now for the case of CC Eq. (3.6b) yields by $\pi = \alpha + \bar{\beta}$,

$$\pi\psi + \tfrac{1}{4}\chi\bar{\pi}\phi\bar{\phi} = 0. \tag{3.8}$$

This equation admits two types of solutions, herein termed as free curvature collineation (i.e., $\pi = 0$) and matter curvature collineation (i.e., $\pi \neq 0$).

*(a) Free Curvature collineation (Free CC)*

By virtue of (3.7b), we have from 3.4a) and (3.4b)

$$\mathscr{L}_n \underset{(F)}{R}{}_{dcb}{}^a = 0 \text{ and } \mathscr{L}_n \underset{(M)}{R}{}_{dcb}{}^a = 0$$

when

$$\pi = 0. \tag{3.9}$$

Thus we get

$$\mathscr{L}_n R_{dcb}{}^a = 0.$$

This case is referred by us as free CC since this symmetry is induced by the Weyl conformal tensor, ie., the symmetry of the curvature field due to the nonlocal matter.

*(b) Matter Curvature Collineation (Matter CC)*

This symmetry is induced by the matter part of the curvature tensor, when $\pi \neq 0$. In fact in RC (3.3), $\pi$ is unrestricted. Now Eq. (3.8) implies

$$\psi = -\tfrac{1}{4}\chi\phi\bar{\phi}(\bar{\pi}/\pi), \tag{3.10a}$$

and so

$$\psi\bar{\psi} = \tfrac{1}{16}\chi^2(\phi\bar{\phi})^2. \tag{3.10b}$$

This satisfies Maxwell equation (2.2b) and Bianchi identities (2.3b). However, Eqs. (2.2c) and (2.3c) with (3.10a) give, by using NP equations (2.5g), and (2.5h).

$$\pi\bar{\phi}(\delta\phi - 2\phi\beta) + \tfrac{1}{4}\bar{\pi}\phi(\bar{\delta}\bar{\phi} - 2\bar{\phi}\bar{\beta}) = 0,$$

which is of the form $A + \frac{1}{4}\bar{A} = 0$, where $A$ is complex. Consequently we infer that

$$\delta\phi = 2\phi\beta \tag{3.10c}$$

for $\pi \neq 0$.

*Remarks*: If Im $\pi = 0$ (i.e., $\pi = \bar{\pi}$), $\pi \neq 0$, it follows from (3.8) that the Weyl scalar $\psi$ is real, i.e.,

$$\psi = -\tfrac{1}{4}\chi\phi\bar{\phi}. \tag{3.11}$$

Since $\pi$ is real, the NP equations (2.5g), and (2.5h) yield $\alpha - \bar{\beta} = 0$ and hence, (2.2c) and (2.3c) give $\pi\phi\bar{\phi} = 0$ which implies $\phi = 0$. This is incompatible with the existence of the source-free null electromagnetic field. Thus equations (3.8), (3.9), and (3.10) yield the following:

*Theorem 2.* The PR fields having $n$ as a gradient field, admit (a) a free CC iff $\pi = 0$ and (b) a matter CC iff $\delta\phi = 2\phi\beta$.

### (iii) Weyl conformal collineation (WCC) and its degeneracy

The Weyl conformal collineation with respect to $n$ is defined

$$\mathscr{L}_n C_{dcb}{}^a = 0. \tag{3.12}$$

As a sequel to (2.7) and (3.4a), we get for the PR fields

$$\mathscr{L}_n C_{dcb}{}^a = 0$$

if and only if

$$\gamma + \bar{\gamma} = 0, \quad \pi + \alpha + \bar{\beta} = 0. \tag{3.13}$$

For a gradient $n$, we have

$$\pi = \alpha + \bar{\beta} = 0. \tag{3.14}$$

Thus, WCC is a trivial symmetry, since it degenerates to CC (3.9).

### (iv) Special Curvature Collineation

A space-time is said to admit a SCC generated by $\xi^a$ if and only if

$$(\mathscr{L}_\xi \Gamma^a_{bc})_{;d} = 0, \tag{3.15}$$

where $\Gamma^a_{bc}$ is the Christoffel symbol of the second kind and

$$\mathscr{L}_\xi \Gamma^a_{bc} = \xi^a{}_{;cb} + R_{dcb}{}^a \xi^d.$$

With the choice $\xi^a = n^a$ for the PR fields these equations (3.15) reduce to

$$n^a{}_{;cbd} = 0 \tag{3.16}$$

since $R_{dcb}{}^a n^d = 0$. On covariantly differentiating $n^a{}_{;cb}$ given in the Appendix and using the NP expressions [*vide* Ref. 22b)] for the covariant derivatives of the tetrad vectors, we infer after a tedious but straightforward computation that (3.16) are equivalent to 36 complex equations. If the symmetry vector $n$ is a gradient field (3.7) these 36 equations for SCC reduce to

$$\pi = \alpha + \bar{\beta} = 0, \tag{3.17a}$$

$$\Delta(\epsilon + \bar{\epsilon}) = \delta(\epsilon + \bar{\epsilon}) = \delta D(\epsilon + \bar{\epsilon}) = 0, \tag{3.17b}$$

$$DF - 3F(\epsilon + \bar{\epsilon}) = 0, \tag{3.17c}$$

where

$$F = -D(\epsilon + \bar{\epsilon}) + 2(\epsilon + \bar{\epsilon})^2.$$

Thus, PR fields admitting SCC are nontrivial. It should be noted that we have used, in getting (3.17), the condition $\delta(\epsilon + \bar{\epsilon}) = 0$, obtainable from NP equations (2.5d) and (2.5e).

### (v) Affine collineation

A space-time is said to admit an AC if there exists a vector field $\xi^a$, such that

$$\mathscr{L}_\xi \Gamma^a_{bc} = \xi^a{}_{;cb} + R_{dcb}{}^a \xi^d = 0. \tag{3.18}$$

For the PR fields, we choose $\xi^a = n^a$ and so (3.18) becomes

$$\mathscr{L}_n \Gamma^a_{bc} = n^a{}_{;cb} = 0, \tag{3.19}$$

by virtue of (2.11). Now the translation of (19) in the NP spin coefficients gives eight complex equations including $\pi = 0$, which is the coefficient of the term $l^a n_b n_c$ (*vide* Appendix I). By (3.7) these eight equations reduce to

$$\pi = \alpha + \bar{\beta} = 0, \tag{3.20a}$$

$$\Delta(\epsilon + \bar{\epsilon}) = \delta(\epsilon + \bar{\epsilon}) = 0, \tag{3.20b}$$

$$D(\epsilon + \bar{\epsilon}) - 2(\epsilon + \bar{\epsilon})^2 = 0. \tag{3.20c}$$

Thus AC is nontrivial for the PR fields and exists when (3.20) is valid.

### (vi) Degeneracy of Projective Collineation to AC

A projective collineation with respect to $n$ is defined by

$$\mathscr{L}_n \Gamma^a_{bc} = \delta_b{}^a A_{;c} + \delta_c{}^a A_{;b}, \tag{3.21}$$

where $A$ is an arbitrary function and $A_{;c}$ can be written as

$$A_{;c} \equiv A_{;i} g_c{}^i = DAn_c + \Delta A l_c - \delta A \bar{m}_c - \bar{\delta} A m_c. \quad (3.22)$$

Since $R_{dcb}{}^a n^d = 0$, for the PR fields, we have

$$n^a{}_{;cb} = \delta_b{}^a A_{;c} + \delta_c{}^a A_{;b}. \quad (3.23)$$

Now these tensor equations are equivalent to

$$\pi = \alpha + \bar{\beta} = 0, \quad (3.24a)$$

$$DA = \Delta A = \delta A = \bar{\delta} A = 0, \quad (3.24b)$$

$$\Delta (\epsilon + \bar{\epsilon}) = \bar{\delta}(\epsilon + \bar{\epsilon}) = 0, \quad (3.24c)$$

$$D(\epsilon + \bar{\epsilon}) - 2(\epsilon + \bar{\epsilon})^2 = 0 \quad (3.24d)$$

by (3.7) and (3.22). The condition (3.24b) implies that $A$ is constant, and it follows from (3.24a)–(3.24d) that the projective collineation for the PR fields degenerates to AC. Similarly one can show the degeneracy of the following five collineations to AC:

**(vii) Special Projective Collineation**

$$\mathscr{L}_n \Gamma^a_{bc} = \delta_b{}^a A_{;c} + \delta_c{}^a A_{;b}, \; A_{;bc} = 0.$$

**(viii) Conformal Collineation**

$$\mathscr{L}_n \Gamma^a_{bc} = \delta_b{}^a B_{;c} + \delta_c{}^a B_{;b} - g_{bc} g^{ad} B_{;d},$$

where $B$ is an arbitrary function.

**(ix) Special Conformal Collineation**

$$\mathscr{L}_n \Gamma^a_{bc} = \delta_b{}^a B_{;c} + \delta_c{}^a B_{;b} - g_{bc} g^{ad} B_{;d}, B_{;bc} = 0.$$

**(x) Null Geodesic Collineation**

$$\mathscr{L}_n \Gamma^a_{bc} = g_{bc} g^{ad} E_{;d},$$

where $E$ is an arbitrary function.

**(xi) Special Null Geodesic Collineation**

$$\mathscr{L}_n \Gamma^a_{bc} = g_{bc} g^{ad} E_{;d}, E_{;dc} = 0.$$

**(xii) Motion**

A motion with respect to $n$ is described by

$$\mathscr{L}_n g_{ab} = 0, \text{ i.e., } n_{a;b} + n_{b;a} = 0. \quad (3.25)$$

For the PR fields with (2.2a) and (2.3a), the translation of (3.25) in terms of the spin coefficients is, by using (2.4),

$$\pi + \alpha + \bar{\beta} = \epsilon + \bar{\epsilon} = \gamma + \bar{\gamma} = 0. \quad (3.26)$$

Thus with (3.7) we have

$$\mathscr{L}_n g_{ab} = 0 \quad \text{iff } \pi = \alpha + \bar{\beta} = \epsilon + \bar{\epsilon} = 0. \quad (3.27)$$

Now it is interesting to note that

**(xiii) Conformal motion**

$$\mathscr{L}_n g_{ab} = h g_{ab}, \quad h \text{ is a scalar.}$$

**(xiv) Special conformal motion**

$$\mathscr{L}_n g_{ab} = h g_{ab}, \quad h_{;ab} = 0.$$

**(xv) Homothetic motion**

$$\mathscr{L}_n g_{ab} = h g_{ab}, \quad h \text{ is constant}$$

all degenerate to motion.

**(xvi) Weyl projective collineation (WPC) and its degeneracy**

For the PR fields, we get from (2.10)

$$\mathscr{L}_n W_{dcb}{}^a = 2[(\gamma + \bar{\gamma})\psi \bar{m}_{[d} n_{c]} \bar{m}_b n^a - \tfrac{1}{4}\chi\phi\bar{\phi}(\gamma + \bar{\gamma}) m_{[d} n_{c]} \bar{m}_b n^a + \{\psi(\pi + \alpha + \bar{\beta}) + \tfrac{1}{4}\chi(\bar{\pi} + \bar{\alpha} + \beta)\phi\bar{\phi}\} m_{[d} n_{c]} n_b n^a] + [\text{c.c.}] - \tfrac{1}{3}\chi\phi\bar{\phi}[(\pi + \alpha + \bar{\beta}) m_{(d} n_{b)} + (\bar{\pi} + \bar{\alpha} + \beta)\bar{m}_{(d} n_{b)} - (\epsilon + \bar{\epsilon}) n_d n_b - (\gamma + \bar{\gamma}) m_{(d} \bar{m}_{b)}] n_c n^a, \quad (3.28)$$

by virtue of (2.2a), (2.2d), (2.3a), and (2.3b). Consequently the Weyl projective collineation with respect to $n$ described by

$$\mathscr{L}_n W_{dcb}{}^a = 0 \quad (3.29)$$

is equivalent to

$$\pi + \alpha + \bar{\beta} = \epsilon + \bar{\epsilon} = \gamma + \bar{\gamma} = 0.$$

Thus WPC is trivial since it degenerates to motion (*vide* 3.26 and 3.27).

Now we conclude that for the PR fields the sixteen geometrical symmetries are not all independent when $n$ is a gradient field. They reduce to five nontrivial symmetries *viz.*, RC, CC, SCC, AC, and $M$. Here we summarize them in a tabular form

| Symmetry | | $\pi$ | $\alpha + \bar{\beta}$ | $\epsilon + \bar{\epsilon}$ |
|---|---|---|---|---|
| (i) | RC | – | —— | – |
| (ii) | Matter CC | – | (3.10c) | – |
| | Free CC | 0 | 0 | – |
| (iii) | SCC | 0 | 0 | (3.17b), (3.17c) |
| (iv) | AC | 0 | 0 | (3.20b), (3.20c) |
| (v) | M | 0 | 0 | 0 |

where ' – ' denotes unrestricted. The table shows how the symmetries become stronger and stronger. In order to establish their existence, we determine a corresponding class of space-times explicitly under certain conditions in the following section.

## 4. METRICS CORRESPONDING TO THE NONTRIVIAL SYMMETRIES

Since the field equations (2.2b), (2.2c), (2.3b), (2.3c), and (2.5a)–(2.5o) are too cumbersome for analytical work, we assume that $l$ is a shear-free and that $n$ satisfies freedom conditions. In terms of NP scalars.

(a) the real null vector $l$ is shear-free:

$$\sigma = 0, \quad (4.1a)$$

(b) the complex null tetrad $Z_\alpha{}^a = \{l^a, m^a, \bar{m}^a, n^a\}$ is parallelly propagated[22] along $n$:

$$\nu = \gamma = \tau = 0. \quad (4.1b)$$

(c) $n$ is a gradient field[23]:

$$\pi = \alpha + \bar{\beta}, \mu = \bar{\mu}. \quad (4.1c)$$

However, (4.1c) are already taken care of in equation (3.7).

### The null tetrad

Since we consider the real null vector $n$ to be an infinitesimal generator in the study of symmetries, we choose the complex null tetrad $Z_\alpha{}^a$ as follows:

$$Z_\alpha{}^a = \begin{pmatrix} 1 & U & X^2 & X^3 \\ 0 & \omega & \xi^2 & \xi^3 \\ 0 & \bar{\omega} & \bar{\xi}^2 & \bar{\xi}^3 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \tag{4.2}$$

where $\omega$, $\xi^i$, $U$, and $X^i$ $(i = 2,3)$ are six arbitrary functions of coordinates. The intrinsic derivative operators $D$, $\delta$, $\bar{\delta}$, $\Delta$ take the following forms:

$$\begin{aligned} D &= \frac{\partial}{\partial u} + \frac{U\partial}{\partial r} + \frac{X^j\partial}{\partial x^j}, \\ \delta &= \frac{\omega\partial}{\partial r} + \frac{\xi^j\partial}{\partial x^j}, \\ \bar{\delta} &= \frac{\bar{\omega}\partial}{\partial r} + \frac{\bar{\xi}^j\partial}{\partial x^j}, \\ \Delta &= \frac{\partial}{\partial r}, \end{aligned} \tag{4.3}$$

where $j = 2,3$. (*Note*: The operators $D$,$\Delta$ correspond respectively to $\Delta$, $D$ of the NP formalism[21]). Then the completeness relation is

$$g^{ab} = l^a n^b + n^a l^b - m^a \bar{m}^b - \bar{m}^a m^b. \tag{4.4}$$

### Metric Equations

Under the conditions (4.1) we obtain the so-called metric equations by using (4.2) in the commutation relation ($D$-4) of Ref. 25 as follows:

$$\Delta U = -\pi\omega - \bar{\pi}\bar{\omega} + (\epsilon + \bar{\epsilon}), \tag{4.5a}$$

$$\Delta X^i = -\pi\xi^i - \bar{\pi}\bar{\xi}^i, \tag{4.5b}$$

$$\Delta\omega = \bar{\pi}, \tag{4.5c}$$

$$\Delta\xi^i = 0, \tag{4.5d}$$

$$\delta U - D\omega = \kappa - (\bar{\rho} + \epsilon - \bar{\epsilon})\omega, \tag{4.5e}$$

$$\delta X^i - D\xi^i = -(\bar{\rho} + \epsilon - \bar{\epsilon})\xi^i, \tag{4.5f}$$

$$\delta\bar{\omega} - \bar{\delta}\omega = (\bar{\beta} - \alpha)\omega + (\bar{\alpha} - \beta)\bar{\omega} + (\rho - \bar{\rho}), \tag{4.5g}$$

$$\delta\bar{\xi}^i - \bar{\delta}\xi^i = (\bar{\beta} - \alpha)\xi^i + (\bar{\alpha} - \beta)\bar{\xi}^i. \tag{4.5h}$$

### (i) A class of metrics admitting RC

For solving Bianchi identities (2.3b), (2.3c), the Maxwell's Eqs. (2.2b), (2.2c), the NP Eqs. (2.5a)–(2.5o), and the metric Eqs. (4.5a)–(4.5h) under the conditions (2.2a), (2.3a) and (4.1a)–(4.1c) we follow the method described by Newman and Tamburino,[26] and Collinson and Morris.[27] The solution of these equations is

$$\begin{aligned} &\nu = \gamma = \tau = \sigma = \lambda = \mu = 0, \\ &\beta = \beta_0, \\ &\alpha = \alpha_0 = \bar{\beta}_0 + 2\bar{P}(\log P)_{,z}, \\ &\pi = \pi_0 = 2\bar{\beta}_0 + 2\bar{P}(\log P)_{,z} \\ &\rho = \rho_0 = \frac{1}{f(u)} \exp\left[\int\left(\frac{\pi_0}{2P}\right)d\bar{z}\right], \end{aligned} \tag{4.6}$$

$$\begin{aligned} &\kappa = \kappa_0 - r\bar{\pi} \exp\left[\int\left(\frac{\pi_0}{2P}\right)d\bar{z}\right]/f(u), \\ &\epsilon = \epsilon_0 - r(\bar{\pi}_0\alpha_0 + \pi_0\beta_0), \\ &U = U_0 + r(\epsilon_0 + \bar{\epsilon}_0) - 4r^2\pi_0\bar{\pi}_0. \\ &X^2 = -r(\pi_0 P + \bar{\pi}_0\bar{P}),\ X^3 = -ir(\pi_0 P - \bar{\pi}_0\bar{P}), \\ &\omega = r\bar{\pi}_0,\ \xi^2 = -i\xi^3 = P, \\ &X_0^2 = X_0^3 = \omega_0 = 0, \end{aligned} \tag{4.7}$$

$$\phi = \phi_0 = PA(u), \tag{4.8a}$$

$$\psi = \psi_0 = 2(\beta_0 - P\partial/\partial\bar{z})(2PU_{0,\bar{z}} - U_0\bar{\pi}_0), \tag{4.8b}$$

where $A, f$ are functions of $u$ only, a subscript 0 denotes independence with respect to $r$ and

$$\begin{aligned} &P = P(u,x^2,x^3),\ z = x^2 + ix^3, \\ &\kappa_0 = 2PU_{0,\bar{z}} - U_0\bar{\pi}_0. \end{aligned} \tag{4.9}$$

Hence the components of the metric (4.4) determining the RC which is not a CC are

$$\begin{aligned} &g^{10} = g^{01} = 1,\ g^{23} = g^{00} = g^{02} = g^{03} = 0, \\ &g^{11} = 2[U_0 + r(\epsilon_0 + \bar{\epsilon}_0) - 4r^2\pi_0\bar{\pi}_0], \\ &g^{12} = -2r(\pi_0 + \bar{\pi}_0)\phi_0/A, \\ &g^{13} = -2ir(\pi_0 - \bar{\pi}_0)\phi_0/A, \\ &g^{22} = g^{33} = -2(\phi_0/A)^2,\ (A \neq 0). \end{aligned} \tag{4.10}$$

Here $\phi_0$ characterizes the pure electromagnetic-radiation field and $U_0$ is related to the pure gravitational-radiation field by (4.8a) and (4.8b).

### (ii) Space-time admitting nontrivial Matter CC

The metric determining matter CC is the same as in RC except for the relations [*vide* (3.10c) and (2.3c) with (2.2c) and (3.10c), respectively]

$$\phi = Q \exp\int\left(\frac{1}{P}\beta_0\right)d\bar{z}, \tag{4.11a}$$

$$\psi = \bar{\phi}P^3H, \tag{4.11b}$$

where $Q$, $H$ are functions of $u$ alone.

### (iii) Metrics admitting Free-CC, SCC, AC

Using the tetrad rotation[26] $m^{a'} = m^a \exp(i\theta)$, where $\theta$ is real and independent of $r$, we set

$$P = \bar{P} \tag{4.12}$$

such that we infer from (4.5f)

$$\epsilon = \bar{\epsilon}. \tag{4.13}$$

Then as a sequel to (4.5d), (4.5e), and (4.13), we obtain

$$\epsilon = \epsilon(u), \tag{4.14}$$

where $\epsilon$ is an arbitrary function of $u$ only. Therefore, the components of the metric tensor determining free CC which is not a SCC, have the form

$$\begin{aligned} &g^{10} = g^{01} = 1,\ g^{11} = 2(U_0 + 2r\epsilon),\ g^{00} = g^{02} = g^{03} = 0, \\ &g^{12} = g^{13} = g^{23} = 0,\ g^{22} = g^{33} = -2(\phi_0/A)^2, \end{aligned} \tag{4.15}$$

where $U_0$ is related to the pure gravitational-radiation field as follows:

$$\psi_0 = -4(P^2U_{0,\bar{z}})_{,\bar{z}} \tag{4.16}$$

and the electromagnetic field is given by (4.8a). Then the nontrivial nature of the metrics corrresponding to the three collineations (3.9), (3.17), and (3.20) can be distinguished as follows:

$$\text{Free CC: } \epsilon \text{ is an arbitrary function of } u. \quad (4.17)$$

$$\text{SCC: } \epsilon'' - 14\epsilon\epsilon' + 24\epsilon^3 = 0, \ \epsilon' \neq 4\epsilon^2. \quad (4.18)$$

$$\text{AC: } \epsilon' - 4\epsilon^2 = 0, \quad \text{or} \quad \epsilon = -B(4uB+1)^{-1}, \quad (4.19)$$

where $\epsilon' = d\epsilon/du$ and $B$ is a nonzero constant. It may be noted that, when $B = 0$, AC will degenerate to motion.

### (iv) Metric admitting motion

We observe from (3.27) that the salient feature of motion is

$$\epsilon = 0, \quad (4.20)$$

and the line element is

$$ds^2 = -2U_0(du)^2 - 2\,du\,dr - \tfrac{1}{2}(A/\phi_0)^2\{(dx^2)^2 + (dx^3)^2\}, \quad (4.21)$$

where $U_0, A, \phi_0, \psi_0$ are the same as in AC (*vide* 4.19).

## 5. DISCUSSION

The choice of PR fields in this paper is in consonance with the Tariq and Tupper's theorem,[7] "The only curvature collineations admitted by null source-free Einstein–Maxwell fields, not of Petrov-type $N$ or $O$, are conformal motions." However, they did not aim at getting exact solutions of Einstein–Maxwell field equations. We have obtained nontrivial metrics describing the PR fields propagating along the real null vector $n$.

If one considers the real null vector $l$ of the complex null tetrad $Z_\alpha{}^a$ as a symmetry vector instead of $n$ in the above investigation, the corresponding pure radiation fields are characterized by the two scalars $\psi_4 \neq 0$ and $\phi_2 \neq 0$, since $l$ will then be the common propagation vector of both the radiation fields. Accordingly instead of $\kappa, \epsilon, \pi, \rho, \alpha, \beta, \sigma$ we have to consider the nonzero spin coefficients, $\nu, \gamma, \tau, \mu, \alpha, \beta, \lambda$ in the case of $l$. However, it may be noted that the form of the nontrivial metrics given in Sec. 4 is essentially unaltered.

For an isolated system, the gravitational-radiation field (with the propagation vector $n$) is given[21] by

$$\psi \equiv \psi_0 = O(r^{-5}).$$

Since $\psi$ is independent of $r$ [*vide* Eq. (2.3b)], we infer that the metrics obtained in this paper do not represent self-gravitating systems.

## ACKNOWLEDGMENTS

The authors are thankful to the referee for valuable comments. Discussions with Mrs. S. P. Gumaste have been illuminating. One of the authors (N.I.S.) is grateful to the University Grants Commission, New Delhi for the award of a Research Fellowship.

## APPENDIX I

The Newman–Penrose concomitant of $n^a{}_{;cb}$ under the conditions $\nu = \lambda = \mu = 0$ are given below:

$$\begin{aligned}
n^a{}_{;cb} &= 2\pi\bar\pi l^a n_b n_c + [\Delta(\gamma+\bar\gamma)l_b l_c + \{2(\epsilon+\bar\epsilon)(\gamma+\bar\gamma) - \Delta(\epsilon+\bar\epsilon) - \pi\tau - \bar\pi\bar\tau - (\alpha+\bar\beta)\tau - (\bar\alpha+\beta)\bar\tau\}l_b n_c \\
&\quad + \{D(\gamma+\bar\gamma) + (\alpha+\bar\beta)\bar\pi + (\bar\alpha+\beta)\pi\}n_b l_c + \{\Delta(\alpha+\bar\beta) + 2\bar\gamma(\alpha+\bar\beta) - \bar\tau(\gamma+\bar\gamma)\}l_b m_c \\
&\quad + \{\Delta(\bar\alpha+\beta) + 2\gamma(\bar\alpha+\beta) - \tau(\gamma+\bar\gamma)\}l_b\bar m_c - \delta(\gamma+\bar\gamma)\bar m_b l_c - \bar\delta(\gamma+\bar\gamma)m_b l_c \\
&\quad + \{-\bar\delta(\alpha+\bar\beta) + 2\bar\beta(\alpha+\bar\beta) + (\gamma+\bar\gamma)\bar\sigma\}m_b m_c + \{-\delta(\bar\alpha+\beta) + 2\beta(\bar\alpha+\beta) + (\gamma+\bar\gamma)\sigma\}\bar m_b\bar m_c \\
&\quad + \{-\bar\delta(\bar\alpha+\beta) + 2(\bar\alpha+\beta)\alpha + \rho(\gamma+\bar\gamma)\}m_b\bar m_c \\
&\quad + \{-\delta(\alpha+\bar\beta) + 2(\alpha+\bar\beta)\bar\alpha + \bar\rho(\gamma+\bar\gamma)\}\bar m_b m_c + \{D(\bar\alpha+\beta) \\
&\quad - 2\epsilon(\bar\alpha+\beta) - \bar\pi(\epsilon+\bar\epsilon) - \kappa(\gamma+\bar\gamma)\}n_b\bar m_c + \{D(\alpha+\bar\beta) - 2\bar\epsilon(\alpha+\bar\beta) \\
&\quad - \pi(\epsilon+\bar\epsilon) - \bar\kappa(\gamma+\bar\gamma)\}n_b m_c + \{\bar\delta(\epsilon+\bar\epsilon) - 2(\epsilon+\bar\epsilon)(\alpha+\bar\beta) + \pi\rho + \bar\pi\bar\sigma + \rho(\alpha+\bar\beta) \\
&\quad + (\bar\alpha+\beta)\bar\sigma\}m_b n_c + \{\delta(\epsilon+\bar\epsilon) - 2(\epsilon+\bar\epsilon)(\bar\alpha+\beta) + \bar\pi\bar\rho + \pi\sigma + \bar\rho(\bar\alpha+\beta) + (\alpha+\bar\beta)\sigma\}\bar m_b n_c \\
&\quad + \{-D(\epsilon+\bar\epsilon) + 2(\epsilon+\bar\epsilon)^2 - (\bar\alpha+\beta)\bar\kappa - (\alpha+\bar\beta)\kappa \\
&\quad - \bar\pi\bar\kappa - \pi\kappa\}n_b n_c]\mathrm{n}^a + [\{D\pi - \pi(3\bar\epsilon+\epsilon)\}n_b n_c + \{\Delta\pi - 2\pi\bar\gamma\}l_b n_c + \pi(\gamma+\bar\gamma)n_b l_c \\
&\quad + \{-\delta\pi + \pi(\bar\alpha+\beta)\}\bar m_b n_c + \{-\bar\delta\pi + 2\pi\bar\beta\}m_b n_c \\
&\quad + \{\pi^2 + (\alpha+\bar\beta)\pi\}n_b m_c + \{\pi\bar\pi + 2\pi(\bar\alpha+\beta)\}n_b\bar m_c]m^a + [\text{c.c.}],
\end{aligned}$$

where the symbol [c.c.] denotes the complex conjugate of the terms of the preceeding bracket.

## APPENDIX II

### NP equivalent of $n^a{}_{;cbd} = 0$

Under the condition $\nu = \mu = \lambda = 0$, we get that

$$n^a{}_{;cbd} = 0$$

are equivalent to

$$\pi = 0,$$

$$DA + A(\epsilon+\bar\epsilon) = 0,$$

$$\Delta A + A(\gamma+\bar\gamma) = 0,$$

$$\delta A + A(\bar\alpha+\beta) = 0,$$

$$DC - C(\epsilon+\bar\epsilon) - L\bar\kappa - \bar L\kappa = 0,$$

$$\Delta C - C(\gamma+\bar\gamma) - L\bar\tau - \bar L\tau = 0,$$
$$\delta C - C(\bar\alpha+\beta) - L\bar\rho - \bar L\sigma = 0,$$
$$\bar\delta C - C(\alpha+\bar\beta) - L\bar\sigma - L\rho = 0,$$
$$DL + L(\bar\epsilon - \epsilon) - A\kappa = 0,$$
$$\Delta L + L(\bar\gamma - \gamma) - A\tau = 0,$$
$$\delta L - L(\beta - \bar\alpha) - A\sigma = 0,$$
$$\bar\delta L - L(\alpha - \bar\beta) - A\rho = 0,$$
$$DB - B(\epsilon+\bar\epsilon) - E\kappa - \overline{E\kappa} = 0,$$
$$\Delta B - B(\gamma+\bar\gamma) - E\tau - \overline{E\tau} = 0,$$
$$\delta B - B(\alpha+\bar\beta) - E\rho - \overline{E\sigma} = 0,$$
$$\bar\delta B - B(\bar\alpha+\beta) - E\sigma - \overline{E\rho} = 0,$$
$$DE = E(\epsilon-\bar\epsilon) - A\bar\kappa = 0,$$
$$\Delta E + E(\gamma-\bar\gamma) - A\bar\tau = 0,$$
$$\delta E - E(\bar\alpha-\beta) - A\bar\rho = 0,$$
$$\bar\delta E - E(\bar\beta-\alpha) - A\bar\sigma = 0,$$
$$DI - I(3\epsilon+\bar\epsilon) - C\kappa - H\kappa - \overline{G\kappa} = 0,$$
$$\Delta I - I(3\gamma+\bar\gamma) - C\tau - H\tau - \overline{G\tau} = 0,$$
$$\delta I - I(\bar\alpha+3\beta) - C\sigma - H\sigma - \overline{G\rho} = 0,$$
$$\bar\delta I - I(\bar\beta+3\alpha) - C\rho - H\rho - \overline{G\sigma} = 0,$$
$$DG + G(\epsilon-3\bar\epsilon) - (\bar L+E)\bar\kappa = 0,$$
$$\Delta G + G(\gamma-3\bar\gamma) - (\bar L+E)\bar\tau = 0,$$
$$\delta G + G(3\bar\alpha-\beta) - (\bar L+E)\bar\rho = 0,$$
$$\bar\delta G - G(3\bar\beta-\alpha) - (\bar L+E)\bar\sigma = 0,$$
$$DH - H(\epsilon+\bar\epsilon) - \bar L\kappa - \overline{E\kappa} = 0,$$
$$\Delta H - H(\gamma+\bar\gamma) - \bar L\tau - \overline{E\tau} = 0,$$
$$\delta H - H(\bar\alpha+\beta) - \bar L\sigma - \bar E\rho = 0,$$
$$\bar\delta H - H(\alpha+\bar\beta) - \bar L\rho - \overline{E\sigma} = 0,$$
$$DF - 3F(\epsilon+\bar\epsilon) - (\bar I+J)\kappa - (I+\bar J)\bar\kappa = 0,$$
$$\Delta F - 3F(\gamma+\bar\gamma) - (\bar I+J)\tau - (I+\bar J)\bar\tau = 0,$$
$$\delta F - 3F(\bar\alpha+\beta) - (\bar I+J)\sigma - (I+\bar J)\bar\rho = 0,$$
$$\bar\delta F - 3F(\alpha+\bar\beta) - (\bar I+J)\rho - (I+\bar J)\bar\sigma = 0,$$

where

$$A = \Delta(\gamma+\bar\gamma),\ C = D(\gamma+\bar\gamma),\ L = -\delta(\gamma+\bar\gamma)$$
$$B = -\Delta(\epsilon+\bar\epsilon) + 2(\epsilon+\bar\epsilon)(\gamma+\bar\gamma) - (\alpha+\bar\beta)\tau - (\bar\alpha+\beta)\bar\tau,$$
$$E = \Delta(\alpha+\bar\beta) + 2\bar\gamma(\alpha+\bar\beta) - \bar\tau(\gamma+\bar\gamma),$$
$$G = \bar\delta(\alpha+\bar\beta) + 2\bar\beta(\alpha+\bar\beta) + (\gamma+\bar\gamma)\bar\sigma,$$
$$H = -\delta(\bar\alpha+\beta) + 2\alpha(\bar\alpha+\beta) + \rho(\gamma+\bar\gamma),$$
$$I = D(\bar\alpha+\beta) - 2\epsilon(\bar\alpha+\beta) - (\gamma+\bar\gamma)\kappa,$$
$$J = \bar\delta(\epsilon+\bar\epsilon) - 2(\epsilon+\bar\epsilon)(\alpha+\bar\beta) + \rho(\alpha+\bar\beta) + (\bar\alpha+\beta)\bar\sigma,$$
$$F = -D(\epsilon+\bar\epsilon) + 2(\epsilon+\bar\epsilon)^2 - (\bar\alpha+\beta)\bar\kappa - (\alpha+\bar\beta)\kappa.$$


[1]S. J. Campbell and J. Wainwright, Gen. Relativ. Gravit. **8**, 987 (1977).
[2]G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969); **11**, 1578 (1970); Tensor N.S. **21**, 51 (1970).
[3]G. H. Katzin and J. Levine, Tensor. (N.S.) **22**, 64 (1971); Colloq. Math. **36**, 22 (1972).
[4]J. Levine and G. H. Katzin, Tensor (N.S.) **19**, 317 (1968); **20**, 251 (1969); **21**, 319 (1970).
[5]A. N. Petrov, *Einstein Spaces*, edited by J. Woodrow (Pergamon, Oxford, 1969).
[6]C. D. Collinson, J. Math. Phys. **11**, 818 (1970).
[7]N. Tariq and B. O. J. Tupper, Tensor (N.S.) **31**, 42 (1977).
[8]C. B. G. McIntosh, "Symmetries and exact solutions of Einstein's Equations," in *Gravitational Radiation, Collapsed Objects and Exact Solutions*: Proceedings of the Einstein Centenary Summer School, Perth, Australia, *Lecture Notes in Physics*, Vol. 128 edited by C. Edwards, (Springer-Verlag, Berlin, 1980), pp. 469–476.
[9]W. D. Halford, C. B. G. McIntosh, and E. H. van Lecuwen, J. Phys. A: Math. Gen. **13**, 2995 (1980).
[10]J. Leroy, "Motions and conformal motions in Petrov-type *N* vacuum metrics with twist," in *Abstracts of Contributed Papers*, the 9th International Conference on General Relativity and Gravitation: 14–19, July 1980, Jena, German Democratic Republic, Vol. 1, paper GR-9, pp. 49–50.
[11]B. Lukacs, Z. Parjes, and A. Sebestyan, J. Math. Phys. **22**, 1246 (1981).
[12]C. B. G. McIntosh, Gen. Relativ. Gravit. **7**, 199 (1976).
[13]G. G. Asgekar and T. H. Date, J. Math. Phys. **18**, 738 (1977).
[14]L. Radhakrishna and A. B. P. Rao, "Collineations in collapsing radiative perfect fluid," in *Proceedings of the Indian Association*, edited by M. Nagaraj (R. V. Prabhakara, Bangalore, 1982), pp. 149–173.
[15]G. S. Hall, "Curvature collineations and the determination of the metric from the curvature in General Relativity," Gen. Relativ. Gravit. **15**, 581 (1983).
[16]W. R. Davis, L. H. Green, and L. K. Norris, Nuovo Cimento B **34**, 256 (1976).
[17]P. Szekeres, Ph.D. thesis, University of London, 1964 (unpublished).
[18]S. R. Roy and L. Radhakrishna, Proc. R. Soc. London, Ser. A **275**, 245 (1963).
[19]W. Kundt, Z. Phys. **163**, 77 (1961).
[20]C. B. G. McIntosh and W. D. Halford, J. Math. Phys. **23**, 436 (1982).
[21]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[22]G. C. Debney and J. D. Zund, Tensor (N.S.) a) **22**, 333 (1971); b) **24**, 47 (1972).
[23]E. J. Flaherty, *Hermitian and Kahlerian Geometry in Relativity*: *Lecture Notes in Physics*, Vol. 46, (Springer-Verlag, Berlin, 1976).
[24]K. P. Singh, L. Radhakrishna, and R. Sharan, Ann. Phys. **1**, 46 (1965).
[25]M. Carmeli, *Group Theory and General Relativity*, (McGraw-Hill, New York, 1977), p. 333.
[26]E. T. Newman and L. A. Tamburino, J. Math. Phys. **3**, 902 (1962).
[27]C. D. Collinson and P. B. Morris, Int. J. Theor. Phys. **5**, 293 (1972).

# Kaluza–Klein theory derived from a Riemannian submersion

P. A. Hogan
*Mathematical Physics Department, University College, Belfield, Dublin 4, Ireland and School of Theoretical Physics, Dublin Institute for Advanced Studies, Dublin 4, Ireland*

(Received 30 August 1983; accepted for publication 2 December 1983)

Taking a Riemannian submersion as our starting point, we obtain some formulas derived from O'Neill's fundamental equations of a submersion and compare them with the basic equations of Bergmann's approach to Kaluza–Klein theory in five dimensions. Having imposed Hermann's sufficient conditions for the submersion to be a principal fiber bundle, we study the conclusions that can be drawn from the derived formulas.

PACS numbers: 04.50 + h, 02.40.Ky

## 1. INTRODUCTION

The use of principal fiber bundles possessing Riemannian metrics for studying the Kaluza–Klein[1–3] approach to the Einstein–Maxwell theory, and its generalization to the Einstein–Yang–Mills theory, was initiated by Trautman,[4,5] worked through by Kerner[6] and Cho,[7] and further studied and developed by Kopczyński,[8] Bradfield and Kantowski,[9] Cho and Freund,[10] and others. In all of these cases the bundle considered resembled a Riemannian submersion. We note that if $M$ and $B$ are $C^\infty$ Riemannian manifolds, then a Riemannian submersion is a $C^\infty$ map $\pi: M \to B$ having the properties that (i) $\pi$ is of maximal rank and (ii) $\pi_*$ preserves the lengths of horizontal vectors, i.e., vectors orthogonal to the fiber $\pi^{-1}$ (b) for $b \in B$. Here $\pi_*$ is the derivative map induced by $\pi$.

The purpose of the present paper is to study Kaluza–Klein theory taking a Riemannian submersion as starting point. In Sec. 2 we establish some consequences of the fundamental equations of a submersion developed by O'Neill.[11] This is followed in Sec. 3 by a comparison with Bergmann's[3] approach to Kaluza–Klein theory in the special case of five dimensions. In Sec. 4 we invoke the theorem of Hermann,[12] giving sufficient conditions for the submersion to be a fiber bundle, and study the consequences of the equations obtained in Sec. 2. The paper ends with a discussion in Sec. 5.

## 2. DEDUCTIONS FROM THE FUNDAMENTAL EQUATIONS

The fibers of a submersion $\pi: M \to B$, denoted $\pi^{-1}$ (b) for $b \in B$, are submanifolds of $M$ of dimension dim $M$ − dim $B$ as a consequence of property (i) of a submersion.[11] Vector fields on $M$ which are tangent to the fibers will be called "vertical" while vector fields orthogonal to the fibers are "horizontal." If $E$ is a vector field on $M$, it may be decomposed into its horizontal and vertical parts, which we write as

$$E = \mathscr{H}E + \mathscr{V}E. \tag{2.1}$$

O'Neill[11] defines the tensors $A$ and $T$ by

$$T_E F = \mathscr{H} D_{\mathscr{V}E} \mathscr{V}F + \mathscr{V} D_{\mathscr{V}E} \mathscr{H}F, \tag{2.2a}$$

$$A_E F = \mathscr{V} D_{\mathscr{H}E} \mathscr{H}F + \mathscr{H} D_{\mathscr{H}E} \mathscr{V}F, \tag{2.2b}$$

where $E$ and $F$ are vector fields on $M$ and $D$ is the Riemannian connection on $M$. Both $T$ and $A$ are tensors of type (1,2). If $V, W$ are vertical vector fields, then[11]

$$D_V W = \mathscr{V} D_V W + T_V W, \tag{2.3}$$

showing that $T$ is the second fundamental tensor (cf. Ref. 13, p. 75) of the fibers while if $X$ and $Y$ are horizontal vector fields[11]

$$A_X Y = \tfrac{1}{2}\mathscr{V}[X,Y], \tag{2.4}$$

indicating that $A$ is the integrability tensor of the horizontal distribution $\mathscr{H}$ on $M$. Many useful properties of the tensors $T$ and $A$ are derived in Ref. 11.

Denoting by $\langle\ ,\ \rangle$ the Riemannian metric on $M$, the Riemann–Christoffel curvature tensor $R$ is given by

$$R(E,F,P,L) = \langle E, R_{PL}(F)\rangle, \tag{2.5}$$

with

$$R_{PL}(F) = D_{[P,L]}F - D_P D_L F + D_L D_P F, \tag{2.6}$$

where $E$, $F$, $P$, and $L$ are vector fields on $M$. O'Neill's[11] fundamental equations of a submersion consist of the components of the curvature tensor $R$ expressed in terms of the tensors $T$ and $A$ and their covariant derivatives. If $X$, $Y$, $Z$, and $H$ are horizontal vector fields and $U$, $V$, $W$, and $F$ are vertical vector fields, then he finds that

$$R(F,W,U,V) = \langle F, \hat{R}_{UV}(W)\rangle - \langle T_U W, T_V F\rangle + \langle T_V W, T_U F\rangle, \tag{2.7a}$$

$$R(X,W,U,V) = \langle (D_V T)_U W, X\rangle - \langle (D_U T)_V W, X\rangle, \tag{2.7b}$$

$$R(H,Z,X,Y) = \langle H, R^*_{XY}(Z)\rangle - 2\langle A_X Y, A_Z H\rangle + \langle A_Y Z, A_X H\rangle + \langle A_Z X, A_Y H\rangle, \tag{2.7c}$$

$$R(V,Z,X,Y) = \langle (D_Z A)_X Y, V\rangle + \langle A_X Y, T_V Z\rangle - \langle A_Y Z, T_V X\rangle - \langle A_Z X, T_V Y\rangle, \tag{2.7d}$$

$$R(W,Y,X,V) = \langle (D_X T)_V W, Y\rangle + \langle (D_V A)_X Y, W\rangle - \langle T_V X, T_W Y\rangle + \langle A_X V, A_Y W\rangle. \tag{2.7e}$$

In (2.7a) the first term on the right-hand side is the curvature tensor of the fiber metric while the first term on the right-hand side of (2.7c) is the horizontal lift of the curvature tensor of $B$. The covariant derivative of $T$ and $A$ appearing in (2.7b), (2.7d), and (2.7e) is given, for example, by

$$(D_V T)_U W = D_V(T_U W) - T_{D_V U}(W) - T_U(D_V W). \tag{2.8}$$

We shall find it convenient to specify basis vector fields on $M$ as follows: Suppose $\dim B = n$, $\dim \pi^{-1}(b) = m$ for $b \in B$; then $\dim M = m + n$. Let Greek indices take values $1,2,3,\ldots,n$ and Latin indices take values $1,2,3,\ldots m$. Let $\{e^i\}$ be a set of $m$ linearly independent vertical vector fields and let $\{e_\mu\}$ be a set of $n$ linearly independent horizontal vector fields which are $\pi$-related to a set of $n$ linearly independent vector fields on $B$. Such horizontal vector fields on $M$ are called *basic* vector fields by O'Neill. If $\{\theta^i,\theta^\mu\}$ is a dual basis of 1-form fields on $M$, then the metric tensor $g = \langle\, , \rangle$ on $M$ may be written

$$g = g_{ij}\theta^i \otimes \theta^j + g_{\mu\nu}\theta^\mu \otimes \theta^\nu, \tag{2.9}$$

with $g_{ij} = \langle e_i, e_j\rangle$, $g_{\mu\nu} = \langle e_\mu, e_\nu\rangle$.

The vector fields $A_{e_\mu} e_\nu$ are vertical and so we may write

$$A_{e_\mu} e_\nu = -\tfrac{1}{2}F^i{}_{\mu\nu} e_i, \tag{2.10}$$

with $F^i{}_{\mu\nu} = -F^i{}_{\nu\mu}$ following from (2.4). The vector fields $A_{e_\mu} e_i$ are horizontal and thus we have

$$A_{e_\mu} e_i = \tfrac{1}{2}W_{i\mu}{}^\sigma e_\sigma. \tag{2.11}$$

We can show that $F^i{}_{\mu\nu} = W^i{}_{\mu\nu}$, where indices are raised and lowered with the use of the metric (2.9). This follows using (2.2b) and the fact that $D$ is Riemannian, i.e., torsion-free and compatible with the metric (2.9), since

$$F_{j\mu\nu} = -2\langle e_j, D_{e_\mu} e_\nu\rangle, \tag{2.12}$$

and also

$$W_{j\mu\nu} = 2\langle e_\nu, D_{e_\mu} e_j\rangle = -2\langle D_{e_\mu} e_\nu, e_j\rangle = F_{j\mu\nu}. \tag{2.13}$$

For future reference we note that since $\{e_\mu\}$ are basic, if $V$ is vertical then

$$V\langle e_\mu, e_\nu\rangle = 0 \tag{2.14a}$$

and, since $\pi_* V = 0$, $[V, e_\mu]$ is vertical and thus, in particular,

$$\mathscr{H}[e_i, e_\mu] = 0. \tag{2.14b}$$

We can now prove the following:

*Lemma 1:* If the submersion $\pi: M \to B$ has totally geodesic fibers, then the Ricci scalar of $M$ may be written

$$R = R^* + \hat{R} - \tfrac{1}{4}\|F\|^2, \tag{2.15}$$

where $R^*$ is the Ricci scalar of $B$ lifted to $M$ via $\pi$, $\hat{R}$ is the Ricci scalar of the fibers, and

$$\|F\|^2 = F_{i\mu\nu}F^{i\mu\nu}. \tag{2.16}$$

*Proof*: Since the fibers are totally geodesic (cf. Ref. 14, p. 180), the tensor $T$ vanishes and thus (2.7) gives the following components of the curvature tensor on the basis $\{e_i, e_\mu\}$:

$$R_{ijkl} = \hat{R}_{ijkl}, \tag{2.17a}$$

$$R_{\mu ijk} = 0, \tag{2.17b}$$

$$R_{\mu\nu\rho\sigma} = R^*_{\mu\nu\rho\sigma} - 2\langle A_{e_\rho} e_\sigma, A_{e_\nu} e_\mu\rangle + \langle A_{e_\sigma} e_\nu, A_{e_\rho} e_\mu\rangle + \langle A_{e_\nu} e_\rho, A_{e_\sigma} e_\mu\rangle, \tag{2.17c}$$

$$R_{i\mu\nu\rho} = \langle (D_{e_\mu} A)_{e_\nu} e_\rho, e_i\rangle, \tag{2.17d}$$

$$R_{i\nu\rho j} = \langle (D_{e_j} A)_{e_\rho} e_\nu, e_i\rangle + \langle A_{e_\rho} e_j, A_{e_\nu} e_i\rangle, \tag{2.17e}$$

where $R_{ABCD} = R(e_A, e_B, e_C, e_D)$, with capital letters taking values $1,2,3,\ldots,n+m$. From (2.10), (2.17a), and (2.17c)

$$R = R^* - \tfrac{3}{4}\|F\|^2 + 2g^{\nu\rho}g^{ij}R_{i\nu\rho j} + \hat{R}. \tag{2.18}$$

In (2.17e) the first term on the right-hand side is skew-symmetric in $\rho,\nu$ (this is a consequence of Lemma 6 of Ref. 11) and so

$$g^{ij}g^{\nu\rho}R_{i\nu\rho j} = g^{ij}g^{\nu\rho}\langle A_{e_\rho} e_j, A_{e_\nu} e_i\rangle = \tfrac{1}{4}\|F\|^2, \tag{2.19}$$

using (2.11) and (2.13). Thus, combining (2.18) and (2.19), we find that the Ricci scalar of $M$ is given by (2.15) which has the general *form* of the Lagrangian density appearing in the Kaluza–Klein approach to the Einstein–Yang–Mills theory.[7–9]

We end this section by giving two further lemmas. The first is O'Neill's[11] Lemma 7 which we state without proof:

*Lemma 2:* If $V$ is vertical and $\mathfrak{S}$ denotes the cyclic sum over the horizontal vector fields $X$, $Y$, and $Z$, then

$$\mathfrak{S}\langle (D_Z A)_X Y, V\rangle = \mathfrak{S}\langle A_X Y, T_V Z\rangle. \tag{2.20}$$

We shall see in what follows that this equation coincides with the Bianchi identities satisfied by the Yang–Mills field when we make the necessary specializations.

*Lemma 3:* If $X, Y$ are basic and $V, W$ are vertical vector fields, then

$$\langle D_V(A_X Y), W\rangle + \langle D_W(A_X Y), V\rangle = \langle (D_Y T)_W V, X\rangle - \langle (D_X T)_W V, Y\rangle. \tag{2.21}$$

*Proof*: The proof follows from the observation of O'Neill[11] that identities involving the derivatives of $T$ and $A$ can be obtained from (2.7) using the symmetries of the curvature tensor. We begin, however, with the identity

$$\langle (D_V A)_X Y, W\rangle = \langle D_V(A_X Y), W\rangle - \langle A_{D_V X}(Y), W\rangle - \langle A_X(D_V Y), W\rangle. \tag{2.22}$$

Using the properties of $A$, the fact that $D$ is Riemannian, and also that $\mathscr{H}[X,W] = 0 = \mathscr{H}[Y,W]$ since $X, Y$ are basic, we have

$$\begin{aligned}\langle A_{D_V X}(Y), W\rangle &= -\langle A_Y(D_V X), W\rangle \\ &= -\langle A_Y(D_X V), W\rangle \\ &= -\langle A_Y A_X V, W\rangle \\ &= \langle A_X V, A_Y W\rangle.\end{aligned} \tag{2.23}$$

Thus (2.22) may be written

$$\langle (D_V A)_X Y, W\rangle = \langle D_V(A_X Y), W\rangle - \langle A_X V, A_Y W\rangle + \langle A_Y V, A_X W\rangle. \tag{2.24}$$

From the symmetry of the curvature tensor

$$R(V,X,Y,W) = R(W,Y,X,V), \tag{2.25}$$

together with (2.24) and

$$\langle (D_Y T)_W V, X\rangle = \langle (D_Y T)_V W, X\rangle \tag{2.26}$$

(cf. Ref. 11, Lemma 6), we arrive at (2.21) above.

If we denote covariant differentiation in the fibers by a caret, i.e., if $V, W$ are vertical vector fields,

$$\hat{D}_V W = \mathscr{V} D_V W, \tag{2.27}$$

we see that the left-hand side of (2.21) may be written

$$\langle \hat{D}_V(A_X Y), W\rangle + \langle \hat{D}_W(A_X Y), V\rangle.$$

Hence, if the fibers are totally geodesic the right-hand side of (2.21) vanishes and we obtain Killing's equations (cf. Ref. 15, p. 88) satisfied by $A_X Y$, i.e., *$A_X Y$ is a Killing vector field of* *the fiber metric*. This result is due to Bishop,[16] who gave an elegant geometrical proof.

## 3. BERGMANN'S FIVE-DIMENSIONAL THEORY

In his approach to the five-dimensional Kaluza–Klein theory, Bergmann[3] first developed some useful formulas for the study of a unit vector field in a five-dimensional Riemannian space. Making use of the notation of Sec. 2, we take $m = 1$, $n = 4$, choose a local coordinate system $\{x^A\}$ with $A = 1,2,3,4,5$, and write

$$e_\mu = e_\mu^A \frac{\partial}{\partial x^A}, \quad e_1 = A^B \frac{\partial}{\partial x^B}, \tag{3.1}$$

and assume that $\langle e_1, e_1\rangle = 1$. We raise and lower capital indices using the metric tensor components

$$g_{AB} = \left\langle \frac{\partial}{\partial x^A}, \frac{\partial}{\partial x^B}\right\rangle, \tag{3.2}$$

and define

$$A_{BC} = A_{B|C} - A_{C|B}, \quad B_A = A_{A|B}A^B, \tag{3.3}$$

with the stroke indicating covariant differentiation with respect to the metric (3.2). Bergmann[3] obtains the following formulas [his Eqs. (17.15), (17.24), (17.29), and (17.51), respectively]:

$$e_B^\mu e_C^\nu g_{\mu\nu,A}A^A = A_{B|C} + A_{C|B} - A_B B_C - A_C B_B, \tag{3.4a}$$

$$\varphi_{\mu\nu,B}A^B = e_\mu^A e_\nu^B (B_{A|B} - B_{B|A}), \tag{3.4b}$$

$$e_\rho(\varphi_{\mu\nu}) + e_\nu(\varphi_{\rho\mu}) + e_\mu(\varphi_{\nu\rho}) = e_\mu^A e_\nu^B e_\rho^C (B_A A_{CB} + B_B A_{AC} + B_C A_{BA}), \tag{3.4c}$$

$$R = R^* - A^{D|C}A_{D|C} - (A^D{}_{|D})^2 + B_D B^D - 2(A^D{}_{|D})_{|C}A^C + 2B^D{}_{|D}. \tag{3.4d}$$

Here $g_{\mu\nu} = \langle e_\mu, e_\nu\rangle$ with Greek indices being raised and lowered with this metric. Partial differentiation is indicated by a comma. Also

$$\varphi_{\mu\nu} = e_\mu^A e_\nu^B A_{AB}, \tag{3.5}$$

and these quantities are related to Maxwell's electromagnetic tensor in this theory. In deriving (3.4c) we have to assume $\mathscr{H}[e_\mu, e_\nu] = 0$. This is equivalent to Bergmann's Eq. (17.28) and can always be guaranteed to hold at a point, which is sufficient for our purposes. In (3.4d) $R$ is the Ricci scalar of the metric $g_{AB}$. The scalar $R^*$ (denoted $\delta_n^i\, g^{kl}R_{ikl}{}^n$ in Bergmann's Eq. (17.51), in which the opposite sign convention to ours in (2.6) is used) is interpreted from the submersion viewpoint following (3.20).

We will now assume that we are working on the space $M$ of the submersion discussed in Sec. 2 and study the validity and interpretation of formulas (3.4) in that case.

With $i = 1$ and $e_\mu$ and $e_1$ given by (3.1), Eq. (2.14b) can be written in the form

$$e_{\mu|A}^B A^A e_{\nu B} = A^B{}_{|A} e_\mu^A e_{\nu B}. \tag{3.6}$$

Multiplying by $e_C^\nu$ and $e_D^\mu$ and using $e_C^\nu e_{\nu D} = g_{CD} - A_C A_D$ yields

$$e_D^\mu e_{\mu C|A}A^A = A_{C|D} - B_D A_C - B_C A_D. \tag{3.7}$$

A direct calculation gives

$$e_B^\mu e_C^\nu g_{\mu\nu,A}A^A = e_B^\mu e_{\mu C|A}A^A - e_B^\mu e_{\mu D|A}A^A A^D A_C + e_C^\mu e_{\mu B|A}A^A - e_C^\mu e_{\mu D|A}A^A A^D A_B. \tag{3.8}$$

Substitution of (3.7) into this and use of $B_C A^C = 0$ results in (3.4a). However, (2.14a) implies in the present case

$$0 = e_1(g_{\mu\nu}) = g_{\mu\nu,A}A^A, \tag{3.9}$$

and so $A^B$ must satisfy

$$A_{B|C} + A_{C|B} - A_B B_C - A_C B_B = 0. \tag{3.10}$$

This means that the integral curves of $A^B$ constitute a rigid congruence.

We next look at (3.4b). That this question is a special case of (2.21) can be seen as follows: Putting $V = W = e_1$, $X = e_\mu$, and $Y = e_\nu$ and using (2.10), the left-hand side of (2.21) becomes

$$-\langle D_{e_1}(F^1_{\mu\nu}e_1), e_1\rangle = -e_1(F^1_{\mu\nu}) = -A^B F^1{}_{\mu\nu,B}. \tag{3.11}$$

Using (2.4), (2.10), (3.3), and (3.5), we find

$$F^1{}_{\mu\nu} = -e_\mu^A e_\nu^B A_{AB} = -\varphi_{\mu\nu}. \tag{3.12}$$

From the definition (2.2a) of the tensor $T$,

$$T_{e_1}e_\mu = -B_A e_\mu^A e_1, \quad T_{e_1}e_1 = B^A e_A^\mu e_\mu. \tag{3.13}$$

Substituting into the right-hand side of (2.21) and using (2.8), we obtain the right-hand side of (3.4b).

Consider now (3.4c). This is a special case of (2.20) for using (2.10), (3.3), (3.12) and the first of (3.13), we have

$$\langle A_{e_\mu}e_\nu, T_{e_1}e_\rho\rangle = -\tfrac{1}{2}e_\mu^A e_\nu^B e_\rho^C A_{AB}B_C. \tag{3.14}$$

Hence, if $\mathfrak{S}$ denotes the cyclic sum over $\mu$, $\nu$, and $\rho$, we find

$$\mathfrak{S}\langle A_{e_\mu}e_\nu, T_{e_1}e_\rho\rangle = \tfrac{1}{2}e_\mu^A e_\nu^B e_\rho^C (B_A A_{CB} + B_B A_{AC} + B_C A_{BA}). \tag{3.15}$$

On the other hand, assuming $\mathscr{H}[e_\mu, e_\nu] = 0$,

$$\mathfrak{S}\langle (D_{e_\rho}A)_{e_\mu}e_\nu, e_\rho\rangle = \mathfrak{S}\langle D_{e_\rho}(A_{e_\mu}e_\nu), e_1\rangle = -\tfrac{1}{2}\{e_\rho(F^1{}_{\mu\nu}) + e_\nu(F^1{}_{\rho\mu}) + e_\mu(F^1{}_{\nu\rho})\}. \tag{3.16}$$

The first equality here is established in Ref. 11, Lemma 7. Substituting (3.12), (3.15), and (3.16) into (2.22) yields (3.4c).

Finally, turning to (3.4d), the fundamental equations (2.7) together with (2.10) yield, in this five-dimensional case,

$$R = R^* - \tfrac{3}{4}\|F\|^2 + 2g^{\mu\nu}R_{\mu 1 1 \nu}, \tag{3.17}$$

with $\|F\|^2 = F^1{}_{\mu\nu}F^1{}^{\mu\nu}$. However, using (2.8) and (3.13), we have

$$\begin{aligned} g^{\mu\nu}R_{\mu 1 1 \nu} &= g^{\mu\nu}\langle (D_{e_\mu}T)_{e_1}e_1, e_\nu\rangle - g^{\mu\nu}\langle T_{e_1}e_\mu, T_{e_1}e_\nu\rangle + \tfrac{1}{4}\|F\|^2 \\ &= g^{\mu\nu}\langle (D_{e_\mu}T)_{e_1}e_1, e_\nu\rangle - B^A B_A + \tfrac{1}{4}\|F\|^2 \\ &= B^A{}_{|A} + \tfrac{1}{4}\|F\|^2, \end{aligned} \tag{3.18}$$

and so (3.17) becomes

$$R = R^* - \tfrac{1}{4}\|F\|^2 + 2B^A{}_{|A}. \tag{3.19}$$

On the other hand, (3.10) and (3.12) can be used to show that

$$\tfrac{1}{4}\|F\|^2 = A^{D|C}A_{D|C} - B^A B_A, \tag{3.20}$$

and, when this is substituted into (3.19), we obtain agreement with (3.4d) on account of (3.10). From the submersion point of view, we see that $R^*$ in (3.4d) is the Ricci scalar of $B$ lifted to $M$ via $\pi$.

To obtain the Kaluza–Klein theory from this formalism, Bergmann[3] begins by assuming that $A^B$ is a Killing vector and thus $B_A = 0$ and (3.10) is satisfied. We see from (3.13) that from the submersion point of view Bergmann's assumption is equivalent to the vanishing of the tensor $T$, i.e., the fibers are totally geodesic. In this case (3.19) has the form of the Lagrangian density appearing in the Einstein–Maxwell theory and, of course, it is a special case of Lemma 1 in Sec. 2. In addition, with $B_A = 0$, (3.4b) facilitates the introduction of a special coordinate system[3] in which $\varphi_{\mu\nu}$ is independent of the fifth coordinate. The relationship between (3.4c) and the Bianchi identities for Maxwell's electromagnetic tensor emerges as a special case from the argument of Sec. 4.

## 4. THE BUNDLE VIEWPOINT

Sufficient conditions for a Riemannian submersion to be a fiber bundle are given in the following theorem due to Hermann[12]:

**Theorem**: If $M$ is complete as a Riemannian space, so is $B$. $M$ is then a locally trivial fiber space. If in addition the fibers of $\pi$ are totally geodesic submanifolds of $M$, then $\pi$: $M \to B$ is a fiber bundle with structure group the Lie group of isometries of the fiber.

We shall henceforth assume that the conditions of this theorem are satisfied. Thus in Sec. 2 we take the tensor $T = 0$ and if $\{e_i^\times\}$ are a basis for the Lie algebra of the structure group we may take $\{e_i\}$ to be the corresponding fundamental vector fields on $M$. Thus in addition to having $[e_i, e_j]$ vertical, we now have

$$[e_i, e_j] = C^k{}_{ij} e_k, \tag{4.1}$$

where $C^k{}_{ij}$ are the structure constants of the Lie algebra of the structure group with respect to the basis $\{e_i^\times\}$. We also have, from the theorem above, that $\{e_i\}$ are Killing vector fields of the fiber metric. The analytical form of this property is given by

$$e_k(g_{ij}) + 2g_{l(j} C^l{}_{i)k} = 0, \tag{4.2}$$

where the subscript parentheses denote symmetrization over $i$ and $j$.

Defining the Lie algebra-valued 1-form field on $M$,[7,8]

$$\omega = \theta^i e_i^\times, \tag{4.3}$$

one easily shows that

$$\omega(e_i) = e_i^\times, \quad \mathscr{L}_{e_i}\omega = [\omega, e_i^\times], \tag{4.4}$$

where the left-hand side of the second equation is the Lie derivative of $\omega$ with respect to the vector field $e_i$ and the bracket on the right-hand side is the Lie algebra bracket. The second equation in (4.4) states, in infinitesimal form, that $\omega$ is type Ad. Thus (4.3) is a (pseudotensorial) connection 1-form on the bundle. Defining the (tensorial) curvature 2-form of $\omega$ in the usual way

$$\Omega = d\omega + \tfrac{1}{2}[\omega, \omega] = \tfrac{1}{2}\Omega^i{}_{\mu\nu}\theta^\mu \wedge \theta^\nu e_i^\times, \tag{4.5}$$

we have from (4.3)

$$d\theta^i + \tfrac{1}{2}C^i{}_{jk}\theta^j \wedge \theta^k = \tfrac{1}{2}\Omega^i{}_{\mu\nu}\theta^\mu \wedge \theta^\nu = \Omega^i. \tag{4.6}$$

Using the fact that

$$d\theta^i(e_\mu, e_\nu) = -\tfrac{1}{2}\theta^i([e_\mu, e_\nu]) = \tfrac{1}{2}F^i{}_{\mu\nu} \tag{4.7}$$

[the last equality coming from (2.4) and (2.10)], we have

$$\Omega^i{}_{\mu\nu} = F^i{}_{\mu\nu}. \tag{4.8}$$

Taking the covariant exterior derivative of the second of (4.4), we obtain

$$\mathscr{L}_{e_i}\Omega = [\Omega, e_i^\times], \tag{4.9}$$

which can be rewritten in the form

$$e_j(F^i{}_{\mu\nu}) = C^i{}_{kj}F^k{}_{\mu\nu}. \tag{4.10}$$

This states, in infinitesimal form, that $\Omega$ is of type Ad.

When $T = 0$, we can write (2.20), at a point at which $\mathscr{H}[e_\mu, e_\nu] = 0$, in the form

$$\mathfrak{S} e_\rho(F^i{}_{\mu\nu}) = 0. \tag{4.11}$$

Using (4.6), we find that at a point at which $\mathscr{H}[e_\mu, e_\nu] = 0$ we have

$$d\Omega^i(e_\mu, e_\nu, e_\rho) = (1/3!)\mathfrak{S} e_\rho(F^i{}_{\mu\nu}), \tag{4.12}$$

and so (4.11) is equivalent to the Bianchi identities

$$d\Omega^i + C^i{}_{jk}\theta^j \wedge \Omega^k = 0 \tag{4.13}$$

(cf. Ref. 14, p. 78). Since this is a tensorial equation, the special choice of horizontal vector fields $\{e_\mu\}$ used to obtain it is legitimate.

When $T = 0$, (2.21) can be written, using (4.1) and the fact that $D$ is Riemannian, in the form

$$e_i(F^k{}_{\mu\nu})g_{kj} + e_j(F^k{}_{\mu\nu})g_{ik} + F^k{}_{\mu\nu}\{e_k(g_{ij}) + 2g_{l(i}C^l{}_{j)k}\} = 0. \tag{4.14}$$

Using (4.10), we find

$$F^k{}_{\mu\nu}e_k(g_{ij}) = 0. \tag{4.15}$$

Hence, to have no restriction on $F^k{}_{\mu\nu}$, could take

$$e_k(g_{ij}) = 0. \tag{4.16}$$

Then $g_{ij}$ must be *constant* along the fiber, and by (4.2), $C_{ijk} = g_{il}C^l{}_{jk}$ is skew-symmetric under interchange of any pair of indices.

## 5. DISCUSSION

When the conditions of Sec. 4 are satisfied (2.15) becomes the Lagrangian density of the Kaluza–Klein theory. The Ricci scalar $\hat{R}$ is then calculated with the invariant metric $g_{ij}$ and plays the role of a cosmological constant. It has been pointed out by Bradfield and Kantowski[9] that $\hat{R} = 0$ for certain Lie algebras and Kopczyński[8] has described a mechanism for removing it in general. The 1-form $\omega$ and the 2-form $\Omega$, pulled back to the base space $B$ via a local cross section, are the gauge potential and the Yang–Mills field, respectively.


## ACKNOWLEDGMENTS

I wish to thank Professor B. Smyth for drawing my attention to Riemannian submersions and Dr. J. D. McCrea and Dr. W. Kopczyński for helpful discussions at different stages of this work.



[1]T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **2**, 966 (1921).

[2]O. Klein, Z. Phys. **37**, 895 (1929).

[3]P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Englewood Cliffs, NJ, 1966), p. 254.
[4]A. Trautman, "The Applications of Fiber Bundles in Physics," notes on lectures given at King's College, London, September 1967, p. 42.
[5]A. Trautman, Rep. Math. Phys. **1**, 29 (1970).
[6]R. Kerner, Ann. Inst. H. Poincaré **9**, 143 (1968).
[7]Y. M. Cho, J. Math. Phys. **16**, 2029 (1975).
[8]W. Kopczyński, "A Fiber Bundle Description of Coupled Gravitational and Gauge Fields," in *Proceedings of the Conference on Differential Geometrical Methods in Mathematical Physics, Salamanca 1979*, Lecture Notes in Mathematics (Springer-Verlag, Berlin, 1980).
[9]T. Bradfield and R. Kantowski, J.Math. Phys. **23**, 128 (1982).
[10]Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975).
[11]B. O'Neill, Michigan Math. J. **13**, 459 (1966).
[12]R. Hermann, Proc. Am. Math. Soc. **11**, 236 (1960).
[13]N. J. Hicks, *Notes on Differential Geometry* (Van Nostrand, Princeton, NJ, 1965).
[14]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry I* (Interscience, New York, 1963).
[15]Y. Matsushima, *Differentiable Manifolds* (Marcel Dekker, New York, 1972).
[16]R. L. Bishop, in *Differential Geometry*, papers in honor of K. Yano (Kinokuniya, Tokyo, 1972), p. 21.

# "Polynomial constants" for the quantized NLS equation

K. M. Case
*The Rockefeller University, New York, New York 10021*

(Received 29 December 1983; accepted for publication 16 March 1984)

The classical nonlinear Schrödinger equation (NLS) is known to have an infinite number of polynomial constants. While recursion relations to compute these are available, no general expressions in terms of the fields have been found. However, general expressions have been obtained in terms of the reflection coefficients. When we turn to the quantum case where the fields become operators with conventional commutation relations, the polynomials with *suitable ordering* are still constants. The classical expression for the constants in terms of the reflection coefficients strongly suggests what the quantum form should be. This conjecture is proved for the repulsive case. The expression is significantly simpler than the classical one. It is $I_n = (1/2\pi)\int_{-\infty}^{\infty}(k)^n R^*(k)R(k)dk$.

PACS numbers: 11.10.Lm, 02.90. + p

## I. INTRODUCTION

Classically, the nonlinear Schrödinger equation

$$i\Psi_t = -\Psi_{xx} - 2\sigma\kappa\Psi^*\Psi\Psi \tag{1}$$

($\sigma = +1$ for the attractive case and $\sigma = -1$ for the repulsive case) is known to have an infinite number of polynomial constants of motion.[1] Thus, the first five are

$$I_0' \equiv N = \int_{-\infty}^{\infty} \Psi^*\Psi\, dx,$$

$$I_1' \equiv P = \int_{-\infty}^{\infty} \Psi^* \frac{\partial_x}{i} \Psi\, dx,$$

$$I_2' \equiv H = \int_{-\infty}^{\infty} \{\partial_x\Psi^*\,\partial_x\Psi - \sigma\kappa\Psi^{*2}\Psi^2\}dx,$$

$$I_3' = i\int_{-\infty}^{\infty} \{\Psi^*\,\partial_x^3\Psi + \tfrac{3}{2}\kappa\sigma\Psi^{*2}\,\partial_x\Psi^2\}dx, \tag{2}$$

$$I_4' = \int_{-\infty}^{\infty} \{\Psi^*_{xx}\Psi_{xx} - 2\sigma\kappa(\Psi^{*2})_x(\Psi^2)_x - \sigma\kappa\Psi^{*2}(\Psi_x)^2 - \sigma\kappa(\Psi^*_x)^2\Psi^2 + \kappa^2[\Psi^{*3}\Psi^3 + \Psi^{*2}\Psi\Psi^*\Psi^2]\}dx.$$

(Of course classically the order in which the $\Psi^*$ and $\Psi$ are written is unimportant. However, it will be seen that the order given will be useful later when the $\Psi$'s are operators.)

We present some remarks.

(1) These constants are in involution.

(2) They can be obtained from the coefficients in the Laurent expansion of $a(\xi)$ (defined below).

(3) While recursion relations permit us to calculate these polynomials successively, no general closed form expression for these seems to be available. However, a relatively simple closed form expression does exist in terms of reflection coefficients.

(4) Comments on the construction of $I_n'$, $n > 4$, are given in Appendix B.

Here we wish to investigate the quantum case.[2–4] Thus, $\Psi$, $\Psi^*$ are assumed to be operators satisfying

$$[\Psi(x),\Psi(x')] = 0 = [\Psi^*(x),\Psi^*(x')], \tag{3}$$

and

$$[\Psi(x),\Psi^*(x')] = \delta(x - x').$$

It is known that $I_0'$, $I_1'$, and $I_2'$ are again constants. Further, when expressed[4] in terms of the reflection operators $R(k)$, $R^*(k)$ they have the form

$$I_0' = \frac{1}{2\pi}\int_{-\infty}^{\infty} R^*(k)R(k)dk,$$

$$I_1' = \frac{1}{2\pi}\int_{-\infty}^{\infty} kR^*(k)R(k)dk, \tag{4}$$

$$I_2' = \frac{1}{2\pi}\int_{-\infty}^{\infty} k^2R^*(k)R(k)dk.$$

Our purpose is the following: From the commutation relations of the reflection coefficients it is readily shown that if we define $I_n$ by

$$I_n = \frac{1}{2\pi}\int_{-\infty}^{\infty} k^nR^*(k)R(k)dk, \tag{5}$$

then

$$[I_n,I_m] = 0, \quad \text{all } n,m. \tag{6}$$

Thus, we have an infinite set of commuting constants of motion. Here we will show that the $I_n$ defined by Eq. (5) are precisely the quantum analog of the classical polynomial constants, i.e., when expressed in the field variables they are polynomials.

It will be seen that the following hold.

(i) When $a(\xi)$ is expanded in powers of $1/\xi$, we obtain in each successive term a new constant.

(ii) The quantum form of the constants when expressed in terms of the reflection operators are significantly simpler than the classical form.

The program to be followed is so. In Sec. II, we briefly summarize well-known results to have them in the notation we want to use. Section III recalls what is a dispersion relation for the Zakharov–Shabat function $\psi$. (Here we are restricted to $\sigma = -1$.) Passing to the limit $x \to -\infty$, gives a singular integral equation for $a(\xi)$ in terms of the reflection coefficients. In the classical case, this is solved in closed form. Expanding ln $a$ in terms of $1/\xi$ gives the well-known result.[5] From this we can readily conjecture what the quantum result should be. However, to treat the quantum case rigorously, we solve the integral equation by a Neumann series. The constants are then obtained by further expanding

 0022-2488/84/072306-09$02.50  

in $1/\xi$. An essential simplification occurs. Only a finite number of terms of the Neumann series contribute to the coefficient of a given power of $1/\xi$. As we go from the coefficient of $\xi^{-n}$ to that of $\xi^{-n-1}$, we obtain precisely *one* new constant $I_n$. The other terms in the coefficient are merely polynomials in the lower-order constants.

It is then shown that the (quantum) $I_n'$ of Eqs. (2) give precisely the same result when acting on a large class of states as do the $I_n$ in Eqs. (5). For the *repulsive* case these states are complete. Hence, we have the identification.

## II. SUMMARY OF NEEDED FORMULAS

To the quantized nonlinear Schrödinger equation, we associate an operator Zakharov–Shabat eigenvalue problem

$$v_{1x} - (i\xi/2)v_1 = \kappa^{1/2} v_2 \Psi, \qquad v_{2x} + (i\xi/2)v_2 = -\sigma\kappa^{1/2}\Psi^* v_1. \tag{7}$$

Conventionally, one defines four different solutions by boundary conditions at $\pm\infty$. Thus, $\phi$ is defined by

$$\lim_{x\to-\infty} \phi e^{-i\xi x/2} = \begin{pmatrix}1\\0\end{pmatrix}, \tag{8}$$

$\overline{\phi}$ by

$$\lim_{x\to-\infty} \overline{\phi} e^{+i\xi x/2} = \begin{pmatrix}0\\-1\end{pmatrix}, \tag{9}$$

$\psi$ by

$$\lim_{x\to+\infty} \psi e^{i\xi x/2} = \begin{pmatrix}0\\1\end{pmatrix}, \tag{10}$$

and $\overline{\psi}$ by

$$\lim_{x\to\infty} \overline{\psi} e^{-i\xi x/2} = \begin{pmatrix}1\\0\end{pmatrix}. \tag{11}$$

The boundary conditions and the differential equation can be combined in the integral equations

$$\phi_1(x) = e^{i\xi x/2} + \kappa^{1/2}\int_{-\infty}^{x} e^{+i\xi(x-x')/2}\phi_2\Psi\,dx', \qquad \phi_2(x) = -\sigma\kappa^{1/2}\int_{-\infty}^{x} e^{-i\xi(x-x')/2}\Psi^*\phi_1\,dx', \tag{12}$$

$$\overline{\phi}_1(x) = \kappa^{1/2}\int_{-\infty}^{x} e^{i\xi(x-x')/2}\overline{\phi}_2(x')\Psi(x')dx', \qquad \overline{\phi}_2(x) = -e^{-i\xi x/2} - \sigma\kappa^{1/2}\int_{-\infty}^{x} e^{-i\xi(x-x')/2}\Psi^*(x')\overline{\phi}_1\,dx', \tag{13}$$

$$\psi_1(x) = -\kappa^{1/2}\int_{x}^{\infty} e^{i\xi(x-x')/2}\psi_2(x')\Psi(x')dx', \qquad \psi_2(x) = e^{-i\xi x/2} + \sigma\kappa^{1/2}\int_{x}^{\infty} e^{-i\xi(x-x')/2}\Psi^*(x')\psi_1(x')dx', \tag{14}$$

$$\overline{\psi}_1(x) = e^{i\xi x/2} - \kappa^{1/2}\int_{x}^{\infty} e^{i\xi(x-x')/2}\overline{\psi}_2(x')\Psi(x')dx', \qquad \overline{\psi}_2(x) = \sigma\kappa^{1/2}\int_{x}^{\infty} e^{-i\xi(x-x')/2}\Psi^*(x')\overline{\psi}_1(x')dx'. \tag{15}$$

From these integral equations and the commutation relations of Eqs. (3), we readily find the following.

(1) At the same $x$ the solutions of Eqs. (7) defined with boundary conditions at $-\infty$ commute with those defined by conditions at $+\infty$, e.g.,

$$[\phi_i(x),\psi_j(x)] = 0. \tag{16}$$

(2)

$$[\phi_1(x),\Psi(x)] = 0 = [\phi_2(x),\Psi^*(x)], \tag{17}$$

$$[\phi_2(x),\Psi(x)] = (\sigma\kappa^{1/2}/2)\phi_1(x), \tag{18}$$

$$[\phi_1(x),\Psi^*(x)] = (\kappa^{1/2}/2)\phi_2(x), \tag{19}$$

$$[\psi_1(x),\Psi(x)] = 0 = [\psi_2(x),\Psi^*(x)], \tag{20}$$

$$[\psi_1(x),\Psi^*(x)] = (-\kappa^{1/2}/2)\psi_2(x), \tag{21}$$

$$[\psi_2(x),\Psi(x)] = (-\sigma\kappa^{1/2}/2)\psi_1(x), \tag{22}$$

$$[\overline{\psi}_1(x),\Psi(x)] = 0 = [\overline{\psi}_2(x),\Psi^*(x)], \tag{23}$$

$$[\overline{\psi}_1(x),\Psi^*(x)] = (-\kappa^{1/2}/2)\overline{\psi}_2(x), \tag{24}$$

$$[\overline{\psi}_2(x),\Psi(x)] = (-\sigma\kappa^{1/2}/2)\overline{\psi}_1(x). \tag{25}$$

Comments about the derivation of these results are given in Appendix C. The scattering data $a,b$ are defined by

$$\lim_{x\to\infty} \phi_1(x,\xi)e^{-i\xi x/2} = a(\xi), \tag{26}$$

$$\lim_{x\to\infty} \phi_2(x,\xi)e^{i\xi x/2} = b(\xi). \tag{27}$$

It follows that

$$\overline{b}(\xi) = \lim_{x\to\infty} \overline{\phi}_1(x,\xi)e^{-i\xi x/2}, \tag{28}$$

and

$$\overline{a}(\xi) = -\lim_{x\to\infty} \overline{\phi}_2(x,\xi)e^{+i\xi x/2}, \tag{29}$$

where $\overline{a} = a^*$, $\overline{b} = \sigma b^*$, and $*$ denotes the complex conjugate classically and Hermitian conjugate quantum mechanically.

More generally we have

$$a(\xi) = \phi_1(x,\xi)\psi_2(x,\xi) - \phi_2(x,\xi)\psi_1(x,\xi), \tag{30}$$

from which we can also obtain a formula that will be very useful

$$a(\xi) = \lim_{x\to-\infty} \psi_2(x,\xi)e^{i\xi x/2}. \tag{31}$$

Similarly, we have

$$b(\xi) = \phi_2(x,\xi)\overline{\psi}_1(x,\xi) - \phi_1(x,\xi)\overline{\psi}_2(x,\xi). \tag{32}$$

Given Eqs. (30) and (32) and the commutation relations of Eqs. (16)–(25), one readily computes the commutation relations of the scattering data with $\Psi$ and $\Psi^*$. These are conveniently summarized as follows.

Let $v,v'$ be commuting solutions of Eq. (7). Then the three-vector $\chi$ constructed as

$$\chi_1 = v_2' v_2, \qquad \chi_2 = -v_1' v_1, \qquad \chi_3 = (v_1' v_2 + v_2' v_1)/2, \tag{33}$$

satisfies the equations

$$\partial_x\chi_1 + i\xi\chi, = -2\sigma\kappa^{1/2}\Psi^*\chi_3, \qquad -\partial_x\chi_2 + i\xi\chi_2 = 2\kappa^{1/2}\chi_3\Psi, \qquad \partial_x\chi_3 = \kappa^{1/2}\chi_1\Psi + \sigma\kappa^{1/2}\Psi^*\chi_2. \tag{34}$$

Then for $\Lambda = a,b,\overline{a},\overline{b}$ we can associate a $\chi^{(\Lambda)}$ such that

$$[\Lambda,\Psi(x)] = \sigma\kappa^{1/2}\chi_2^{(\Lambda)}(x), \qquad [\Lambda,\Psi^*(x)] = \kappa^{1/2}\chi_1^{(\Lambda)}(x), \tag{35}$$

and

$$[\chi_2^{(A)}(x),\Psi(x)] = [\chi_1^{(A)}(x),\Psi^*(x)] = [\chi_3^{(A)}(x),\Psi(x)] = [\chi_3^{(A)}(x),\Psi^*(x)] = 0. \quad (36)$$

Further

$$[\chi_1^{(A)},\Psi] = (\sigma\kappa^{1/2}/2)A,$$

and (37)

$$[\chi_2^{(A)},\Psi^*] = (\kappa^{1/2}/2)A.$$

It is amusing to note that in virtue of these commutation relations, the Eqs. (34) for $\chi^{(A)}$ are quite *insensitive* to ordering. Indeed if $\alpha+\beta=1$, we see that

$$\begin{aligned}\partial_x\chi_1^{(A)} + i\xi\chi_1^{(A)} &= -2\sigma\kappa^{1/2}(\alpha\Psi^*\chi_3^{(A)} + \beta\chi_3^{(A)}\Psi^*),\\ -\partial_x\chi_2^{(A)} + i\xi\chi_2^{(A)} &= 2\kappa^{1/2}(\alpha\chi_3^{(A)}\Psi + \beta\Psi\chi_3^{(A)}),\\ \partial_x\chi_3^{(A)} &= \kappa^{1/2}(\alpha\chi_1^{(A)}\Psi + \beta\Psi\chi_2^{(A)})\\ &\quad + \sigma\kappa^{1/2}(\alpha\Psi^*\chi_2^{(A)} + \beta\chi_2^{(A)}\Psi^*).\end{aligned} \quad (38)$$

Explicitly we have

$$\chi^{(a)}(x,\xi) = \begin{pmatrix}\phi_2(x,\xi)\psi_2(x,\xi)\\ -\phi_1(x,\xi)\psi_1(x,\xi)\\ \frac{1}{2}\{\phi_1(x,\xi)\psi_2(x,\xi) + \phi_2(x,\xi)\psi_1(x,\xi)\}\end{pmatrix}, \quad (39)$$

$$\chi^{(b)}(x,\xi) = \begin{pmatrix}-\phi_2(x,\xi)\bar\psi_2(x,\xi)\\ \phi_1(x,\xi)\bar\psi_1(x,\xi)\\ -\frac{1}{2}\{\phi_1(x,\xi)\bar\psi_2(x,\xi) + \phi_2(x,\xi)\bar\psi_1(x,\xi)\}\end{pmatrix}.$$

Other important commutation relations obtained in the referenced work[2–4] (for the case $\sigma=-1$) are the following. Let

$$R^*(\xi) = (i/\sqrt{\kappa})b(\xi)a^{-1}(\xi), \quad (40)$$

then

$$\begin{aligned}R(\xi)R(\xi') &= S^{-1}(\xi,\xi')R(\xi')R(\xi),\\ R^*(\xi)R^*(\xi') &= S^{-1}(\xi,\xi')R^*(\xi')R^*(\xi),\\ R(\xi)R^*(\xi') &= S(\xi,\xi')R^*(\xi')R(\xi) + 2\pi\delta(\xi-\xi').\end{aligned} \quad (41)$$

Here $S(\xi,\xi')$ is a $c$-number whose only properties we need here are

$$S^{-1}(\xi,\xi') = S(\xi',\xi) = S^*(\xi,\xi'), \quad (42)$$

and $S(\xi,\xi) = -1$.

## III. THE INTEGRAL EQUATION

In Ref. 3, a dispersion relation for the repulsive case has been obtained. With the present notation this is

$$e^{-i\xi x/2}\bar\psi = \begin{pmatrix}1\\0\end{pmatrix} + \frac{\sqrt{\kappa}}{2\pi}\int_{-\infty}^{\infty}\frac{R^*(\xi')\psi(x,\xi')e^{-i\xi' x/2}\,d\xi'}{\xi'-\xi-i\epsilon}. \quad (43)$$

Since in the case $\sigma=-1$, we have

$$\bar\psi = \begin{pmatrix}\psi_2^*\\ \psi_1^*\end{pmatrix}, \quad (44)$$

the Eqs. (43) are

$$e^{-i\xi x/2}\psi_2^*(x,\xi) = 1 + \frac{\sqrt{\kappa}}{2\pi}\int_{-\infty}^{\infty}\frac{R^*(\xi')\psi_1(x,\xi')e^{-i\xi' x/2}d\xi'}{\xi'-\xi-i\epsilon},$$

$$e^{-i\xi x/2}\psi_1^*(x,\xi) = \frac{\sqrt{\kappa}}{2\pi}\int_{-\infty}^{\infty}\frac{R^*(\xi')\psi_2(x,\xi')e^{-i\xi' x/2}d\xi'}{\xi'-\xi-i\epsilon}. \quad (45)$$

Taking the limit $x\to-\infty$ and noting the expression of Eq. (31) for $a(\xi)$ gives

$$a(\xi) = 1 - \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty}\frac{R^*(\xi')a(\xi')R(\xi')d\xi'}{\xi'-\xi+i\epsilon}. \quad (46)$$

We now plan to solve this for $a(\xi)$ as a function of $R^*,R$. It is well known that $a(\xi)$ is a constant for all $\xi$. (Below this will be seen to follow as one of a set of relations.) In particular, then if we expand $a(\xi)$ [or any function of $a(\xi)$] in a Laurent series in $\xi$, each coefficient will be a constant.

## IV. THE CLASSICAL SOLUTION

Nothing in the derivation of Eq. (46) is changed if all quantities are treated classically, i.e., they commute. It is interesting to treat this case to see that this singular integral equation does indeed have a solution. In addition, this leads to a rather convincing (if heuristic) "proof" of our general result.

The solution is as follows. Since all quantities are now classical, we can write Eq. (46) as

$$a(\xi) = 1 - \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty}\frac{f(\xi')a(\xi')d\xi'}{\xi'-(\xi-i\epsilon)}, \quad (47)$$

where

$$f(\xi') = R^*(\xi')R(\xi'). \quad (48)$$

Let

$$N(z) = 1 - \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty}\frac{f(\xi')a(\xi')d\xi'}{\xi'-z}, \quad (49)$$

then (i) $N(z)$ is analytic in the complex $z$ plane cut along the real axis; (ii) $N(z)\to 1$ as $|z|\to\infty$; (iii) the boundary value $(N_-(\xi))$ as $z$ approaches the real axis from below is

$$N_-(\xi) \equiv 1 - \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty}\frac{f(\xi')a(\xi')d\xi'}{\xi'-(\xi-i\epsilon)};$$

and (iv) The difference of the boundary values of $N$ is

$$N_+(\xi) - N_-(\xi) = -\kappa a(\xi)f(\xi).$$

Thus, Eq. (47) says that

$$N_+ - (1-\kappa f)N_- = 0. \quad (50)$$

Let

$$X(z) = \exp\left\{-\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\ln[1-\kappa f(\xi')]d\xi'}{\xi'-z}\right\},$$

then $X(z)$ is analytic and nonzero in the cut plane, and $X(z)\to 1$ as $|z|\to\infty$. Further $X_-(\xi)/X_+(\xi) = 1-\kappa f(\xi)$. Equation (50) reads $X_+N_+ - X_-N_- = 0$.

$\therefore M(z) = X(z)N(z)$ is analytic everywhere and goes to 1 at $\infty$. We conclude $M(z)=1$ and thus $N(z)=1/X(z)$. But Eq. (47) then tells us that $a(\xi) = N_-(\xi)$. We conclude that

$$\ln a(\xi) = -\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\ln[1-\kappa f(\xi')]d\xi'}{\xi'-(\xi-i\epsilon)}. \quad (51)$$

If we write $\ln a(\xi) = -\Sigma_{n=0}^{\infty} d_n/\xi^{n+1}$, expand the right side of Eq. (51) in powers of $1/\xi$, and equate coefficients we obtain

$$d_n = -\frac{1}{2\pi i}\int_{-\infty}^{\infty}(\xi')^n \ln[1-\kappa R^*(\xi')R(\xi')]d\xi'. (52)$$

What should the quantum form for these constants be? A reasonable conjecture is that it should be Eq. (52) with normal ordering, i.e., all $R$ to the right of all $R^*$. Then Eq. (52) becomes

$$d_n = \frac{1}{2\pi i}\int_{-\infty}^{\infty}(\xi')^n\left\{\kappa R^*R + \frac{\kappa^2}{2}R^{*2}R^2 + \frac{\kappa^3}{3}R^{*3}R^3 + \cdots\right\}d\xi'.$$

However, the commutation rules of Eqs. (41) and (42) tell us that $R^{*2}(\xi') = 0 = R^2(\xi')$. Therefore, we expect the quantum constants $d_n$ to be $d_n = (\kappa/i)(1/2\pi)$ $\times\int_{-\infty}^{\infty}(\xi')^n R^*(\xi')R(\xi')d\xi'$. We verify this in the next section.

## V. THE QUANTUM SOLUTION

We now want to solve Eq. (46) when $a$, $R$, and $R^*$ are operators. The simple approach using the theory of functions of a complex variable does not seem to be applicable. However, formally we can obtain a solution by a Neumann series. Thus, we imagine $\kappa$ being replaced by $\epsilon\kappa$ and iteratively obtain a power series in $\epsilon$. Aside from the possible complicated structure of the solution so obtained, we have to consider whether the series so found for quantities of interest converges as $\epsilon\to 1$. We write

$$a(\xi) = \sum_{n=0}^{\infty} a_n(\xi). \quad (53)$$

Choose $a_0(\xi) = 1$, and obtain from Eq. (46) the recursion relation

$$a_n(\xi) = \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty}\frac{R^*(k_n)a_{n-1}(k_n)R(k_n)dk_n}{\xi-k_n-i\epsilon}, \quad n\geqslant 1.$$

The general term is obviously

$$a_n(\xi) = \left(\frac{\kappa}{2\pi i}\right)^n\int_{-\infty}^{\infty}\iint\frac{dk_1\cdots dk_n R^*(k_n)\cdots R^*(k_1)R(k_1)\cdots R(k_n)}{\Pi_{j=1}^{n-1}(k_{j+1}-k_j-i\epsilon)(\xi-k_n-i\epsilon)}. \quad (54)$$

If we expand in powers of $1/\xi$, we obtain

$$a(\xi) = 1 + \sum_{m=0}^{\infty}\frac{C_m}{\xi^{m+1}}, \quad (55)$$

where [using Eqs. (53) and (54)]

$$C_m = \sum_{n=1}^{\infty} C_m^{(n)}, \quad (56)$$

with

$$C_m^{(n)} = \left(\frac{\kappa}{2\pi i}\right)^n \times\int_{-\infty}^{\infty}\iint\frac{dk_1\cdots dk_n\,(k_n)^m F_n(k_1,\ldots,k_n)}{\Pi_{j=1}^{n-1}(k_{j+1}-k_j-i\epsilon)}.$$

Here

$$F_n = R^*(k_n)\cdots R^*(k_1)R(k_1)\cdots R(k_n). \quad (57)$$

## VI. PROPERTIES OF THE $C_m^{(n)}$

Notice that $F_n$ is a symmetric function. Indeed interchanging two $R^*$ gives a factor just inverse to that obtained by interchanging the corresponding two $R$'s. Further $F_n = 0$ when any two arguments are equal since

$$R^2(k) = 0 = R^{*2}(k). \quad (58)$$

Two immediate consequences are that the $i\epsilon$ can be omitted and the singularities can be interpreted as principal values. Also orders of integration can be arbitrarily interchanged.

From this a basic theorem follows. It is

$$C_m^{(n)} = 0 \quad \text{if } n > m+1. \quad (59)$$

Thus, the question of convergence of the series for $C_m$ is answered. It is the sum of a finite number of terms. Some other consequences are as follows.

(i) $$C_m^{(1)} = \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty}k^m R^*(k)R(k)dk \equiv \frac{\kappa}{i}I_m. \quad (60)$$

(Here the second line is a definition of $I_m$.)

(ii) For $n > 1$, $C_m^{(n)}$ is a sum of products of $C_{m_i}^{(1)}$, where $\Sigma_i m_i \leqslant m-1$. In particular then when we go from $m$ to $m+1$, we obtain the one *new* constant $I_{m+1}$.

(iii) For $m \geqslant n-1$,

$$C_m^{(n)} = \left(\frac{\kappa}{2\pi i}\right)^n\int_{\infty}^{\infty}\iint dk_1\cdots dk_n \times F_n(k_1,\ldots,k_n)S_m(k_1,\ldots,k_n), \quad (61)$$

where $S_m$ is a homogeneous symmetric polynomial of degree $m-(n-1)$. For example,

$$S_{n-1} = 1/n!,$$

$$S_n = \sum_{i=1}^{n}\frac{k_i}{n!},$$

$$S_{n+1} = \frac{\{\Sigma_{j=1}^{n}(k_i)^2 + \Sigma_{i<j}k_ik_j\}}{n!}.$$

The proofs of these properties are somewhat tedious if straightforward. Accordingly, we relegate them to an Appendix. However, they are essentially based on the simple lemma.

*Lemma*: If $g(k_1,\ldots,k_n)$ is symmetric and vanishes when any two arguments are equal, then

$$\int\cdots\int_{-\infty}^{\infty}dk_1\cdots dk_n\frac{g(k_1,\ldots,k_n)}{\Pi_{j=1}^{n-1}(k_{j+1}-k_j)} = 0. \quad (62)$$

*Proof*: Let the integrand in Eq. (62) be $\{\ \}_1$. With our assumptions relabel with $j\to j+1$, $n\to 1$. The integrand is then

$$\{\ \}_2 = \frac{k_2-k_1}{k_1-k_n}\{\ \}_1.$$

Next do the same permutation on $\{\quad\}_2$. We obtain

$$\{\quad\}_3 = \frac{k_3 - k_2}{k_1 - k_n}\{\quad\}_1.$$

Do this $n-1$ times and average the $n$ equivalent integrands. Then

$$\{\quad\}_1 = \frac{1}{n}\left\{1 + \sum_{j=1}^{n-1}\frac{(k_{j+1}-k_j)}{k_1 - k_n}\right\}\{\quad\}_1$$
$$= \frac{1}{n}\left\{1 + \frac{k_n - k_1}{k_1 - k_n}\right\}\{\quad\}_1 \approx 0.$$

## VII. EXPLICIT EXPRESSIONS FOR THE $C_m$

With the properties obtained we have for lowest-order constants: for $m=0$,

$$C_0^{(1)} = \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty} F_1\, dk_1;$$

for $m=1$

$$C_1^{(1)} = \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty} k_1 F_1\, dk_1,$$

$$C_1^{(2)} = \left(\frac{\kappa}{2\pi i}\right)^2 \int_{-\infty}^{\infty}\!\!\int dk_1\, dk_2\, F_2;$$

for $m=2$

$$C_2^{(1)} = \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty} k_1^2 F_1\, dk_1,$$

$$C_2^{(2)} = \left(\frac{\kappa}{2\pi i}\right)^2 \iint_{-\infty}^{\infty} dk_1\, dk_2\, k_2 F_2,$$

$$C_2^{(3)} = \left(\frac{\kappa}{2\pi i}\right)^3 \frac{1}{3!}\iint_{-\infty}^{\infty}\!\!\int dk_1\, dk_2\, dk_3\, F_3;$$

and for $m=3$

$$C_3^{(1)} = \frac{\kappa}{2\pi i}\int_{-\infty}^{\infty} k_1^3\, dk_1\, F_1,$$

$$C_3^{(2)} = \left(\frac{\kappa}{2\pi i}\right)^2 \iint_{-\infty}^{\infty} dk_1\, dk_2 \left\{k_2^2 + \frac{k_1 k_2}{2}\right\} F_2,$$

$$C_3^{(3)} = \left(\frac{\kappa}{2\pi i}\right)^3 \iint_{-\infty}^{\infty}\!\!\int dk_1\, dk_2\, dk_3\, k_3 F_3,$$

$$C_3^{(4)} = \left(\frac{\kappa}{2\pi i}\right)^4 \iiint_{-\infty}^{\infty}\!\!\int dk_1\, dk_2\, dk_3\, dk_4\, F_4.$$

The commutation relations (Eqs. 41) show that $R^*(k_1)R(k_1)R(k_2) = R(k_2)R^*(k_1)R(k_1) - 2\pi\delta(k_1-k_2)R(k_1)$. From this it follows that $F_2 = R(k_2)R(k_2)R^*(k_1)R(k_1) - 2\pi\delta(k_1-k_2)R^*(k_1)R(k_1)$ and

$$\begin{aligned}F_3 = {} & R^*(k_3)R(k_3)R^*(k_2)R(k_2)R^*(k_1)R(k_1)\\ & - 2\pi R^*(k_3)R(k_3)R^*(k_1)R(k_1)\delta(k_3-k_2)\\ & - 2\pi R^*(k_3)R(k_3)R^*(k_2)R(k_2)\delta(k_3-k_1)\\ & - 2\pi R^*(k_3)R(k_3)R^*(k_1)R(k_1)\delta(k_2-k_1)\\ & + (2\pi)^2 R^*(k_3)R(k_3)\delta(k_1-k_3)\delta(k_2-k_3)\\ & + (2\pi)^2 R^*(k_3)R(k_3)\delta(k_1-k_2)\delta(k_1-k_3).\end{aligned}$$

Inserting these expressions into the $C_m^{(n)}$ then yields

$$C_0 = (\kappa/i)I_0,$$

$$C_1 = \frac{\kappa}{i}I_1 + \left(\frac{\kappa}{i}\right)^2\frac{I_0(I_0-1)}{2},$$

$$C_2 = \frac{\kappa}{i}I_2 + \left(\frac{\kappa}{i}\right)^2 I_1(I_0-1) + \left(\frac{\kappa}{i}\right)^3\frac{I_0(I_0-1)(I_0-2)}{3!},$$

$$\begin{aligned}C_3 = {} & \frac{\kappa}{i}I_3 + \left(\frac{\kappa}{i}\right)^2\{I_2(I_0-1) + \tfrac{1}{2}(I_1^2 - I_2)\}\\ & + \left(\frac{\kappa}{i}\right)^3\frac{I_1(I_0-1)(I_0-2)}{2}\\ & + \left(\frac{\kappa}{i}\right)^4\frac{I_0(I_0-1)(I_0-2)(I_0-3)}{4!}.\end{aligned}$$

[Remember: Our definition is $I_n = (1/2\pi)\int_{-\infty}^{\infty} k_n \times R^*(k)R(k)dk$.]

## VIII. IDENTIFICATION OF THE $I_n$ and $I'_n$

We want to identify the $I_n$ just introduced with the "polynomials" described in the Introduction. To do so, first let us see the effect of the $I_n$ on a complete set of states. For the repulsive case such a set are the vacuum $|0\rangle$ and

$$|k_1\cdots k_m\rangle \equiv R^*(k_1)\cdots R^*(k_n)|0\rangle,$$
$$m = 1,2,\ldots,\quad -\infty < k_i < \infty. \tag{63}$$

From the commutation relations of Eqs. (41), we see that

$$[I_n, R^*(k_i)] = (k_i)^n R^*(k_i), \tag{64}$$

and [using $R(k)|0\rangle = 0$] that

$$I_n|k_1\cdots k_m\rangle = \left(\sum_{i=1}^m (k_i)^n\right)|k_1\cdots k_m\rangle. \tag{65}$$

What is the result of applying the $I'_n$ of the Introduction to these states? To find this, we need the commutators

$[R^*(k_i), I'_n]$,

i.e., in virtue of the definition of Eq. (40), we need

$[\Lambda, I'_n]$, for $\Lambda = a,b$.

We maintain that the fundamental relation

$$[\Lambda(\xi), I'_n] = \xi^n\{\chi_3^{(\Lambda)}(\xi)\}_{-\infty}^{\infty} \tag{66}$$

holds. Since the $\chi_3^{(\Lambda)}$ are combinations of $\phi$, $\bar{\phi}$, $\psi$, and $\bar{\psi}$ we know the limits at $\pm\infty$. In particular,

$$\chi_3^{(a)}(\infty,\xi) = \chi_3^{(a)}(-\infty,\xi) = a(\xi)/2, \tag{67}$$

and

$$\chi_3^{(b)}(\infty,\xi) = -\chi_3^{(b)}(-\infty,\xi) = -b(\xi)/2. \tag{68}$$

Therefore,

$$[a(\xi), I'_n] = 0, \tag{69}$$

and

$$[b(\xi), I'_n] = -\xi^n b(\xi). \tag{70}$$

From Eq. (69) with $n=2$ we conclude that $a(\xi)$ is a constant for all $\xi$. For general $n$ we then see that

$$[I_m, I'_n] = 0,\quad \text{all } n,m. \tag{71}$$

Combining Eqs. (69) and (70) yields

$$[I'_n, R^*(k_i)] = (k_i)^n R^*(k_i), \tag{72}$$

and thus

$$I'_n|k_1\cdots k_m\rangle = \left(\sum_{i=1}^m (k_i)^n\right)|k_1\cdots k_m\rangle. \tag{73}$$

*Remark*: This result holds for both the repulsive and attractive case. However, only in the repulsive case are the states $|k_1 \cdots k_m\rangle$ complete. Thus, only in the repulsive case can we conclude that

$$I'_n = I_n. \tag{74}$$

## IX. JUSTIFICATION OF THE FUNDAMENTAL RELATION

To verify Eq. (66) one can proceed so: Note that in the classical case (commutators replaced by Poisson brackets) the equation is easily proved. Thus, one shows the analog of Eq. (66) holds for $n = 0$ and using a recursion relation for the $I'_n$ that

$$\{A(\xi), I'_n\} = \xi\{A(\xi), I'_{n-1}\}. \tag{75}$$

The analog of Eq. (66) is thus obtained by induction.

In the quantum case one would expect there are polynomial constants of the classical form with *some* suitable ordering. This is how the quantum constants $I'_n$ of Eqs. (2) were constructed. Thus with the ordering given they satisfy

$$[A(\xi), I'_0] = \{\chi_3^{(A)}(x,\xi)\}_{x=-\infty}^{x=+\infty} \tag{76}$$

and

$$[A(\xi), I'_n] = \xi[A(\xi), I'_{n-1}]. \tag{77}$$

Let us see how this comes about. We have

$$[A, I'_0] = \int_{-\infty}^{\infty} [A, \Psi^*(x)\Psi(x)]dx$$
$$= \int_{-\infty}^{\infty} \{\Psi^*(x)[A,\Psi(x)] + [A,\Psi^*(x)]\Psi(x)\}dx.$$

Using the commutation relations of Eqs. (35) this becomes

$$[A, I'_0] = \int_{-\infty}^{\infty} \{\Psi^*\sigma\kappa^{1/2}\chi_2^{(A)} + \kappa^{1/2}\chi_1^{(A)}\Psi\}dx. \tag{78}$$

The last of Eq. (34) then shows that

$$[A, I'_0] = \int_{-\infty}^{\infty} \partial_x \chi_3^{(A)} dx = \chi_3^{(A)}(x,\xi)|_{-\infty}^{\infty}. \tag{79}$$

Next consider

$$[A, I'_1] = \int_{-\infty}^{\infty} \left[A, \Psi^* \frac{\partial_x}{i}\Psi\right]dx$$
$$= \int_{-\infty}^{\infty} \left\{\Psi^* \frac{\partial_x}{i}[A,\Psi] - \left(\frac{\partial_x}{i}[A,\Psi^*]\right)\Psi\right\}dx$$
$$= \int_{-\infty}^{\infty} \left\{\Psi^* \frac{\sigma\kappa^{1/2}}{i}\partial_x\chi_2^{(A)} - \frac{\kappa^{1/2}}{i}(\partial_x\chi_1^{(A)})\Psi\right\}dx. \tag{80}$$

The first two of Eqs. (34) then show that

$$[A, I'_1] = \int_{-\infty}^{\infty} \left\{\Psi^* \frac{\sigma\kappa^{1/2}}{i}(i\xi\chi_2^{(A)} - 2\kappa^{1/2}\chi_3^{(A)}\Psi) \times \frac{-\kappa^{1/2}}{2}(-i\xi\chi_1^{(A)} - 2\sigma\kappa^{1/2}\Psi^*\chi_3^{(A)})\Psi\right\}dx$$
$$= \xi\int_{-\infty}^{\infty} \{\Psi^*\sigma\kappa^{1/2}\chi_2^{(A)} + \kappa^{1/2}\chi_1^{(A)}\Psi\}dx.$$

Comparing this with Eq. (78) indeed shows that

$$[A, I'_1] = \xi[A, I'_0]. \tag{81}$$

Consider one more example:

$$[A, I'_2] \equiv \int_{-\infty}^{\infty} [A, \partial_x\Psi^*\partial_x\Psi - \sigma\kappa\Psi^{*2}\Psi^2]dx.$$

(a) $$\left[A, \int_{-\infty}^{\infty} \partial_x\Psi^*\partial_x\Psi\,dx\right]$$
$$= \int_{-\infty}^{\infty} \{\partial_x\Psi^*\partial_x[A,\Psi] + (\partial_x[A,\Psi^*])\partial_x\Psi\}dx$$
$$= \int_{-\infty}^{\infty} \{\partial_x\Psi^*\sigma\kappa^{1/2}\partial_x\chi_2^{(A)} + \kappa^{1/2}(\partial_x\chi_1^{(A)})\partial_x\Psi\}dx$$
$$= \int_{-\infty}^{\infty} \{\partial_x\Psi^*\sigma\kappa^{1/2}(i\xi\chi_2^{(A)} - \sigma\kappa^{1/2}\chi_3^{(A)}\Psi) + \kappa^{1/2}(-i\xi\chi_1^{(A)} - 2\sigma\kappa^{1/2}\Psi^*\chi_3^{(A)})\partial_x\Psi\}dx,$$

where Eqs. (34) have been used.

Integrating by parts we obtain

$$\left[A, \int_{-\infty}^{\infty} \partial_x\Psi^*\partial_x\Psi\,dx\right]$$
$$= \xi\int_{-\infty}^{\infty} \left\{\Psi^* \frac{\sigma\kappa^{1/2}}{i}\partial_x\chi_2^{(A)} - \frac{\kappa^{1/2}}{i}(\partial_x\chi_1^{(A)})\Psi\right\}dx + 2\sigma\kappa\int_{-\infty}^{\infty} \Psi^*\partial_x\chi_3\Psi.$$

Comparing with Eq. (80) we see that

$$\left[A, \int_{-\infty}^{\infty} \partial_x\Psi^*\partial_x\Psi\,dx\right] = \xi[A, I'_1] + 2\sigma\kappa\int_{-\infty}^{\infty} \Psi^*\partial_x\chi_3^{(A)}\Psi\,dx. \tag{82}$$

Also,

(b) $$\left[A, \int_{-\infty}^{\infty} \Psi^{*2}\Psi^2\,dx\right]$$
$$= \int_{-\infty}^{\infty} \{\Psi^{*2}[A,\Psi^2] + [A,\Psi^{*2}]\Psi^2\}dx$$
$$= 2\int_{-\infty}^{\infty} \Psi^*\{\sigma\kappa^{1/2}\Psi^*\chi_2^{(A)} + \kappa^{1/2}\chi_1^{(A)}\Psi\}\Psi\,dx$$
$$= 2\int_{-\infty}^{\infty} \Psi^*\partial_x\chi_3^{(A)}\Psi\,dx. \tag{83}$$

Combining Eqs. (82) and (83) we then see

$$[A, I'_2] = \xi[A, I'_1]. \tag{84}$$

Similarly, one shows that

$$[A, I'_3] = \xi[A, I'_2]. \tag{85}$$

Thus, we have shown that the classical polynomial constants $I'_0$, $I'_1$, $I'_2$ and $I'_3$ (when fields are replaced by operators and properly ordered) are operators which satisfy Eqs. (66). It may be noted that the ordering is that which one might have guessed.The order is normal, i.e., all creation operators are to the left of all annihilation operators. This does *not* persist to higher order as seen in the last term of $I'_4$ in Eqs. (2).

It can be shown that with the choice given

$$[A, I'_4] = \xi[A, I'_3]. \tag{86}$$

Note: The calculation is a little delicate. One must regard the field operators as tempered distributions.

The main point is that for the higher-order polynomials, non-normal ordered terms arise in the commutator

$[A, I'_n]$ from the commutator with normal ordered terms in $I'_n$ with sufficiently high numbers of derivations. However, it can be shown that by suitably arranging the order of terms with fewer derivatives in $I'_n$ these can be canceled. For example: In the commutator of $I'_4$ with $A$ the commutator of

$$-\sigma\kappa\int_{-\infty}^{\infty}\{2(\Psi^{*2})_x(\Psi^2)_x+\Psi^{*2}\Psi_x^2+(\Psi_x^*)^2\Psi^2\}dx,$$

with $A$ gives rise to non-normal ordered terms which are canceled by the commutator of $A$ with the non-normal terms in $I'_4$ which have no derivatives. Again suitable delicacy is needed.

## X. CONCLUSION

It has been shown that the polynomial constants of the classical nonlinear Schrödinger equation become quantum constants when the fields are promoted to operators and are appropriately ordered.

As in the classical case these constants have a particularly simple general form when expressed in terms of reflection coefficients. Indeed the classical expression strongly suggests the quantum form. Surprisingly, the quantum form is much simpler than the classical one.


## ACKNOWLEDGMENTS

This work was partially supported by the Department of Energy under Contract No. DE-AC02-81ER10807.


## APPENDIX A: PROOF OF PROPERTIES OF $C_m^{(n)}$

In Sec. VI various properties of the $C_m^{(n)}$ were listed. Here it is shown how these can be derived.

The general integral we encounter is

$$J_m^{(n)}=\int\cdots\int_{-\infty}^{\infty}(k_n)^m\frac{g_n(k_1,k_3,\ldots,k_n)dk_1\cdots dk_n}{\Pi_{j=1}^{n-1}(k_{j+1}-k_j)},\qquad(\mathrm{A1})$$

where $g_n$ is a symmetric function of its arguments which vanishes when any two are equal.

The most important result is the following theorem.

**Theorem**:

$$J_m^{(n)}=0,\quad n>m+1.\qquad(\mathrm{A2})$$

*Proof*: This proceeds by induction.

(1) The lemma of Sec. VI (Eq. 62) tells us that

$$J_0^{(n)}=0,\quad n>1.\qquad(\mathrm{A3})$$

(2) We show that for $n>m+1$, $J_m^{(n)}$ can be expressed in terms of $J_{m'}^{(n')}$ with $n'>m'+1$, $n>n'$, $m>m'$. Therefore, the integrals can be successively reduced to $J_0^{(n)}$ and are thus zero.

To show the reduction property, we write

$$K_1=J_m^{(n)}=\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n(k_n)^m g_n}{\Pi_{j=1}^{n-1}(k_{j+1}-k_j)},\qquad(\mathrm{A4})$$

$n$ times. Thus,

$$K_2\equiv K_1=\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n[(k_n)^m-(k_{n-1})^m]g_n}{\Pi(\ )}+\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n(k_{n-1})^m g_n}{\Pi(\ )},$$

$$K_3\equiv K_1=\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n[(k_n)^m-(k_{n-1})^m]g_n}{\Pi(\ )}+\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n(k_{n-1})^m-(k_{n-2})^n g_n}{\Pi(\ )}+\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n(k_{n-2})^m g_n}{\Pi(\ )},$$

$$\vdots$$

$$K_n=\int\int_{-\infty}^{\infty}\frac{dk_1\cdots dk_n\{[(k_n)^m-(k_{n-1})]^m+[(k_{n-1})^m-(k_{n-2})^m]\}}{\Pi(\ )}+\cdots+\frac{\{[(k_2)^m-(k_1)^m]+(k_1)^m\}g_n}{\Pi(\ )}.$$

Now average the $n$ expressions for $K_1$, i.e.,

$$J_m^{(n)}=\frac{1}{n}\sum_{i=1}^{n}K_i.\qquad(\mathrm{A5})$$

Four types of terms appear in Eq. (A5):

$$T_1=\int\int_{-\infty}^{\infty}dk_1\cdots dk_n\sum_{i=1}^{n}\frac{(k_i)^m g_n}{\Pi(\ )},\qquad(\mathrm{A6})$$

$$T_2=\int\int_{-\infty}^{\infty}dk_1\cdots dk_n\frac{[(k_n)^m-(k_{n-1})^m]g_n}{\Pi(\ )},\qquad(\mathrm{A7})$$

$$T_3=\int\int_{-\infty}^{\infty}dk_1\cdots dk_n\frac{[(k_2)^m-(k_1)^m]g_n}{\Pi(\ )},\qquad(\mathrm{A8})$$

and (if $n\geqslant 4$)

$$T_4=\int\int_{-\infty}^{\infty}dk_1\cdots dk_n\frac{[(k_l)^m-(k_{l-1})^m]g_n}{\Pi(\ )},$$

$$n-1\geqslant l\geqslant 3.\qquad(\mathrm{A9})$$

By the lemma $T_1=0$. To treat $T_2$, we eliminate the common factor $k_n-k_{n-1}$ in the numerator and denominator. Then

$$T_2=\sum_{r=0}^{m-1}\int\int_{-\infty}^{\infty}dk_1\cdots dk_n\frac{(k_n)^{m-1-r}(k_{n-1})^r g_n}{\Pi_{j=1}^{n-2}(k_{j+1}-k_j)}.$$

The generic term is

$$\int\int_{-\infty}^{\infty}dk_1\cdots dk_n\frac{(k_n)^{m-1-r}(k_{n-1})^r g_n}{\Pi_{j=1}^{n-2}(k_{j+1}-k_j)}=\int\int_{-\infty}^{\infty}dk_1\cdots dk_{n-1}\frac{(k_{n-1})^r g_{n-1}}{\Pi_{j=1}^{n-2}(k_{j+1}-k_j)},$$

where $g_{n-1}=\int_{-\infty}^{\infty}dk_n(k_n)^{m-1-r}g_n$. Thus, the expression is of the form $J_r^{(n-1)}$ where $r\leqslant m-1$, i.e., this is $J_{m'}^{(n')}$, where $n'=n-1<n$, $m'=r<m$. Here $T_3$ is clearly the same as $T_2$ with relabeling.

We are left with $T_4$. If we divide out the common factor of $k-k_{l-1}$, $T_4$ then becomes

$$T_4=\sum_{r=0}^{\infty}\int\cdots\int_{-\infty}^{\infty}dk_1\cdots dk_n\times\frac{(k_l)^{m-1-r}(k_{l-1})^r g_n}{\Pi_{j=1}^{l-2}(k_{j+1}-k_j)\Pi_{j=1}^{n-1}(k_{j+1}-k_j)},$$

with $n\geqslant 4$, $n-1\geqslant l\geqslant 3$.

This is of the general form

$$T_4=\sum_{r=0}^{m-1}J_r^{(l-1)}J_{m-1-r}^{(n-l+1)}.$$

We now use the induction argument to show that at least one of the two factors in each of the terms in the sum are zero.

Thus, consider $J_r^{(l-1)}$. This is of the form $J_{m'}^{(n')}$ where $n' < n$, $m' < m$. Therefore, it is zero unless $n' - m' \leqslant 1$, i.e., $l - 1 - r \leqslant 1$ or $l = r + 2 - \alpha$, where $\alpha \geqslant 0$. The second factor $J_{m-1-r}^{n-l+1}$, is of the form $J_{m''}^{n''}$, where $n'' < n$, $m'' < m$. But $n'' - m'' = n + 1 - (r + 2 - \alpha) - (m - 1 - r) = n - m + \alpha > 1 + \alpha$. Thus, $J_{m-1-r}^{n-l+1}$ is zero.

*The nonzero* $J_m^{(n)}$: As indicated in the main text, general results can be obtained. However, it is probably more informative to see how simply these can be computed for the small $m$ values. Thus we have the following.

For $m = 0$, the only nonzero integral is

$$J_0^{(1)} = \int_{-\infty}^{\infty} dk_1\, g_1(k_1).$$

For $m = 1$, there are two nonzero integrals

$$J_1^{(1)} = \int_{-\infty}^{\infty} k_1\, dk_1\, g_1(k_1),$$

and

$$J_1^{(2)} = \iint_{-\infty}^{\infty} \frac{k_2 g_2(k_1,k_2) dk_1\, dk_2}{k_2 - k_1}.$$

To evaluate $J_1^{(2)}$, we write this

$$J_1^{(2)} = \iint_{-\infty}^{\infty} \left(\frac{k_2 - k_1 + k_1}{k_2 - k_1}\right) g_2\, dk_1\, dk_2$$
$$= \iint_{-\infty}^{\infty} g_2\, dk_1\, dk_2 - J_1^{(2)},$$

i.e.,

$$J_1^{(2)} = \frac{1}{2} \iint_{-\infty}^{\infty} g_2\, dk_1\, dk_2.$$

For $m = 2$, there are now three nonzero integrals:

$$J_2^{(1)} = \int_{-\infty}^{\infty} k_1^2\, dk_1\, g_1,$$

$$J_2^{(2)} = \iint_{-\infty}^{\infty} \frac{k_2^2\, dk_1\, dk_2\, g_2}{k_2 - k_1},$$

$$J_2^{(3)} = \iint_{-\infty}^{\infty}\!\!\int \frac{k_3^2\, dk_1\, dk_2\, dk_3\, g_3}{(k_3 - k_2)(k_2 - k_1)}.$$

We write

$$J_2^{(2)} = \iint_{-\infty}^{\infty} \frac{(k_2^2 - k_1^2 + k_1^2) g_2\, dk_1\, dk_2}{k_2 - k_1}$$
$$= \iint_{-\infty}^{\infty} (k_2 + k_1) g_2\, dk_1\, dk_2 - J_2^{(2)},$$

$$\therefore J_2^{(2)} = \iint_{-\infty}^{\infty} k_2 g_2\, dk_1\, dk_2.$$

[Here we have interchanged some dummy labels and used the symmetry $g_2(k_1,k_2) = g_2(k_2,k_1)$.]

The evaluation of $J_2^{(3)}$ is most instructive since it illustrates the full range of tricks needed in the general case. We write $J_2^{(3)}$ in three equivalent forms:

$$J_2^{(3)} = \iint_{-\infty}^{\infty}\!\!\int \frac{k_3^2 g_3\, dk_1\, dk_2\, dk_3}{(k_3 - k_2)(k_2 - k_1)},$$

$$J_2^{(3)} = \iint_{-\infty}^{\infty}\!\!\int \frac{(k_3^2 - k_2^2 + k_2^2) g_3\, dk_1\, dk_2\, dk_3}{(k_3 - k_2)(k_2 - k_1)},$$

$$J_2^{(3)} = \iint_{-\infty}^{\infty}\!\!\int \frac{(k_3^2 - k_2^2 + k_2^2 - k_1^2 + k_1^2) g_3\, dk_1\, dk_2\, dk_3}{(k_3 - k_2)(k_2 - k_1)}.$$

Now take the average of these three expressions and note that

$$\iint_{-\infty}^{\infty}\!\!\int \frac{[(k_1^2 + k_2^2 + k_3^2)] g_3}{(k_3 - k_2)(k_2 - k_1)} = 0,$$

in virtue of our Lemma. Therefore,

$$J_2^{(3)} = \frac{1}{3} \iint_{-\infty}^{\infty}\!\!\int \frac{2(k_3^2 - k_2^2) + (k_2^2 - k_1^2) g_3\, dk_1\, dk_2\, dk_3}{(k_3 - k_2)(k_2 - k_1)}$$
$$= \frac{1}{3} \iint_{-\infty}^{\infty}\!\!\int \left(\frac{2(k_3 + k_2)}{k_2 - k_1} + \frac{(k_2 + k_1)}{(k_3 - k_2)}\right)$$
$$\times g_3\, dk_1\, dk_2\, dk_3.$$

The terms proportional to $k_3$ and $k_1$ are zero in virtue of the antisymmetry of the denominator.

Interchanging the labels 1 and 3 in the second term we see it is of the same form as the first (with a minus sign). Therefore,

$$J_2^{(3)} = \frac{1}{3} \iint_{-\infty}^{\infty}\!\!\int \frac{k_2}{k_2 - k_1} g_3\, dk_1\, dk_2\, dk_3$$
$$= \frac{1}{3} \iint_{-\infty}^{\infty}\!\!\int \frac{k_2 - k_1 + k_1}{k_2 - k_1} g_3\, dk_1\, dk_2\, dk_3$$
$$= \frac{1}{3} \iint_{-\infty}^{\infty}\!\!\int g_3\, dk_1\, dk_2\, dk_3 - J_2^{(3)},$$

$$\therefore J_2^{(3)} = \frac{1}{3!} \iint_{-\infty}^{\infty}\!\!\int g_3\, dk_1\, dk_2\, dk_3.$$

The calculation of the nonzero $J_m^{(n)}$ for higher $m$ proceeds in exactly the same fashion, e.g., for $m = 3$ we have the four nonzero integrals:

$$J_3^{(1)} = \int_{-\infty}^{\infty} (k_1)^3 g_1\, dk_1,$$

$$J_3^{(2)} = \iint_{-\infty}^{\infty} \left(k_2^2 + \frac{k_1 k_2}{2}\right) g_2\, dk_1\, dk_2,$$

$$J_3^{(3)} = \frac{1}{2} \iint_{-\infty}^{\infty}\!\!\int k_3 g_3\, dk_1\, dk_2\, dk_3,$$

$$J_3^{(4)} = \frac{1}{4!} \iiint_{-\infty}^{\infty}\!\!\int g_4\, dk_1\, dk_2\, dk_3\, dk_4.$$

## APPENDIX B: THE HIGHER-ORDER $I'_n$

We give a heuristic procedure to calculate these.

First introduce $\bar{I}_n$. (These are essentially the classical constants put in normal ordered form.) Define

$$Q^{(n)} = (\mathscr{L})^n Q^{(0)}, \tag{B1}$$

where

$$Q^{(0)} = \begin{pmatrix} \Psi^* \\ -\sigma\Psi \end{pmatrix}, \tag{B2}$$

and

$$\mathscr{L}\begin{pmatrix}\Phi_1\\ \Phi_2\end{pmatrix} = i\begin{pmatrix}\partial_x\Phi_1 + 2\sigma\kappa\Psi^*\partial_x^{-1}(\Phi_1\Psi) + 2\kappa\Psi^*\partial_x^{-1}(\Psi^*\Phi_2)\\ -\partial_x\Phi_2 - 2\sigma\kappa[\partial_x^{-1}(\Psi^*\Phi_2)]\Psi - 2\kappa[\partial_x^{-1}(\Phi_1\Psi)]\Psi\end{pmatrix}. \tag{B3}$$

Then we compute $\overline{I}_n$ from

$$Q_1^{(n)} = [\overline{I}_n, \Psi^*],$$
$$Q_2^{(n)} = \sigma[\overline{I}_n, \Psi]. \tag{B4}$$

The $I'_n$ are then to be obtained from the $\overline{I}_n$ by reordering. The rules are the following.

(1) At least one annihilation operator appears at the extreme right in all terms. (This guarantees $I'_n|0\rangle = 0$.) In particular, the term of highest order in the derivatives is uniquely determined by this requirement.

(2) Choose the ordering of the remaining terms such that $[\Lambda, I'_n] = \xi[\Lambda, I'_{n-1}]$. It is very tedious, but possible, to show this.

### APPENDIX C: COMMENTS ON THE COMMUTATION RELATIONS BETWEEN FIELD OPERATORS AND ZS FUNCTIONS

A "conventional"[3] derivation of these runs so: Consider, for example, $[\phi_1(x), \Psi^*(x)]$. Using Eqs. (12) and (3), we obtain

$$\begin{aligned}[\phi_1(x), \Psi^*(x)] &= \kappa^{1/2}\int_{-\infty}^{x} e^{+i\xi(x-x')/2}\phi_2(x') \times [\Psi(x'), \Psi^*(x)]dx \\ &= \kappa^{1/2}\int_{-\infty}^{x} e^{i\xi(x-x')/2}\phi_2(x')\delta(x'-x)dx' \\ &= (\kappa^{1/2}/2)\phi_2(x).\end{aligned} \tag{C1}$$

Here we have used: (i) the fact that $\phi_2(x')$ involves only $\Psi, \Psi^*$ for arguments less than $x$ and so $[\phi_2(x'), \Psi^*(x)] = 0$; and (ii) the convention that

$$\int_{-\infty}^{0} \delta(x)dx = \frac{1}{2}. \tag{C2}$$

While this seems very reasonable, a purist might ask for a further justification. There are two approaches.

The first, using the theory of tempered distributions is very technical, rigorous, and tedious.

The second is slightly heuristic but very convincing. Thus, we note that the only use of the commutation relations of $\Psi, \Psi^*$ with the ZS functions is to obtain Eqs. (35)–(37).

Note that the commutation relations of Eq. (3) are obtained from the classical Poisson brackets via the correspondence principle. Hence we might expect the same for other commutators.

Consider the classical form of Eq. (12) and take the Poisson bracket with $\Psi^*(x)$. We obtain

$$\begin{aligned}\{\phi_1(x), \Psi^*(x)\} &= \kappa^{1/2}\int_{-\infty}^{x} e^{i\xi(x-x')/2}\phi_2(x')\{\Psi(x'), \Psi^*(x)\}dx \\ &= -i\kappa^{1/2}\int_{-\infty}^{x} e^{i\xi(x-x')/2}\phi_2(x')\delta(x'-x)dx'.\end{aligned} \tag{C3}$$

Now using the convention of Eq. (C2) this gives

$$\{\phi_1(x), \Psi^*(x)\} = (-i\kappa^{1/2}/2)\phi_2(x). \tag{C4}$$

Similarly, using the same convention, we obtain the classical analog of our commutators with the other ZS functions. Combining these, we then obtain the classical analog of Eqs. (35)–(37). However, in Ref. 6 these relations are obtained completely rigorously with no use of Eq. (C2).


[1] V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[2] E. Sklyanin and L. D. Faddeev, Dokl. Acad. Nauk. SSSR **243**, 1470 (1978) [Sov. Phys. Dokl. **23**, 902 (1978)].
[3] H. B. Thacker and David Wilkinson, Phys. Rev. D **19**, 3660 (1979); D. B. Creamer, H. B. Thacker, and D. Wilkinson, Phys. Rev. D **21**, 1573 (1980).
[4] B. Davies, J. Phys. A: Math.Gen. **14**, 2631 (1981).
[5] D. J. Kaup, J. Math. Phys. **16**, 2036 (1975).
[6] M. D. Arthur and K. M. Case, J. Math. Phys. **23**, 1771 (1982).

# Conformally invariant wave equations for massless particles

James A. McLennan
*Department of Physics, Lehigh University, Bethlehem, Pennsylvania 18015*

(Received 1 February 1983; accepted for publication 9 March 1984)

The invariance of wave equations for massless particles under conformal transformations of space-time is briefly summarized. Particular attention is given to a recent paper by Bracken and Jessup in which it is claimed that results obtained by the author are in error. Their paper contains several misleading statements based on a misreading of the author's paper, and in addition an argument of theirs, intended to show error, is itself invalid. Their claims of error on the author's part are therefore unfounded.

PACS numbers: 11.10.Qr, 11.30.Ly

## I. INTRODUCTION

It was demonstrated long ago that the scalar wave equation,[1] Maxwell's equations,[2] and the Dirac equation with zero mass[3] are invariant under the conformal transformations of space-time. The conformal invariance in these cases was clearly related to the absence of mass, and the question arose whether other equations for massless particles were also conformally invariant. This question was investigated by the author,[4] the equations considered being those constructed by Gårding[5] for massless particles of arbitrary spin.

Recently Bracken and Jessup[6] have claimed that CI is in error in several respects. However they have not accurately represented the content of CI, and an argument which they employ is lacking in precision and therefore unable to demonstrate their point. Consequently they have not in fact found errors in CI. The purpose of this note is to rectify the inaccuracies in the paper of Bracken and Jessup, and to summarize in elementary terms the situation regarding the conformal invariance of the equations in question.

## II. REPRESENTATIONS OF THE CONFORMAL GROUP

The method used in CI was to extend the standard spinor representations of the Lorentz group to the conformal group, and then to apply the resulting transformations on a case-by-case basis to the equations at hand. The method of construction of representations of the conformal group can be described as follows. Let $C$ denote the conformal group (on space-time) and $L$ the homogeneous Lorentz group; their parts connected to the identity will be denoted by $C_0$ and $L_0$. Under a conformal transformation $x\rightarrow x'$ [where $x=(x_0, x_1, x_2, x_3)$ and we use a metric with signature $+\ -\ -\ -$] the interval is changed by

$$dx'^2 = \mu^2\, dx^2. \tag{1}$$

The scale factor $\mu$ is related to the Jacobian $J$ of the transformation by $|J| = |d^4x'/d^4x| = \mu^4$. If we write

$$dx_i' = \mu Q_i^{\ j}\, dx_j, \tag{2}$$

then $Q$ is a Lorentz matrix. Thus the transformations induced by $C$ on differential forms at $x$ differ by only a scale factor from Lorentz transformations and the same is then true of the transformations induced on the tangent space at $x$. (This was noted in the early work of Bateman and Cunningham.[2]) Hence we can immediately apply the theory of representations of the Lorentz group[7] in the following way.

The standard expression of a point in Minkowski space as a matrix on $C^2$ (complex two-space) is

$$X = \begin{pmatrix} x_0 + x_3 & x_1 - i\,x_2 \\ x_1 + i\,x_2 & x_0 - x_3 \end{pmatrix}. \tag{3}$$

From the fact that $Q$ is a Lorentz matrix it follows that, for every element in $C_0$, there exists a matrix $q$ on $C^2$ (which is determined uniquely except for sign) such that

$$dX' = q^*\, dX\, q. \tag{4}$$

Taking determinants we find $|\det q|^2 = \mu^2$. For the full group $C$ it is necessary to consider transformations from $C^2$ to $\overline{C^2}$ (where the bar denotes the complex conjugate), just as for the Lorentz group. By applying two transformations in succession, it is readily confirmed that $\mu$ and $q$ satisfy the composition laws

$$\mu(g', gx)\,\mu(g, x) = \mu(g'g,x),$$
$$q(g,x)q(g', gx) = q(g'g, x), \tag{5}$$

where $g$, $g'$ are elements of $C_0$. The matrix $r$ which occurs in the transformation for tangent vectors satisfies

$$r(g, x)q(g, x) = 1, \quad r(g', gx)r(g, x) = r(g'g, x). \tag{6}$$

Thus $\mu$ and $r$ are "multipliers" in the sense of Bargmann.[8]

Consider the transformations $u\rightarrow u' = T(g)\,u$ on functions $u(x)$ with values in $C^2$,

$$(T(g)u)(gx) = \mu^{-t}(g, x)r(g, x)u(x). \tag{7}$$

These satisfy $T(g)\,T(g') = T(gg')$ and so provide a representation of $C_0$, or rather a family of representations which is labeled by the parameter $t$. (Actually this is a two-valued representation, or a representation of the covering group, as are also the other representations discussed below for half-integral spin.) The generalization to any of the finite irreducible representations $D(m,n)$ of $L_0$ is immediate. [Here we label the representations by integers, the relation to Cartan's notation being $D(m,n) = \mathscr{D}_{m/2,n/2}$. The $D(m,n)$ can be expressed as a set of transformations on a symmetric spinor with $m$ undotted and $n$ dotted indices.] If $u$ carries the representation $D(m,n)$ of $L_0$, then the corresponding representation of $C_0$ is provided by the transformations which we again denote by $T(g)$,

$$(T(g)u)^{\alpha\ldots\dot\beta\ldots}(gx) = \mu^{-t}(g, x)r^{\alpha}{}_{\lambda}(g, x)\ldots r^{\dot\beta}{}_{\dot\rho}(g, x)\ldots u^{\lambda\ldots\dot\rho\ldots}(x). \tag{8}$$

Here for clarity the spinor indices have been written out and

 0022-2488/84/072315-06$02.50  

$\bar{r}^{\dot\beta}{}_{\dot\rho}$ is the complex conjugate of $r^{\beta}{}_{\rho}$. These representations will be denoted by $C(m,n,t)$.

Except for a factor $\mu$, $r$ has algebraically the same form as its restriction to $L_0$, and so the standard Lorentz scalars are conformal scalars modulo a power of $\mu$. Thus if $u{\cdot}v$ denotes the Lorentz-invariant bilinear form associated with the representation $D(m, n)$, then $u'{\cdot}v' = \mu^{-w}\, u{\cdot}v$, where $w = m + n + 2t$. By choosing $t$, one can make $u{\cdot}v\, d^4x$ a conformal invariant, which leads to the construction of conformally invariant action principles.[9]

The scale factor $\mu$ corresponding to any element of $C$ can be evaluated by calculating the Jacobian. To evaluate $r$, let $U(Q)$ denote the standard mapping of $L_0$ to SL(2). Then since $U$ is unimodular while $r$ (as normalized above) satisfies $\det r = \mu^{-1}$, we have $r(g, x) = \mu^{-1/2}\, U(Q)$, where $Q$ is the Lorentz matrix associated with the pair $g,x$ according to

$$Q_{ij} = \mu^{-1} R_{ij}, \quad R_i{}^j = \frac{\partial x_i'}{\partial x_j}. \tag{9}$$

In CI, $r$ was worked out explicitly for the accelerations (or special conformal transformations) and the inversion. The results are as follows. The acceleration is

$$x_i' = \mu[x_i - a_i x^2], \quad \mu = [1 - 2a{\cdot}x + a^2x^2]^{-1}, \tag{10}$$

and the associated matrix $r$ is given by

$$r^{\alpha}{}_{\lambda} = \delta^{\alpha}_{\lambda} + a^{\alpha}{}_{\dot\beta} X_{\lambda}{}^{\dot\beta}. \tag{11}$$

Here $X_{\alpha\dot\beta}$ denotes the elements of the matrix (3) (e.g., $X_{1\dot2} = x_1 - i\,x_2$) and $a_{\alpha\dot\beta}$ is related to $a_i$ in the same way. [The result (11) is given in CI for infinitesimal $a$, but has the same form for finite $a$.] The inversion is

$$x_i' = -k\,x_i/x^2, \quad \mu = |k/x^2|. \tag{12}$$

This requires a mapping from $C^2$ to $\overline{C}^2$, and the matrix $r$ is given by

$$r_{\alpha\dot\beta} = k^{-1/2} X_{\alpha\dot\beta}. \tag{13}$$

For the dilatations $x_i' = \mu x_i$ ($\mu$ = const), we have $r = \mu^{-1/2}$. For integral spin, Eq. (8) can be written as a transformation on a tensor,

$$u'_{ij\ldots} = \mu^{-t} R_i{}^k R_j{}^m \cdots u_{km\ldots}. \tag{14}$$

## III. GÅRDING'S EQUATIONS AND THEIR BEHAVIOR UNDER CONFORMAL TRANSFORMATIONS

Once the above representations were constructed, they were used in CI to discuss the conformal invariance of Gårding's equations. These can be described as follows. Let $u$, as above, carry the irreducible representation $D(m,n)$ of $L_0$, and put $p_i = \partial/\partial x_i$. From the quantities $p_i\, u$ one can form (by multiplying the representations for $p_i$ and $u$, then reducing) four objects which transform irreducibly according to the representations $D(m+1, n+1)$, $D(m+1, n-1)$, $D(m-1, n+1)$, and $D(m-1, n-1)$. (There are obvious exceptions when $m$ or $n$ vanish.) On setting these four quantities equal to zero, Gårding's irreducible equations are obtained. There are slight modifications in the procedure when equations invariant under $L$ are desired. (The situation here is similar to that in the theory of the neutrino; the equations constructed for $L_0$ are also invariant under $L$ if transformations between $u$ and its complex conjugate are allowed.)

The equations which transform as $D(m+1, n+1)$ are not physically interesting since they do not have plane-wave solutions. This is easily seen when $m = n = 0$ as then the equations for the scalar $u$ become $p_i u = 0$; the proof for arbitrary $m,n$ is not difficult. Hence when $m$ vanishes there is only one interesting equation, which transforms as $D(1, n-1)$; a similar situation occurs if $n = 0$.

Each of Gårding's irreducible equations applies to a massless particle in the sense that they have solutions which are also solutions to the wave equation $\Box u = 0$. Gårding then constructed "minimum sets" which have the property that every solution to the equations in a minimum set is also a solution to the wave equation. It was found that if $m$ or $n$ vanishes then any single irreducible equation is a minimum set; otherwise any two equations form a minimum set, with an exception to be noted below. (Clearly a minimum set which is composed of two irreducible equations does not itself transform according to an irreducible representation.) If the equations without plane-wave solutions are eliminated, the remaining minimum sets fall into only four classes, which we shall proceed to list here.

In case I either $m$ or $n$ vanishes and a minimum set consists of a set of equations which transforms irreducibly according to the representation $D(1, n-1)$, or $D(m-1, 1)$. If neither $m$ nor $n$ vanishes, the following additional cases occur. In case II the minimum sets consist of two irreducible equations transforming according to $D(m+1, n-1)$ and $D(m-1, n-1)$, or according to $D(m-1, n+1)$ and $D(m-1, n-1)$.[10] In case III the minimum sets contain equations transforming according to $D(m+1, n-1)$ and $D(m-1, n+1)$. However there is an exception when $m = n$, as then these two equations do not form a minimum set. These three cases have been described in terms appropriate to $L_0$. The structure of the minimum sets applicable to $L$ is similar, but if $m = n$ there is a special case which will be called case IV: a minimum set consists of two equations, one of which transforms irreducibly (under $L$) as $D(m+1, m-1) \oplus D(m-1, m+1)$, and the other as $D(m-1, m-1)$. In this case it can be shown that $u$ is a symmetric tensor with zero trace and the minimum set reduces to

$$u_{ij\ldots} = p_i p_j \cdots \psi, \quad \Box\psi = 0. \tag{15}$$

The scalar wave equation does not occur in Gårding's minimum sets except through this reduction.

The results obtained in CI regarding the conformal invariance of Gårding's equations and minimum sets can be summarized as follows. First, all of the irreducible equations are conformally invariant. If the irreducible equation is Lorentz invariant when $u$ transforms according to the representation $D(m, n)$, then it is conformally invariant with the transformations corresponding to the representation $C(m,n,t)$ for a particular value of $t$; the value of $t$ is different for each of the irreducible equations which can be written for a given $u$. (These values of $t$ are listed in a table in CI. Note that $m$, $n$ have slightly different meanings there than here.) For the minimum sets, conformal invariance was not considered in those cases which do not have plane-wave solutions. It was also not considered for case IV since these minimum sets can be reduced to the scalar wave equation and so were not felt to be of interest in themselves. In case I, the conformal invariance is an immediate consequence of that for the irreducible equations. In this case $m$ or $n$ vanishes and Table I of CI yields the value $t = 1$, so the representation which yields invariance is $C(m,0,1)$ or $C(0,n,1)$. In case II the transformation law which results in invariance is different for the two irreducible equations (they have different values for $t$), and so neither transformation law leaves the minimum set invariant. However it is invariant with a third transformation law which corresponds to the representation $1_{m+1} \otimes C(0,n,1)$ [or $C(m,0,1) \otimes 1_{n+1}$], where $1_p$ denotes the $p$-dimensional identity representation. In case III, the minimum sets were shown not to be conformally invariant. (Here for brevity no distinction has been made between invariance under $C_0$ and under $C$. Details on this point are available in CI.)

The reducibility of the representations in case II merits further comment. Consider a minimum set composed of equations transforming irreducibly under $L_0$ according to $D(m-1, n+1)$ and $D(m-1, n-1)$, respectively. As shown by Gårding, the combination of equations can be written as

$$p^{\dot\beta_0}{}_{\alpha_1} u^{\alpha_1 \cdots \alpha_m \dot\beta_1 \cdots \dot\beta_n} = 0. \tag{16}$$

This equation does not transform irreducibly under $L_0$ since the left-hand side is not symmetric in the dotted indices; the irreducible equations can be retrieved by forming symmetric parts in the dotted indices. Transformations which leave this equation invariant under $C_0$ are

$$(T(g)u)^{\alpha_1 \cdots \alpha_m \dot\beta_1 \cdots \dot\beta_n}(gx) = \mu^{-1}(g,x) r^{\alpha_1}{}_{\lambda_1}(g,x) \cdots r^{\alpha_m}{}_{\lambda_m}(g,x) u^{\lambda_1 \cdots \lambda_m \dot\beta_1 \cdots \dot\beta_n}(x). \tag{17}$$

Since the indices $\dot\beta_1 \ldots \dot\beta_n$ are not summed over, these transformations correspond to the representation $C(m,0,1) \otimes 1_{n+1}$. If $n > 0$ this representation can immediately be reduced to a direct sum of $n+1$ representations $C(m,0,1)$. If $n = 0$, the indices $\dot\beta_1 \cdots \dot\beta_n$ are absent and we have case I. If $n > 0$, then for fixed values of $\dot\beta_1 \cdots \dot\beta_n$, Eq. (16) is identical to the case I equation, that is, Eq. (16) is a collection of $n+1$ independent case I equations. The conformal invariance in case II, with the transformation law (17), is therefore an immediate consequence of that in case I. Lorentz invariance of the irreducible equations individually requires $u$ to transform (except for a constant factor) according to $D(m,n)$, but as is usually the case the invariance properties of the set of equations are different from those of its individual equations or subsets. Thus Eq. (16) is Lorentz invariant if $u$ transforms according to the representation $D(m,n)$ but it is also invariant with the reducible representation $D(m,0) \otimes 1_{n+1}$, (see Ref. 11) indeed even with $D(m,0) \otimes \mathrm{GL}(n+1)$, as is manifest from the form of the equation. Similar remarks apply when the roles of $m$ and $n$ are interchanged. In such cases the transformation law under $L_0$ is not determined by the free-field equations but might become definite if interactions were included.

Weinberg[12] has shown that any free massless field can be expressed as a linear combination of certain fundamental fields and their derivatives, where the fundamental fields transform under $L_0$ according to $D(m,0)$ or $D(0,n)$. Gårding's minimum sets reduce to equations for the fundamental fields as follows. In case I the fields already transform as fundamental fields. In case II the minimum sets admit the reducible representations described in the previous paragraph so $u$ is a collection of fundamental fields. Case IV has the reduction (15) to a scalar field. In case III there is a similar reduction (which was not realized when CI was written): If for example $n > m$, then it can be shown from the case III equations that $u$ is an $m$th derivative of a quantity which transforms according to $D(0,n-m)$ and satisfies the case I equations. Thus one way or another all of Gårding's minimum sets lead to case I or to the scalar wave equation, in accordance with Weinberg's theorem.

## IV. DISCUSSION

Now we turn to the paper by Bracken and Jessup. First of all, this paper contains several statements regarding CI which are not in accord with the actual content of that reference. In the abstract they state: "...it is confirmed that not all Poincaré-invariant sets of massless Type-Ia field equations are conformal invariant, contrary to some often-quoted results of McLennan, which are shown to be invalid." Contrary to this statement, in CI not all minimum sets were considered (even beyond those without plane-wave solutions), and of those treated some were shown not to be conformally invariant. That some minimum sets are not conformally invariant is stated explicitly in the Introduction, in the section devoted to the minimum sets, and again in the Summary of CI. (Only minimum sets qualify as massless field equations in the sense of the above quotation, as those irreducible equations which are not minimum sets do not imply the wave equation.) Then in their Introduction, Bracken and Jessup say: "McLennan claimed to prove the invariance of each of Gårding's 'irreducible sets'... ." Gårding did not have "irreducible sets," and the term is not used in CI. If they said instead "irreducible equations" the statement would be true; each of the irreducible equations is conformally invariant. If they meant to say "minimum sets," then as already noted the statement would be false.

Particular emphasis is placed by Bracken and Jessup on alleged error with regard to case IV, but as stated above conformal invariance in this case was not considered. They quote the following: (such sets of equations) "...are equivalent to the scalar or pseudo-scalar wave equations," implying incorrectly that the statement of equivalence constituted a claim of conformal invariance. In CI the reduction (15) is given and the sentence immediately following in full quotation is "Thus the minimum sets made up from (3.11) and (3.12) are equivalent [in the sense of (3.15)] to the scalar or pseudo-scalar wave equation." There Eqs. (3.11) and (3.12) constitute the case IV minimum set, and Eq. (3.15) is the same as (15) above. Contrary to the implication of the incomplete quotation by Bracken and Jessup, it is not stated that the case IV minimum sets are conformally invariant. Instead they were removed from further consideration once the equivalence (15) was established, and Sec. III closes with an unambiguous statement to this effect. Bracken and Jessup attribute error in this case to misunderstanding of a point that $p_i$ is not "conformal covariant." What they mean is not clear; the derivative operator always behaves as a contravariant vector under transformations of the coordinates. However it is to be emphasized that the results of CI were based on specific, detailed calculations rather than on a casual application of some unsupported rule of covariance.

Bracken and Jessup are correct that the case IV minimum sets are not conformally invariant, but this fact seems not to have much significance. The natural description of a scalar particle is by the scalar wave equation rather than the more complicated case IV equations.

Apparently Bracken and Jessup confirm the results of CI in regard to case I, as they state "We detected no errors in this part of McLennan's work." However in regard to case II, they assert "This contradicts a claim made by McLennan, but it is easy to find an error in his analysis." As noted above, the conformal invariance in case II is an immediate consequence of that in case I, so there is no additional analysis to be in error. Indeed Bracken and Jessup do not, as they claim, locate any error in analysis, but instead construct an independent argument which leads to what they believe to be a contradiction. They note that the generators of infinitesimal rotations can be obtained by commuting generators of translations and accelerations. The acceleration transformations given in CI, having the form (17), act trivially on the space labeled by the dotted indices, and so will the rotations obtained this way. Contradiction is then claimed because the infinitesimal rotations "will affect all dotted and undotted indices." This argument makes no contact with the equations in question, so the claim is not merely that the equations are not invariant, but that the transformation itself is somehow in contradiction. Indeed they say "McLennan's proposed transformation law is not consistent if $p \neq 0$." However it is nothing more than a nonsingular linear transformation on the components of $u$, which violates no mathematical requirements whatsoever. There is no mathematical necessity for a rotation or a Lorentz transformation to affect all Greek indices, dotted or undotted. The indices occur only as a matter of notation and have no mathematical content in themselves, while the transformation law is determined mathematically by a requirement of invariance.

A referee has maintained that Bracken and Jessup use a different definition of invariance, according to which the behavior of $u$ is regarded as "predetermined, being defined by the spinor indices present," and furthermore that "these equations are not conformal-invariant in the usual sense of the term, when applied to a relativistic wave equation for a field whose Lorentz transformation properties have already been prescribed." The definition of invariance given by Bracken and Jessup contains no clear statement to this effect, and in any case the use of a different definition cannot provide grounds for the claim that the analysis of CI is in error. However there is evidently need for some discussion of the meaning of invariance.

The traditional definition of invariance can be stated as follows.[13] Let $D_x$ be a differential operator. If for an invertible transformation $x \rightarrow x' = gx$ on the coordinates there exists a transformation $u \rightarrow u' = su$ such that $D_x u = 0$ is equivalent to $D_{x'} u' = 0$, then the equation $D_x u = 0$ is said to be invariant under $g$. It is readily confirmed that the set of all such $g$ forms a group $G$, and one speaks of invariance under $G$. The set of all $s$ also forms a group $S$, and there is a homomorphism from $S$ onto $G$ whose kernel consists of gauge transformations. The group $S$ (or more properly, the pair $G$, $S$) is called the symmetry group of the equation.

Thus "invariant" has its everyday meaning of "unchanged"; the equation is invariant if its form after the transformation is the same as before. The phrase "invariant under $g$" means that there exists a corresponding transformation $s$ on $u$ such that the pair $g$, $s$ leaves the equation invariant; the existence of other transformations on $u$ which do not meet the requirement does not disprove invariance under $g$.

In CI the discussion of the invariance under $C_0$ of the case II minimum sets is contained entirely in one sentence which reads "For the transformation (4.6), the minimum set (3.19) is invariant if the wave function transforms like,..." where (4.6) is the infinitesimal acceleration and the equation then displayed is the corresponding form of (17) above. This plain-English statement has the unambiguous meaning that when the transformations are carried out, one recovers the original equation, unchanged in form.

The notion that invariance of an equation entails a "prescribed" or "predetermined" transformation is a confusion of concepts. One can stipulate that $u$ transforms in a certain way for a variety of reasons, such as to illustrate a notation, or to study the transformations themselves, or to construct equations which are invariant with a given representation. However an equation determines its own symmetry group and once the equation is established one has no more freedom. The term would lose all useful meaning if an equation could be invariant or not depending on the prescription or on such fashions as the notation.

Restrictions on the transformations to be allowed can destroy expected group-theoretical properties. The straightforward and general demonstration that $G$ and $S$ are groups depends on the supposition that *all* transformations which leave the equation invariant are included (otherwise the set of transformations might not be closed). Equation (16) has an obvious group of gauge transformations which consist of linear transformations with constant coefficients on the space labeled by the indices $\dot{\beta}_1 \dots \dot{\beta}_n$. The representation $D(m,0) \otimes 1_{n+1}$ can be obtained by combining transformations from $D(m,n)$ with gauge transformations. If it is desired to retain the group property and the gauge transformations are admitted, then Lorentz invariance according to $D(m,n)$ requires the acceptance of $D(m,0) \otimes 1_{n+1}$. The prohibition, for whatever reason, of $D(m,0) \otimes 1_{n+1}$ yields a subset of $S$ which is not a group. In particular, this subset is not homomorphic to $C_0$, whereas the argument by which Bracken and Jessup claim to find a contradiction assumes the existence of a homomorphism.

Electromagnetic theory provides a familiar example with features similar to the case at issue. The theory using Lorentz gauge is Lorentz invariant if the potentials transform as a four-vector, but in Coulomb gauge it is necessary to augment the four-vector transformation law with a gauge transformation. One cannot claim that the potentials transform as a four-vector merely because they are labeled by a four-valued index.

One can infer correctly from the invariance under accelerations as stated in CI that Eq. (16) must be Lorentz invariant with the representation $D(m, 0) \otimes 1_{n+1}$ but this is evident from the form of the equation. Furthermore it is necessary in order for these minimum sets to conform to Weinberg's theorem.

In the Appendix below detailed calculations are provided to confirm that the case II minimum sets are conformally invariant. Detailed calculations were not given in CI for any of the minimum sets since the computations are similar to those which were provided for one of the irreducible equations.

In summary, Bracken and Jessup have found no errors in CI, and they have misrepresented the content of CI. It is hoped that their misunderstanding will not be propagated in the literature.

## APPENDIX: PROOF OF CONFORMAL INVARIANCE FOR CASE I AND CASE II MINIMUM SETS

In this Appendix it will be shown that Eq. (16) is invariant under $C_0$ for arbitrary values of $m$ and $n$.

We recall some of the rules of spinor analysis.[14] Spinor indices are raised and lowered according to $a^1 = a_2$, $a^2 = -a_1$. This can be expressed by

$$a_\alpha = a^\lambda \epsilon_{\lambda\alpha}, \tag{A1}$$

where $\epsilon_{\alpha\lambda}$ is the antisymmetric symbol with $\epsilon_{12} = -\epsilon_{21} = 1$. We have $a^\alpha{}_\alpha = -a_\alpha{}^\alpha$ so if $a$ is symmetric then $a^\alpha{}_\alpha = 0$. In addition we have the identities

$$\begin{aligned} &a_{\alpha\lambda} a^{\beta\lambda} = a_{\lambda\alpha} a^{\lambda\beta} = \delta^\beta_\alpha \det a, \\ &b_{\alpha\dot\beta} b^{\lambda\dot\beta} = \delta^\lambda_\alpha b^2, \\ &b^{\alpha\dot\beta} b_{\lambda\dot\rho} - b^\alpha{}_{\dot\rho} b^{\dot\beta}{}_\lambda = b^2 \delta^\alpha_\lambda \delta^{\dot\beta}_{\dot\rho}. \end{aligned} \tag{A2}$$

Note that $b^{\alpha\dot\beta} b_{\alpha\dot\beta} = 2b^2$. The above identities are easily proven by writing them out fully, for example,

$$a_{1\lambda} a^{1\lambda} = a_{11} a^{11} + a_{12} a^{12} = a_{11} a_{22} - a_{12} a_{21} = \det a. \tag{A3}$$

We first consider the behavior of the derivative operator $p^{\alpha\dot\beta} = \partial / \partial X_{\alpha\dot\beta}$ under conformal transformations. We have

$$p'^{\alpha\dot\beta} = \frac{\partial X_{\lambda\dot\rho}}{\partial X'_{\alpha\dot\beta}} p^{\lambda\dot\rho}. \tag{A4}$$

For the inversion,

$$X_{\alpha\dot\beta} = -k X'_{\alpha\dot\beta} / x'^2, \quad x'^2 = k^2 / x^2. \tag{A5}$$

We then obtain

$$\frac{\partial X_{\lambda\dot\rho}}{\partial X'_{\alpha\dot\beta}} = -k^{-1} [x^2 \delta^\alpha_\lambda \delta^{\dot\beta}_{\dot\rho} - X_{\lambda\dot\rho} X^{\alpha\dot\beta}]. \tag{A6}$$

Using Eq. (19) we get

$$\frac{\partial X_{\lambda\dot\rho}}{\partial X'_{\alpha\dot\beta}} = r^\alpha{}_{\dot\rho} r^{\dot\beta}{}_\lambda, \tag{A7}$$

where $r$ is given by Eq. (13). Hence

$$p'^{\alpha\dot\beta} = r^\alpha{}_{\dot\rho} r^{\dot\beta}{}_\lambda p^{\lambda\dot\rho}, \tag{A8}$$

which confirms the remarks near the beginning of the paper that the derivative operator transforms according to a product of transformations on $C^2$ (or, in this case, between $C^2$ and $\overline{C^2}$). The accelerations can be obtained from an inversion followed by a translation followed by the inversion again:

$$x'_i = -k[(-kx_i/x^2) - t_i][(-kx/x^2) - t]^{-2}. \tag{A9}$$

This reduces immediately to Eq. (10) with $a_i = -t_i/k$. Applying the same sequence of transformations to $r$, we obtain the result (11), with the transformation law for the derivative operator having the form

$$p'^{\alpha\dot\beta} = r^\alpha{}_\lambda r^{\dot\beta}{}_{\dot\rho} p^{\lambda\dot\rho}. \tag{A10}$$

We now proceed with the calculation to show that Eq. (16) is invariant under the accelerations with $u$ transforming according to Eq. (17). Starting out with Eq. (16) in the primed coordinates, we obtain

$$\begin{aligned} &p'^{\dot\beta_0}{}_{\alpha_1} u'^{\alpha_1 \cdots \alpha_m \dot\beta_1 \cdots \dot\beta_n} \\ &= r^{\dot\beta_0}{}_{\dot\rho} r_{\alpha_1\lambda} \mu^{-1} r^{\alpha_1}{}_{\lambda_1} \cdots r^{\alpha_m}{}_{\lambda_m} p^{\lambda\dot\rho} u^{\lambda_1 \cdots \lambda_m \dot\beta_1 \cdots \dot\beta_n} \\ &\quad + r^{\dot\beta_0}{}_{\dot\rho} r_{\alpha_1\lambda} [p^{\lambda\dot\rho} \mu^{-1} r^{\alpha_1}{}_{\lambda_1} \cdots r^{\alpha_m}{}_{\lambda_m}] u^{\lambda_1 \cdots \lambda_m \dot\beta_1 \cdots \dot\beta_n}. \end{aligned} \tag{A11}$$

Now

$$p^{\lambda\dot\rho} r^{\alpha_1}{}_{\lambda_1} = -a^{\alpha_1\dot\rho} \delta^\lambda_{\lambda_1}, \tag{A12}$$

and in addition

$$p^{\lambda\dot\rho} \mu^{-1} = -a^{\lambda\dot\rho} + a^2 X^{\lambda\dot\rho}. \tag{A13}$$

The first of the identities (A2) yields

$$r_{\alpha_1\lambda} r^{\alpha_1}{}_{\lambda_1} = \epsilon_{\lambda_1\lambda} \det r = \mu^{-1} \epsilon_{\lambda_1\lambda}, \tag{A14}$$

so we get

$$r_{\alpha_1\lambda} r^{\alpha_1}{}_{\lambda_1} p^{\lambda\dot\rho} r^{\alpha_2}{}_{\lambda_2} = -\mu^{-1} a^{\alpha_2\dot\rho} \epsilon_{\lambda_2\lambda_1}. \tag{A15}$$

This and similar terms give no contribution in (A11) since $u$ is symmetric in the superscripts $\lambda_1 \ldots \lambda_m$. In addition we have

$$\begin{aligned} &r_{\alpha_1\lambda} p^{\lambda\dot\rho} \mu^{-1} r^{\alpha_1}{}_{\lambda_1} \\ &= r_{\alpha_1\lambda} [\mu^{-1}(-a^{\alpha_1\dot\rho} \delta^\lambda_{\lambda_1}) + r^{\alpha_1}{}_{\lambda_1}(-a^{\lambda\dot\rho} + a^2 X^{\lambda\dot\rho})] \\ &= \mu^{-1} [-a^{\alpha_1\dot\rho} r_{\alpha_1\lambda_1} - a_{\lambda_1}{}^{\dot\rho} + a^2 X_{\lambda_1}{}^{\dot\rho}]. \end{aligned} \tag{A16}$$

When Eq. (11) for $r$ is used, this expression reduces to zero. Thus the last term in Eq. (A11) vanishes, and we are left with

$$\begin{aligned} &p'^{\dot\beta_0}{}_{\alpha_1} u'^{\alpha_1 \cdots \alpha_m \dot\beta_1 \cdots \dot\beta_n} \\ &= \mu^{-1} r^{\dot\beta_0}{}_{\dot\rho} r_{\alpha_1\lambda} r^{\alpha_1}{}_{\lambda_1} \cdots r^{\alpha_m}{}_{\lambda_m} p^{\lambda\dot\rho} u^{\lambda_1 \cdots \lambda_m \dot\beta_1 \cdots \dot\beta_n}. \end{aligned} \tag{A17}$$

Use of Eq. (A14) yields

$$\begin{aligned} &p'^{\dot\beta_0}{}_{\alpha_1} u'^{\alpha_1 \cdots \alpha_m \dot\beta_1 \cdots \dot\beta_n} \\ &= \mu^{-2} r^{\dot\beta_0}{}_{\dot\rho} r^{\alpha_2}{}_{\lambda_2} \cdots r^{\alpha_m}{}_{\lambda_m} [p^{\dot\rho}{}_{\lambda_1} u^{\lambda_1 \cdots \lambda_m \dot\beta_1 \cdots \dot\beta_n}] \\ &= 0. \end{aligned} \tag{A18}$$

This completes the proof of invariance under accelerations.

The above demonstration applies to Eq. (16), which is the same (except for notation) as Eq. (3.17) in CI. A similar argument applies to Eq. (3.19) in CI. Invariance under $C_0$ follows from the (evident) invariance under dilatations, translations, and Lorentz transformations. Equation (3.19) is also invariant under $C$, as follows from its invariance under $L$. The same analysis also applies to the case I equations, it only being necessary to suppress the indices $\dot\beta_1 \cdots \dot\beta_n$.

It may be useful to provide an alternative calculation using the more familiar four-vector notation. The simplest case which illustrates the point involved is when $m = n = 1$ and Eq. (16) becomes

$$p^{\dot\beta}{}_{\alpha} u^{\alpha\dot\rho} = 0. \tag{A19}$$

This is a set of two independent two-component neutrino equations, but $u$ can couple to other fields as a four-vector. Let $u_{\alpha\dot\beta}$ correspond to $u_i$ in the manner of Eq. (3). By a straightforward calculation, Eq. (A19) is converted into

$$p^i u_i = 0, \quad p_j u_i - p_i u_j - i\epsilon_{ijkm} p^m u^k = 0, \tag{A20}$$

where $\epsilon_{ijkm}$ is the completely antisymmetric symbol with $\epsilon_{0123} = 1$. These two equations are (in a different notation) the same as the two irreducible equations which comprise the minimum set. That $u_i$ satisfies the wave equation follows immediately from the two equations; neither equation by itself implies the wave equation. Equations (A20) are not invariant under $L$ (with linear transformations). The corresponding minimum set which is invariant under $L$ has $p_i\, u_j - p_j\, u_i = 0$ in place of the second of Eqs. (A20) and the reduction (15) then follows.

We will demonstrate the invariance of Eqs. (A20) under accelerations. The manipulations are lengthier than when spinor notation is used, and so for simplicity only infinitesimal accelerations will be considered. Then the transformation on the coordinates is

$$x_i' = (1 + 2a\cdot x)x_i - a_i x^2, \tag{A21}$$

from which we get

$$\frac{\partial x_i}{\partial x'^j} = (1 - 2a\cdot x)g_{ij} + 2(a_i x_j - a_j x_i). \tag{A22}$$

The transformation law for $u_i$ under accelerations can be obtained by transcribing Eq. (17) (with $m = n = 1$) into four-vector notation, and for infinitesimal $a$ the result is

$$u_i' = (1 - 3a\cdot x)u_i + x_i a\cdot u - a_i x\cdot u - i\epsilon_{ijkm} a^j x^k u^m. \tag{A23}$$

[This transformation does not have the form (14) since $u$ is being transformed as two independent spin-$\frac{1}{2}$ fields.] We then obtain

$$p_j' u_i' = (1 - 5\, a\cdot x)p_j u_i + (a^m x_i - a_i x^m)p_j u_m + 2(a^m x_j - a_j x^m)p_m u_i - i\epsilon_{imnp} a^m x^n p_j u^p + \Delta_{ij}, \tag{A24}$$

where

$$\Delta_{ij} = g_{ij}\, a\cdot u - 3a_j u_i - a_i u_j + i\epsilon_{ijmn} a^m u^n. \tag{A25}$$

It is readily confirmed that

$$\Delta_i{}^i = 0, \Delta_{ij} - \Delta_{ji} - i\epsilon_{ijmn}\Delta^{mn} = 0, \tag{A26}$$

so $\Delta_{ij}$ drops out when the left-hand sides of Eqs. (A20) are formed. For brevity we introduce

$$U = p^i u_i, \quad V_{ij} = p_j u_i - p_i u_j - i\epsilon_{ijmn} p^n u^m, \tag{A27}$$

and write Eq. (37) as $U = 0$, $V_{ij} = 0$. Equation (A24) yields immediately

$$U' = (1 - 5\, a\cdot x)U - a^i x^j V_{ij}. \tag{A28}$$

Next we calculate $V'_{ij}$, using the identities

$$\epsilon_{imnp} A^p{}_j - \epsilon_{jmnp} A^p{}_i = \epsilon_{ijmn} A^p{}_p + \epsilon_{ijnp} A^p{}_m - \epsilon_{ijmp} A^p{}_n,$$
$$\epsilon_{ijmn}\epsilon^{pqrn} = -\sum(\pm)\delta^p_i \delta^q_j \delta^r_m. \tag{A29}$$

Here the sum extends over the six permutations of $i, j, m$, the sign being positive or negative depending on whether the permutation is even or odd. After a tedious but straightforward manipulation it is found that

$$V'_{ij} = (1 - 5\, a\cdot x)V_{ij} + (a_i x_j - a_j x_i - i\epsilon_{ijmn} a^m x^n)U + i\,\epsilon_{ijmn}(a^m x_k - a_k x^m)V^{kn}. \tag{A30}$$

Equations (A20) then show that $U' = 0$, $V'_{ij} = 0$, which completes the proof of invariance.


[1]H. Bateman, Proc. London Math. Soc. **7**, 70 (1909).

[2]E. Cunningham, Proc. London Math. Soc. **8**, 77 (1910); and H. Bateman, Proc. London Math. Soc. **8**, 223 (1910).

[3]P. A. M. Dirac, Ann. Math. **37**, 429 (1936); and W. Pauli, Helv. Phys. Acta **13**, 204 (1940). Dirac expressed the conformal group by linear transformations on a six-dimensional space, while Pauli considered conformal transformations on the metric. A direct approach with explicit expressions for the transformations on the Dirac spinor was given by S. Bludman, Phys. Rev. **107**, 1163 (1957).

[4]J. A. McLennan and P. Havas, Phys. Rev. **87**, 898 (1952); and J. A. McLennan, Nuovo Cimento **3**, 1360 (1956). The latter reference will be referred to herein as CI.

[5]L. Gårding, Proc. Cambridge Philos. Soc. **41**, 49 (1945).

[6]A. J. Bracken and B. Jessup, J. Math. Phys. **23**, 1925 (1982).

[7]See, for example, E. Cartan, *Lecons sur la Theorie des Spineurs* (Hermann, Paris, 1938); B. L. van der Waerden, *Group Theory and Quantum Mechanics* (Springer, New York, 1974); and I. M. Gelfand, R. A. Minlos, and Z. J. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963).

[8]V. Bargmann, Ann. Math. **48**, 568 (1947). Multipliers are also discussed in R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

[9]J. A. McLennan, Nuovo Cimento **5**, 640 (1957).

[10]Equations for particles with zero mass and arbitrary spin were obtained by P. A. M. Dirac, Proc. R. Soc. London Ser. A **155**, 447 (1936). Dirac eliminated the reducibility of the case II minimum sets so that the resulting equations correspond to case I. However his equations for nonzero mass yield both cases I and II if the mass is set equal to zero. The case I and case II equations were also obtained by M. Fierz, Helv. Phys. Acta **12**, 3 (1939) and **13**, 45 (1940). The case I equations include the massless Dirac equation and Maxwell's equations as special cases. They were used by V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. **34**, 211 (1946), and are referred to as the Bargmann–Wigner equations by Bracken and Jessup.

[11]That an equation can be invariant in more than one way, and that the equation need not be determined by the representation, have been discussed by E. P. Wigner in *Theoretical Physics* (IAEA, Vienna, 1963).

[12]S. Weinberg, Phys. Rev. **134**, B882 (1964); **138**, B988 (1965).

[13]See, for example, P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U. P., London, 1947), 3rd ed. p. 257; A. S. Wightman in *Invariant Wave Equations*, edited by G. Velo and A. S. Wightman (Springer, Berlin, 1978); and in the particular context of conformal invariance, F. Bayen, in *Differential Geometry and Relativity*, edited by M. Cahen and M. Flato (Reidel, Dordrecht, 1976).

[14]See for example O. Laporte and G. E. Uhlenbeck, Phys. Rev. **37**, 1380 (1931); W. L. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953); and the books cited in Ref. 7.

# On integrability properties of SU (2) Yang–Mills fields. I. Infinitesimal part

H. Urbantke
*Institut für Theoretische Physik, Universität Wien, Vienna, Austria*

(Received 22 July 1983; accepted for publication 2 December 1983)

We study the distribution of 2-plane elements on which infinitesimal parallel displacement of isovectors yields identity. This yields to an algebraic and differential classification and, in the generic case, to a quasimetric naturally associated with the field.

PACS numbers: 11.15. – q, 11.30.Jw

## I. INTRODUCTION

The mathematical solution of the instanton problem[1] is based on the fundamental remark of Ward[2] that for self-dual Yang–Mills fields on flat space-time the parallel transport of "iso" vectors is path independent within any totally null anti-self-dual 2-plane of flat space. In a different guise, this was also noted by Yang,[3] who used these planes as coordinate planes in order to simplify the field equations and to choose a convenient gauge. In the case of Coleman's plane-fronted waves,[4] and generalizations thereof,[5] there is path independence within the wave hypersurfaces, permitting a choice of gauge that gets rid of the nonlinearities. This suggested[6] an investigation of possible integrability properties of YM fields in order to find simplifying gauges, and we arrived[7] at a classification scheme which is coarser than the ones published (see, e.g., Ref. 8), but, as it stresses a different aspect, it might be nevertheless quite useful. In fact, while the published schemes aim at separating orbits of field tensors under the action of SU (2)$\times$Lorentz group, our approach uses neither the space-time metric ($\leftrightarrow$Lorentz group) nor the particular structure of SU (2), but only its dimensionality.

The problem is to find submanifolds in space-time, of dimension $\geqslant 2$, on which parallel transport is integrable. In general, this problem has no solution, and the aim is to sort out cases where there is one. There are three steps in the problem: the infinitesimal part, the local part, and the global part. We shall have to say nothing about the third part.The infinitesimal part is to find, at each space-time point, those tangent 2-plane elements on which the YM curvature form vanishes. The local problem then is to try and fit plane elements at different points together to form (local) 2-surfaces.

In this paper we describe the solution of the infinitesimal problem. After some general geometric remarks in Sec. 2, we give in Sec. 3 the classification of YM fields that arises in the process of solving the infinitesimal problem. In the concluding section 4, we give an indication of the work on the local problem whose details will appear elsewhere.

## 2. GENERAL GEOMETRIC REMARKS

When an "iso" vector $\psi(x)$ at the space-time point $x$ is parallelly propagated around a closed infinitesimal loop, its change is

$$\Delta\psi(x) = p^{\mu\nu}F^{a}_{\mu\nu}(x)G_a\psi(x), \tag{1}$$

where $G_a$ are the generators of the representation to which $\psi$ belongs, $F^{a}_{\mu\nu}$ ($a = 1,2,3$) are the YM field strengths, and $p^{\mu\nu} = u^{[\mu}v^{\nu]}$ ([...] means antisymmetrization) is the bivector associated with the 2-plane element spanned by the tangent vectors $u$, $v$ that define the loop as an infinitesimal parallelogram. Changing $u$, $v$ within that plane while keeping them linearly independent changes $p^{\mu\nu}$ by a nonzero scalar factor. The infinitesimal problem, for each $x$, is thus to find all solutions $p^{\mu\nu} = -p^{\nu\mu}$ of the system

$$F^{a}_{\mu\nu}p^{\mu\nu} = 0 \quad (a = 1,2,3), \tag{2}$$

together with the condition that $p^{\mu\nu}$ is *simple*, i.e., can be written as $p^{\mu\nu} = u^{[\mu}v^{\nu]}$:

$$p^{\mu\nu}\tilde{p}_{\mu\nu} = 0. \tag{3}$$

[Here we have defined the dual

$$\tilde{p}_{\mu\nu} := \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}p^{\alpha\beta}, \tag{4}$$

and since we are not using any *a priori* space-time metric, $\epsilon_{\mu\nu\alpha\beta}$ is just the permutation symbol, so that $\tilde{p}_{\mu\nu}$ is only a "relative" covariant tensor, which, however, does not matter, (3) being homogeneous. Note in the following that most of the equality signs are important only up to a nonzero factor, so that we shall drop the specification "relative tensor of weight..." in most cases and just say "tensor."] If (3) holds, the factors $u$,$v$ are determined, up to linear combinations of each other, as solutions of

$$\tilde{p}_{\mu\nu}w^{\nu} = 0. \tag{5}$$

The algebraic problem posed by Eqs. (2) and (3) is a standard problem in line geometry (see, e.g., Ref. 9. Appendix, for an exposition in physicists' notation). We shall describe its solution in the next section, distinguishing several cases. We shall work in the complexified tangent space, although the interpretation of complex $p$ in the sense of (1) would require a complexified space-time. Here we make the following consideration on it. Putting $p^{\mu\nu} = u^{[\mu}v^{\nu]}$, we want to solve

$$F^{a}_{\mu\nu}u^{\mu}v^{\nu} = 0. \tag{6}$$

for $u$, $v$. Fix $u$ and put

$$F^{a}_{\nu}(u) := u^{\mu}F^{a}_{\mu\nu}; \tag{7}$$

then

$$F^{a}_{\nu}v^{\nu} = 0 \tag{8}$$

are three linear equations for $v$ which always possess a solution. One solution is $v = u$ by the antisymmetry of $F^{a}_{\mu\nu}$, and this solution is unique up to proportionality iff $\operatorname{rank}(F^{a}_{\nu}(u)) = 3$, which would make $p^{\mu\nu} = 0$ trivial. Hence we must require rank $(F^{a}_{\nu}(u)) < 3$. Equating to zero all four $3\times 3$ determinants in $F^{a}_{\nu}(u)$ amounts to writing

$$(1/3!)\epsilon_{abc}\epsilon^{\mu\alpha\beta\gamma}F^{a}_{\alpha}F^{b}_{\beta}F^{c}_{\gamma} = 0. \tag{9}$$

  

(Again, $\epsilon^{\mu\alpha\beta\gamma}$ is the permutation symbol that is going to be used to form duals of covariant tensors.) The lhs is a (gauge scalar) vector depending cubically on $u$, and must therefore have the form $g(u,u)u^\mu$ where $g(u,u)$ is a scalar quadratic in $u$, i.e., $g(u,u) = g_{\mu\nu}\, u^\mu, u^\nu$. Assuming $g_{\mu\nu} = g_{\nu\mu}$, we determine $g_{\mu\nu}$ by comparing coefficients:

$$g_{\mu\nu} := (-1/3!)\epsilon_{acb} F^a_{\mu\alpha} \tilde{F}^{c\alpha\beta} F^b_{\beta\nu}. \tag{10}$$

Thus from (9) we get $g(u,u) = 0$, and the same must hold for all linear combinations of $u$ and $v$, implying

$$g_{\mu\nu}u^\mu u^\nu = 0, \quad g_{\mu\nu}v^\mu v^\nu = 0, \quad g_{\mu\nu}u^\mu v^\nu = 0. \tag{11}$$

The solutions $p^{\mu\nu}$ therefore describe 2-plane elements which are totally null in the sense of the quasimetric $g_{\mu\nu}$.

Since (3) also implies that $\tilde{p}_{\mu\nu}$ is simple, i.e., determining, up to linear combinations of each other, a pair of covectors $a_\mu$, $b_\nu$, solutions of

$$p^{\mu\nu} c_\nu = 0, \tag{5'}$$

so that $\tilde{p}_{\mu\nu} = a_{[\mu} b_{\nu]}$, and since $F^a_{\mu\nu} p^{\mu\nu} \equiv \tilde{F}^{a\mu\nu} \tilde{p}_{\mu\nu}$ due to $\epsilon_{\mu\nu\alpha\beta}\epsilon^{\mu\nu\rho\sigma} \equiv 4\delta^\rho_{[\alpha}\delta^\sigma_{\beta]}$, there is a dual calculation starting from

$$\tilde{F}^{a\mu\nu}\tilde{p}_{\mu\nu} = 0 \tag{2'}$$

that leads to

$$\tilde{g}^{\mu\nu}a_\mu a_\nu = 0, \quad \tilde{g}^{\mu\nu}b_\mu b_\nu = 0, \quad \tilde{g}^{\mu\nu}a_\mu b_\nu = 0, \tag{11'}$$

where

$$\tilde{g}^{\mu\nu} := (-1/3!)\epsilon_{abc} \tilde{F}^{a\mu\alpha} F^c_{\alpha\beta} F^{b\beta\nu}. \tag{10'}$$

The same tensors are also encountered in the following consideration. There are always nontrivial (possibly complex) coefficients $\lambda_a$ such that the linear combination $F_{\mu\nu} = \lambda_a F^a_{\mu\nu}$ becomes a *simple* bivector, $\tilde{F}^{\mu\nu}F_{\mu\nu} = 0$: choose $\lambda_a$ to satisfy

$$M^{ab}\lambda_a\lambda_b = 0, \quad M^{ab} := \tilde{F}^{a\mu\nu}F^b_{\mu\nu}. \tag{12}$$

Then also $\tilde{F}^{\mu\nu}$ is simple; $\tilde{F}^{\mu\nu} = s^{[\mu}t^{\nu]}$, where $s$, $t$ are independent solutions for $r^\nu$ of

$$F_{\mu\nu}r^\nu = \lambda_a F^a_{\mu\nu} r^\nu = 0. \tag{13}$$

Regarding this as a system of linear homogeneous equations for the nontrivial $\lambda_a$, we again find that the matrix $F^a_\mu(r) = F^a_{\mu\nu}r^\nu$ must have rank $< 3$, implying $g(r,r) = 0$ for all vectors $r$ of the plane spanned by $s$, $t$. In a dual manner, $F_{\mu\nu} = c_{[\mu}d_{\nu]}$, where $\tilde{g}(c,c) = = 0, \tilde{g}(d,d) = 0, \tilde{g}(c,d) = 0$.

In the nonsingular case, det $g_{\mu\nu} \neq 0$, we can add the following remarks. Since $\tilde{p}$ is characterized uniquely up to a factor by $\tilde{p}_{\mu\nu}w^\nu = 0$ for all vectors $w$ from the plane $p$, and since $g_{\mu\nu}u^\mu w^\nu = 0 = g_{\mu\nu}v^\mu w^\nu$, we may take $a_\mu \propto g_{\mu\nu}u^\nu, b_\mu \propto g_{\mu\nu}v^\nu$ to span $\tilde{p}_{\mu\nu} = a_{[\mu}b_{\nu]}$. This leads to three conclusions:

(1) From $g(a,a) = 0$ whenever $a_\mu = g_{\mu\nu}u^\nu$ and $g(u,u) = 0$ it follows that

$$\tilde{g}^{\mu\nu}g_{\mu\alpha}g_{\nu\beta} \propto g_{\alpha\beta}, \quad \text{i.e.,} \quad \tilde{g} \propto g^{-1}, \tag{14}$$

using an obvious matrix notation. (If a quadratic form vanishes on the set of zeros of another, nondegenerate, quadratic form, it must be a multiple thereof.) A direct inversion of (10) would have been tedious.

(2) We get $\tilde{p}_{\mu\nu} \propto g_{\mu\alpha}g_{\nu\beta}p^{\alpha\beta}$, or

$$p^{\mu\nu} \propto |\det g_{..}|^{1/2} g^{-1\mu\alpha} g^{-1\nu\beta} \tilde{p}_{\alpha\beta} =: {}^*p^{\mu\nu}, \tag{15}$$

i.e., $p$ is self-dual in the sense of the "metric" $g$. Note that this statement contains a convention, since from ${}^{**}p \equiv$ sgn (det $g_{..}$)$p$ it follows that ${}^*p = \pm\sqrt{\text{sgn}(\det g_{..})}\, p$. Also note that the $\sim$ duality carries contravariant into covariant tensors and vice versa, whereas * duality needs an additional metric (up to a nonzero factor; and, strictly speaking, we have not provided more than that) and carries contravariants into contravariants, allowing for the concept of self-duality.

(3) An identical reasoning for the $F_{\mu\nu} = \lambda_a F^a_{\mu\nu}$ introduced above leads to ${}^*\tilde{F} \propto \tilde{F}$, but this time the opposite sign than before has to appear, i.e., the $\tilde{F}$ are anti-self-dual in the sense of $g$. This is because by construction of the $p$ we have

$$0 = \epsilon_{\mu\nu\alpha\beta}\tilde{F}^{\mu\nu}p^{\alpha\beta} = \epsilon_{\mu\nu\alpha\beta}s^\mu t^\nu u^\alpha v^\beta, \tag{16}$$

implying a linear dependence between $s$, $t$, $u$, $v$, which means that all planes $p$ have vectors in common with each of the planes $\tilde{F}$, and vice versa, whereas two self-dual planes with a nonzero vector in common would have to coincide, as is easy to verify.

Thus without having used a space-time metric from the start, we have constructed a "quasimetric" (10), up to a nonzero factor, with respect to which the given YM field is (anti-) self-dual in the generic case. There are degenerate cases, however, which we shall describe in the classification of the next section.

### 3. INFINITESIMAL CLASSIFICATION

*Case 1.* $F^a_{\mu\nu}$ ($a = 1,2,3$) are linearly independent. Form the $3\times 3$ matrix $M^{ab}$, Eq. (12), and determine its rank $m$.

*Case 1.1*: $m = 3$. This is the generic case, for which det $g_{\mu\nu} \neq 0$. For real $F^a_{\mu\nu}$, the signature of $g_{\mu\nu}$ is $++++$ (elliptic) or $++--$ (ultrahyperbolic). For elliptic signature, the $p$ and the $\tilde{F}^a\lambda_a$ above are complex; for ultrahyperbolic signature, both are real. In the complex case, there is a one- (complex) parameter count of solutions for $p$.

*Case 1.2*: $m = 2$. $M^{ab}$ can be written as

$$M^{ab} = A^a B^b + A^b B^a, \tag{17}$$

where $A^a$, $B^a$ are linearly independent and unique up to factors. Then, for arbitrary $L^b$, there are fixed $a_\nu$, $b_\nu$, $u^\nu$, $v^\nu$, such that

$$\begin{aligned} \epsilon_{abc}A^aL^bF^c_{\mu\nu} &= c_{[\mu}(L)a_{\nu]}, \\ \epsilon_{abc}B^aL^bF^c_{\mu\nu} &= d_{[\mu}(L)b_{\nu]}, \\ \epsilon_{abc}A^aL^b\tilde{F}^{c\mu\nu} &= w^{[\mu}(L)u^{\nu]}, \\ \epsilon_{abc}B^aL^b\tilde{F}^{c\mu\nu} &= y^{[\mu}(L)v^{\nu]}. \end{aligned} \tag{18}$$

$a$, $b$ as well as $u$, $v$ are independent, unique up to proportionality, and satisfy

$$a_\mu u^\mu = a_\mu v^\mu = b_\mu u^\mu = b_\mu v^\mu = 0. \tag{19}$$

Equations (10) and (10') become

$$g_{\mu\nu} \propto a_\mu b_\nu + a_\nu b_\mu, \quad \tilde{g}^{\mu\nu} \propto u^\mu v^\nu + u^\nu v^\mu. \tag{20}$$

Thus the rank of the matrices $g$, $\tilde{g}$ has dropped down to 2, and they satisfy $g_{\mu\lambda}\tilde{g}^{\lambda\nu} = 0$. The 2-planes we are interested in are given in this case by

$$p^{\mu\nu} = u^{[\mu}w^{\nu]} \quad \text{and} \quad p^{\mu\nu} = v^{[\mu}w^{\nu]}, \tag{21a}$$

where $w$ is an arbitrary vector satisfying

$$b_\mu w^\mu = 0, \quad a_\mu w^\mu = 0, \tag{21b}$$

or, dually, by

$$\tilde{p}_{\mu\nu} = a_{[\mu} c_{\nu]} \quad \text{and} \quad \tilde{p}_{\mu\nu} = b_{[\mu} c_{\nu]}, \tag{$\widetilde{21a}$}$$

where $c$ is an arbitrary covector satisfying

$$v^\mu c_\mu = 0. \tag{$\widetilde{21b}$}$$

In the real case, $a$ and $b$ may be real or complex conjugate; also, $u$, $v$ will be real or complex conjugate, respectively. In the complex case, we have two one-parameter families of solutions for $p$.

*Case 1.3*: $m = 1$. This case is obtained from 1.2 by putting $b_\mu \propto a_\mu, v^\mu \propto u^\mu$. We get a one-parameter family of solutions.

*Case 1.4*: $m = 0$. Here either

$$\tilde{g}^{\mu\nu} = 0, \quad g_{\mu\nu} = a_\mu a_\nu \neq 0; \tag{22}$$

the $F^a_{\mu\nu}$ can be written

$$F^b_{\mu\nu} = c^b_{[\mu} a_{\nu]}, \quad a_\mu, \vec{c}_\mu \text{ indep.} \tag{23}$$

$p$ is determined by

$$\tilde{p}_{\mu\nu} = a_{[\mu} c_{\nu]}, \quad c_\nu \neq 0 \text{ arbitrary } (\not\propto a_\nu); \tag{24}$$

or there is the dual case

$$g_{\mu\nu} = 0, \quad \tilde{g}^{\mu\nu} = u^\mu u^\nu \neq 0, \tag{$\widetilde{22}$}$$

$$\tilde{F}^{a\mu\nu} = w^{a[\mu} u^{\nu]}, \quad u^\nu, \vec{w}^\nu \text{indep.}, \tag{$\widetilde{23}$}$$

$$p^{\mu\nu} = u^{[\mu} w^{\nu]}, \quad w^\nu \neq 0 \text{ arbitrary } (\not\propto u^\nu). \tag{$\widetilde{24}$}$$

The 2-plane elements $p$ are thus either contained in the hyperplane element whose vectors are annihilated by $a_\mu$, or they all pass through a fixed single tangent vector $u^\mu$. Hence we get a two-parameter family of plane elements in each case.

*Case 2.* The $F^a_{\mu\nu}$ span only a two-dimensional subspace of the tensor space and can be more symmetrically written as

$$F^a_{\mu\nu} = \rho^a_A \phi^A_{\mu\nu}, \tag{25}$$

where capital indices range and sum over $\{1,2\}$, and where $\phi^A_{\mu\nu}$ are independent. We form the $2\times 2$ matrix

$$\mu^{AB} := \phi^A_{\mu\nu} \tilde{\phi}^{B\mu\nu} = \mu^{BA} \tag{26}$$

and determine its rank $\mu$ which equals the rank of $M^{ab} = \mu^{AB} \rho^a_A \rho^b_B$, while $g_{\mu\nu} \equiv 0$, $\tilde{g}^{\mu\nu} \equiv 0$ here.

*Case 2.1*: $\mu = 2$. Here we may pick $\phi^1_{\mu\nu}$, $\phi^2_{\mu\nu}$ such as to satisfy

$$\phi^1_{\mu\nu} \tilde{\phi}^{1\mu\nu} = 0 = \phi^2_{\mu\nu} \tilde{\phi}^{2\mu\nu} \tag{27}$$

by going to suitable linear combinations, i. e., we may pick them to be simple:

$$\tilde{\phi}^{1\mu\nu} = u_1^{[\mu} v_1^{\nu]}, \quad \tilde{\phi}^{2\mu\nu} = u_2^{[\mu} v_2^{\nu]}, \quad u_1, u_2, v_1, v_2 \text{ indep.} \tag{28}$$

The solutions for $p^{\mu\nu}$ are then

$$p^{\mu\nu} = (\alpha_1 u_1 + \beta_1 v_1)^{[\mu} (\alpha_2 u_2 + \beta_2 v_2)^{\nu]}, \tag{29}$$

where the coefficients are arbitrary (not all $= 0$). This gives a two-parameter family of plane elements. In the real case, $\phi^1_{\mu\nu}$ and $\phi^2_{\mu\nu}$ are real or complex conjugates. In the latter case, our formula (29) with $u_2 = \bar{u}_1, v_2 = \bar{v}_1, \alpha_2 = \bar{\alpha}_1, \beta_2 = \bar{\beta}_1$ gives $p^{\mu\nu}$ purely imaginary, but a factor $i$ is irrelevant for the reality of the plane element.

*Case 2.2*: $\mu = 1$. There is only one combination of the $\phi^A_{\mu\nu}$ that can be made to satisfy $\phi_{\mu\nu} \tilde{\phi}^{\mu\nu} = 0$. Take this as $\phi^1_{\mu\nu}$, write

$$\tilde{\phi}^{1\mu\nu} = u^{[\mu} v^{\nu]}, \quad \text{where } \phi^1_{\mu\nu} u^\nu = \phi^1_{\mu\nu} v^\nu = 0, \tag{30}$$

and pick some independent $\phi^2_{\mu\nu}$; then $\mu = 1$ implies $\phi^1_{\mu\nu} \tilde{\phi}^{2\mu\nu} = 0$. The solutions $p^{\mu\nu}$ can be written

$$p^{\mu\nu} = w^{[\mu} (u^{\nu]} v^{[\beta} - v^{\nu]} u^{\beta]}) w^\alpha \phi^2_{\alpha\beta}, \tag{31}$$

where $w$ is arbitrary but $u$, $v$, $w$ independent. This gives a two-parameter family of plane elements which is real in the real case.

*Case 2.3*: $\mu = 0$. Here the $\phi^A_{\mu\nu}$ can be written

$$\phi^A_{\mu\nu} = a_{[\mu} b^A_{\nu]}, \quad a_\mu, b^1_\mu, b^2_\mu \text{ indep.} \tag{32}$$

Putting

$$u^\mu = \epsilon^{\mu\nu\alpha\beta} a_\nu b^1_\alpha b^2_\beta, \tag{33}$$

the plane elements are given by

$$p^{\mu\nu} = \epsilon^{\mu\nu\alpha\beta} a_\alpha b_\beta \quad \text{and} \quad p^{\mu\nu} = u^{[\mu} v^{\nu]}, \tag{34}$$

where $b$, $v$ are arbitrary (indep. of $a$, $u$, resp.) This gives us two-parameter families of plane elements (real in the real case): those passing through the vector $u$ and those being contained in the hyperplane element through $u$ whose vectors are annihilated by an on scalar multiplication.

*Case 3.* The $F^a_{\mu\nu}$ span only a one-dimensional space, $F^a_{\mu\nu} = f^a F_{\mu\nu}$.

*Case3.1$F_{\mu\nu}\tilde{F}^{\mu\nu} \neq 0$.* The conditions $F_{\mu\nu} p^{\mu\nu} = 0$, $p_{\mu\nu} \tilde{p}^{\mu\nu} = 0$ define a three-parameter family of plane elements, real in the real case.

*Case 3.2*; $F_{\mu\nu} \tilde{F}^{\mu\nu} = 0$. Here we can find independent $a_\mu, b_\mu$ such that

$$F_{\mu\nu} = a_{[\mu} b_{\nu]}, \tag{35}$$

and then $p^{\mu\nu}$ is given by

$$p^{\mu\nu} = \epsilon^{\mu\nu\alpha\beta} (\gamma a_\alpha + \delta b_\alpha) c_\beta, \tag{36}$$

where $\gamma$, $\delta$ are arbitrary scalars, $c_\beta$ an arbitrary covector. This is again a three-parameter family of plane elements, real in the real case. It consists of the 2-plane elements which intersect the 2-plane element given by the simple $\tilde{F}^{\mu\nu}$ along any vector and not just at the origin.

*Case 4.* This is the trivial case $F^a_{\mu\nu} = 0$ $(a = 1,2,3)$ where $p^{\mu\nu}$ is arbitrary.

## 4. CONCLUDING REMARKS

We have now determined, at each point, all 2-plane elements on which the YM curvature vanishes, and have distinguished 10 nontrivial qualitatively different cases. The local problem is now to try and select, for each $x$, one $p^{\mu\nu}(x)$ out of the family obtained, in such a way that the corresponding 2-plane elements are tangent to 2-surfaces. Using the Frobenius integrability condition and a convenient parametrization of the family, one can work out further conditions which yield a differential classification of each of the above cases. This will be done elsewhere.

It would be tempting to speculate on the further physical significance of the (conformal) metric (10) which we have distilled out of the SU (2) gauge field strengths in the generic case, with respect to which the gauge field is (anti-) self-dual, and which has to be sharply distinguished from any physical metric. It will, in general, be conformally curved, its Weyl tensor entering the integrability problem mentioned above. Apart from this and its properties associated with its very origin, we have not found any further significance so far.

[1]M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977); M. F. Atiyah, V. G. Drinfeld, N. J. Hitchin, and Yu. G. Manin, Phys. Lett. A **65**, 185 (1978).
[2]R. S. Ward, Phys. Lett. A **61**, 81 (1977).
[3]C. N. Yang, Phys. Rev. Lett. **38**, 1377 (1977); E. Corrigan, D. B. Fairlie, P. Goddard, and R. G. Yates, Commun. Math. Phys. **58**, 223 (1978).
[4]S. Coleman, Phys. Lett. B **70**, 59 (1977).
[5]H. K. Urbantke, J. Math. Phys. **20**, 1851 (1979).
[6]H. K. Urbantke, Acta Phys. Austriaca, Suppl. **XIX**, 875 (1978).
[7]H. K. Urbantke, in *Conference on Differential Geometric Methods in Theoretical Physics*, edited by G. Denardo and H. D. Doebner (World Scientific, Singapore, 1983).
[8]L.-L. Wang and C. N. Yang, Phys. Rev. **17**, 2687 (1978); J. Anandan and R. Roskies, Phys. Rev. **18**, 1152 (1978).
[9]V. Hlavaty, *Geometry of Einstein's Unified Field Theory* (Noordhoff, Groningen, 1958).

# Some series of infinitely many symmetry generators in symmetric space chiral models

Bo-yu Hou[a)]
*Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794*

(Received 4 October 1983; accepted for publication 6 January 1984)

This paper shows that, starting from any conserved current generated by some given infinitesimal symmetry generator, one may use finite dual transformations to induce infinitely many infinitesimal symmetry generators. Thus, besides starting from ordinary isotopic and space-time translation, this paper also discovers the infinitesimal generators for Bäcklund transformation, for dual symmetry itself and other general cases, and then uses them to generate infinitely many local or nonlocal currents, respectively.

PACS numbers: 11.30. – j, 11.10.Lm

## I. INTRODUCTION

In this decade, there has been much interest and considerable progress in the nonlinear physical systems and in the nonlinear mathematics. The two dimensional chiral model[1–6] is one of the nonlinear problems under extensive investigation. The chiral model behaves quite similarly with the four-dimensional Yang–Mills field, e.g., both have topologically nontrivial solutions such as instantons and merons[7] both possess some kind of BT (Bäcklund transformation) with similar structures.[1,5–10] It is reasonable to expect that the thorough investigation of this simpler model will be helpful for deeper understanding of the more complicated Yang–Mills field. As a complete integrable system solvable by inverse scattering method, the chiral model possesses a lot of rather interesting and mutually connected properties such as multisoliton solutions, BT, and sets of infinitely many conserved currents, either local or nonlocal.[11–19] What is the hidden symmetry behind so much conservation laws is a crucial question to answer for understanding the structure of the solution space of the chiral model. A lot of work already shows that this phenomenon is closely related with dual symmetry. Results of previous papers[18] show that speaking more exactly the infinitesimal operator generating nonlocal currents is nothing else but the ordinary isospin generator $T$ transformed by DT (dual transformation) $U(x;\gamma)$. The DT with parameter $\gamma$ is the origin of the existence of infinitely many symmetries. Since a $U(x;\gamma_1)$ with a fixed $\gamma_1$ gives one automorphism in solution space, it maps one known explicit symmetry (e.g., constant $T$) into another hidden symmetry $U^{-1}(x;\gamma)TU(x;\gamma_1)$ generating a conserved current $J_\mu(x;\gamma_1)$ (cf. Sec. III). From the same $T$ but with different parameter $\gamma$ we get different symmetries $U^{-1}(x;\gamma)TU(x;\gamma)$ generating different currents $J_\mu(x;\gamma)$. In summary, dual symmetry is the symmetry which induces infinitely many symmetries and maps different currents, but itself is not the symmetry which generates the conserved currents $J_\mu$.

Accordingly, this paper tries at first to find out the infinitesimal variations which leave the Lagrangian unchanged, then takes the DT and thus gets the corresponding set of infinitely many symmetry operators generating conserved currents. In this way, after review the results about dual transformed $T$ shortly in Sec. III, we give subsequently in Sec. IV the current which corresponds to the infinitesimal generator of dual symmetry itself. We show that it is a Noether current and a dynamical symmetry of the equation of motion. The infinitesimal BT plays an important role in the soliton equation. In Sec. V we find the infinitesimal BT. For chiral model, it is given by the solution of a matrix Riccati equation; we show also the local current is just the related Noether current. In Sec. VI, we give the infinitesimal generators and Noether currents for more general cases, including the ordinary space-time translation and energy momentum density.

Since the finite dual transformation is quite well known now, the main role of the second section consists in introducing notations. By the way, deviating from the current conventions, which deal with gauge transformations within the isotropic subgroup $H$ only, we discuss somehow in detail the gauge transformations in the whole group $G$, so that the different formulations may be treated as gauge equivalent expressions and the distinction and relation between the connections, the second fundamental forms, and the invariantly conserved currents are clarified. We use the local involutive operator $N(x) = g(x)ng^{-1}(x)$ of the symmetric space as the dynamical variable, so that our formulation essentially includes the O($N$) nonlinear $\sigma$-model, the CP($N-1$) model, the Grassman chiral model, and the principle chiral model.

## II. CHIRAL MODEL IN VARIOUS GAUGES, DUAL SYMMETRY

### A. Symmetric space and canonical variable

The chiral field may be defined as a map from space time $x_\mu$ ($\mu = 0,1$) onto a symmetric space ($G$, $H$, $n$), i.e., a coset space $G/H$ with involutive automorphism $n$,

$$H = \{h \in G; nhn = h\}, \quad n^2 = 1, \tag{2.1}$$

where $G$ is a connected Lie group with Lie algebra $\mathfrak{G}$ and $H \subset G$ is a closed subgroup with Lie algebra $\mathfrak{H}$. In the adjoint representation the same matrix $n$ gives involutive automorphism for the Lie algebra also

$$[n,\mathfrak{H}] = 0, \quad \{n,\kappa\} = 0, \tag{2.2}$$

[a)] Permanent address: Northwest University, Xian, People's Republic of China.

where

$$\mathfrak{H} \oplus \kappa = \mathfrak{G},$$

$$[\mathfrak{H},\mathfrak{H}] \subset \mathfrak{H}, \quad [\mathfrak{H},\kappa] \subset \kappa, \quad [\kappa,\kappa] \subset \mathfrak{H}. \tag{2.3}$$

The elements of $G/H$ are represented by canonical variables

$$N(x) = g(x) n g^{-1}(x), \quad N(x)^2 = 1. \tag{2.4}$$

Then, if $g_1$ and $g_2$ are in the same coset class, $g_1 = g_2 h$, hence $g_1 n g_1^{-1} = g_2 n g_2^{-1}$, both correspond to the same $N$.

### B. Gauges with diagonal connections

The left Maurer Cartan form is divided into vertical (connection) and horizontal (second fundamental form) parts and pulled back onto $x$ space

$$a_\mu(x) = g^{-1}(x)\, \partial_\mu\, g(x) = h_\mu(x) + k_\mu(x), \tag{2.5}$$

where

$$[h_\mu(x), n] = 0, \tag{2.6}$$

$$\{k_\mu(x), n\} = 0, \tag{2.7}$$

$$h_\mu(x) = \tfrac{1}{2}[\, g^{-1}(x)\, \partial_\mu\, g(x) + n g^{-1}(x)\, \partial_\mu\, g(x) n\,], \tag{2.8}$$

$$k_\mu(x) = \tfrac{1}{2}[\, g^{-1}(x)\, \partial_\mu\, g(x) - n g^{-1}(x)\, \partial_\mu\, g(x) n\,], \tag{2.9}$$

in this gauge $h_\mu$ is diagonal with respect to $n$. The pure gauge $a_\mu$ has zero curvature $a_{\mu\nu}(x) = \partial_\mu a_\nu - \partial_\nu a_\mu + [a_\mu, a_\nu] = 0$; it may be divided into the Gauss equation

$$\begin{aligned} \tfrac{1}{2}[a_{\mu\nu}(x) + n a_{\mu\nu}(x) n] &= \partial_\mu\, h_\nu(x) - \partial_\nu\, h_\mu(x) \\ &\quad + [h_\mu(x), h_\nu(x)] + [k_\mu(x), k_\nu(x)] \\ &\equiv f_{\mu\nu}(x) + [k_\mu(x), k_\nu(x)] = 0 \end{aligned} \tag{2.10}$$

and the Coddazi equation

$$\begin{aligned} \tfrac{1}{2}[a_{\mu\nu}(x) - n a_{\mu\nu}(x) n] &= \partial_\mu\, k_\nu(x) + [h_\mu(x), k_\nu(x)] \\ &\quad - \partial_\nu\, k_\mu(x) - [h_\nu(x), k_\mu(x)] \\ &\equiv D_\mu\, k_\nu(x) - D_\nu\, k_\mu(x) = 0. \end{aligned} \tag{2.11}$$

### C. General gauge transformation

$$h'_\mu(x) = S^{-1}(x) h_\mu(x) S(x) + S^{-1}(x) \partial_\mu\, S(x),$$

$$k'_\mu(x) = S^{-1}(x) k_\mu(x) S(x). \tag{2.12}$$

Usually $S$ is restricted in $H$, then all relations (2.6)–(2.11) remains unchanged. If we allow $S(x)$ to be any element in $G$, then only (2.5), (2.7), (2.10), and (2.11) still remain valid, but the $n$ therein must be replaced by $n'(x) = S^{-1}(x) n S(x)$; meanwhile, instead of the diagonal of $h_\mu$ (2.6) and $\partial_\mu\, n = 0$, we have a covariant condition

$$D'_\mu\, n'(x) \equiv \partial_\mu\, n'(x) + [h'_\mu(x), n'(x)] = 0, \tag{2.13}$$

i.e., the reducibility condition for $h'_\mu$ [20]: "if there exists on the coset bundle $\{\{x\} G/H, G\}$ a section $n'(x)$ invariant under parallel displacement with respect to $h'_\mu$, then the $h'_\mu$ are reducible to a connection in $H$."

### D. Canonical gauge

Choosing $S^{-1}(x) = g(x)$, it occurs that both expressions in (2.12) are expressed solely by the canonical variable $N(x)$ in (2.4); thus

$$H_\mu(x) = \tfrac{1}{2} N(x)\, \partial_\mu\, N(x), \tag{2.14}$$

$$K_\mu(x) = -\tfrac{1}{2} N(x)\, \partial_\mu\, N(x). \tag{2.15}$$

In summary, we have the flat Gauss Coddazi equation

$$\begin{aligned} F_{\mu\nu}(x) &\equiv \partial_\mu\, H_\nu(x) - \partial_\nu\, H_\mu(x) + [H_\mu(x), H_\nu(x)] \\ &= -[K_\mu(x), K_\nu(x)], \end{aligned} \tag{2.16}$$

$$\epsilon^{\mu\nu} D_\mu\, K_\nu(x) \equiv \epsilon^{\mu\nu}(\partial_\mu\, K_\nu(x) + [H_\mu(x), K_\nu(x)]) = 0,$$

$$\epsilon^{10} = -\epsilon^{01} = 1, \tag{2.17}$$

the reducibility condition

$$D_\mu\, N(x) = \partial_\mu\, N(x) + [H_\mu(x), N(x)] = 0, \quad N(x)^2 = 1, \tag{2.18}$$

and the local involutive condition for $K_\mu$

$$\{K_\mu(x), N(x)\} = 0. \tag{2.19}$$

All equations (2.16)–(2.19) are gauge-covariant under (2.12). In addition we have chosen the canonical gauge condition

$$A_\mu(x) = H_\mu(x) + K_\mu(x) = 0; \tag{2.20}$$

then, from (2.18) and (2.19), we get the expressions of $H_\mu$, $K_\mu$ in terms of $N$ as (2.14) and (2.15).

It is interesting to point out that, complementary to the diagonal gauge (1.6), now

$$\{H_\mu(x), N(x)\} = 0. \tag{2.21}$$

Connection $H_\mu(x)$ is fixed by gauge condition (2.21), but we may further change the canonical gauge without breaking (2.21) by using $S = \exp(i\theta N(x))$, where $\theta$ is a constant parameter; then $K'_\mu(x) = \tfrac{1}{2}(\cos 2\theta N(x)\, \partial_\mu\, N(x) - i \sin 2\theta \partial_\mu\, N(x))$, e.g., $\theta = \tfrac{1}{2}\pi$, $K'_\mu(x) = H'_\mu(x) = \tfrac{1}{2} N(x)\, \partial_\mu\, N(x) = \tfrac{1}{2} A'_\mu(x)$.

### E. Dynamics

Let Lagrangian

$$L(x) = \tfrac{1}{8}\, \mathrm{tr}(\partial_\mu\, N(x) \partial^\mu N(x)), \quad N(x)^2 = 1 \tag{2.22}$$

and with some further constraints. The Euler–Lagrangian equation

$$[\partial_\mu\, \partial^\mu N(x), N(x)] = 0 \tag{2.23}$$

may be expressed in $K_\mu$ as

$$\partial_\mu\, K^\mu(x) = 0, \tag{2.24}$$

or rewritten into covariant form

$$D_\mu\, K^\mu(x) \equiv \partial_\mu\, K(x) + [H_\mu(x), K^\mu(x)] = 0. \tag{2.25}$$

### F. Intermediate DT, $K_\mu(x) \to \widetilde{K}_\mu(x;\gamma)$

Since (2.17) and (2.25) are mutually dual in two-dimensional space-time, it is obvious that (2.16)–(2.19) and (2.25) are invariant under DT:

$$\begin{aligned} K_\mu(x) \to \widetilde{K}_\mu(x;\gamma) &= K_\mu(x)(\gamma + \gamma^{-1})/2 + \epsilon_{\mu\nu}\, K^\nu(x)(\gamma - \gamma^{-1})/2, \\ &\equiv K_\mu(x) \cosh\phi + \epsilon_{\mu\nu}\, K^\nu(x) \sinh\phi, \end{aligned} \tag{2.26}$$

$$H_\mu(x) \to \tilde{H}_\mu(x;y) = H_\mu(x). \tag{2.27}$$

Thus, we have $\tilde{F}_{\mu\nu}(x;\gamma) = \partial_\mu \tilde{H}_\nu(x;\gamma) - \partial_\nu \tilde{H}_\mu(x;\gamma) + [\tilde{H}_\mu(x;\gamma), \tilde{H}_\nu(x;\gamma)] = -[\tilde{K}_\mu(x;\gamma), \tilde{K}_\nu(x,\gamma)]$ as ($2.\tilde{1}6$), and ($2.\tilde{1}7$)–($2.\tilde{1}9$) and ($2.\tilde{2}5$) by replacing $K_\mu(x)$ in (2.17)–(2.19) and (2.25) in terms of $\tilde{K}_\mu(x;\gamma)$. Since the explicitly pure condition (2.20) has been broken, $\tilde{H}_\mu$, $\tilde{K}_\mu$ could not be expressed directly by some $\tilde{N}$ as in (2.14) and (2.15). But from Eqs. ($2.\tilde{1}6$) and ($2.\tilde{1}7$), $\tilde{A}_\mu(x;\gamma) = \tilde{H}_\mu + \tilde{K}_\mu$ are pure gauge, so we may discover some new $N(x;\gamma)$ which satisfies the dynamical Eq. (2.23) as follows.

### G. Final dual transformation $N(x) \to N(x;\gamma)$

Equations ($2.\tilde{1}6$) and ($2.\tilde{1}7$) show that $\tilde{A}_\mu(x;\gamma)$ are pure gauge; therefore there exists an $U(x;\gamma)$ such that

$$U^{-1}(x;\gamma)\,\partial_\mu\, U(x;\gamma) = \tilde{A}_\mu(x;\gamma) \equiv \tilde{K}_\mu(x;\gamma) + \tilde{H}_\mu(x;\gamma) \tag{2.28}$$

or

$$\partial_\mu\, U(x;\gamma) = U(x;\gamma)(\tilde{K}_\mu(x;\gamma) - K_\mu(x)). \tag{2.29}$$

If we gauge transform $\tilde{H}_\mu(x;\gamma)$ with $S(x) = U^{-1}(x;\gamma)$, i.e., let

$$H_\mu(x;\gamma) = U(x;\gamma)\tilde{H}(x;\gamma)U^{-1}(x;\gamma) + U(x;\gamma)\,\partial_\mu\, U^{-1}(x;\gamma), \tag{2.30}$$

$$K_\mu(x;\gamma) = U(x;\gamma)\tilde{K}(x;\gamma)U^{-1}(x;\gamma). \tag{2.31}$$

Then, using (2.28), we get

$$A_\mu(x;\gamma) = H_\mu(x;\gamma) + K_\mu(x;\gamma) = 0. \tag{2.20$\gamma$}$$

Now, gauge covariant equations ($2.\tilde{1}6$)–($2.\tilde{1}9$), ($2.\tilde{2}5$) become (2.16$\gamma$)–(2.19$\gamma$), (2.25$\gamma$) after substituting:

$$\begin{aligned}&\tilde{H}_\mu(x;\gamma) \to H_\mu(x;\gamma), \quad \tilde{K}_\mu(x;\gamma) \to K_\mu(x,\gamma),\\ &\tilde{D}_\mu = D_\mu \to D_\mu(\gamma) \equiv \partial_\mu + [H_\mu(x;\gamma)],\\ &\tilde{N}(x;\gamma) \equiv N(x) \to N(x;\gamma),\end{aligned} \tag{2.32}$$

where $N(x;\gamma) \equiv U(x;\gamma)N(x)U^{-1}(x;\gamma)$. In gauge (2.20$\gamma$), Eq. (2.25$\gamma$), $D_\mu(\gamma)K^\mu(x;\gamma) = 0$, may be simplified as

$$\partial^\mu K_\mu(x;\gamma) = 0. \tag{2.24$\gamma$}$$

From (2.20$\gamma$) and (2.18$\gamma$) we have

$$H_\mu(x;\gamma) = \tfrac{1}{2} N(x;\gamma)\,\partial_\mu N(x;\gamma), \tag{2.14$\gamma$}$$

$$K_\mu(x;\gamma) = -\tfrac{1}{2} N(x;\gamma)\,\partial_\mu N(x;\gamma). \tag{2.15$\gamma$}$$

We may check (2.14$\gamma$) and (2.15$\gamma$) directly by substituting on their right-hand sides (2.32) and then use (2.29), ($2.\tilde{1}9$), (2.19), (2.30), or (2.31) to attain the left-hand side. Compare (2.32) with (2.4); we see that if $g(x;\gamma) = U(x;\gamma)g(x)$, then $N(x;\gamma) = g(x;\gamma)ng^{-1}(x;\gamma)$. Finally from (2.24$\gamma$) and (2.15$\gamma$) we get the dual transformed EL equation (2.23$\gamma$). (Equations labeled with $\gamma$, are just the same equation, only with $N$, $H_\mu$, $K_\mu$ replaced by $N\langle\gamma\rangle$, $H_\mu\langle\gamma\rangle$, $K_\mu\langle\gamma\rangle$.)

In the latter we shall adopt following abbreviations:

$H_\mu \equiv H_\mu(x)$, $K_\mu \equiv K_\mu(x)$, $N \equiv N(x)$, $\tilde{K}_\mu \equiv \tilde{K}_\mu(x;\gamma)$,
$N\langle\gamma\rangle \equiv N(x;\gamma)$, $H_\mu\langle\gamma\rangle \equiv H_\mu(x;\gamma)$,
$K_\mu\langle\gamma\rangle \equiv K_\mu(x;\gamma)$, $U \equiv U(x;\gamma)$.

## III. DUAL TRANSFORMATION OF ISOTOPIC SYMMETRY OPERATOR

### A. Ordinary generator for conserved current

Let

$$\delta N(x) = -[N(x), \Lambda(x)]\delta\epsilon. \tag{3.1}$$

(For simplicity, we omit the infinitesimal constant $\delta\epsilon$ in the future.) Then

$$\delta L = \operatorname{tr}(K_\mu\, \partial^\mu \Lambda). \tag{3.2}$$

Define

$$j^\mu(x) = \frac{\delta L}{\delta\partial_\mu N}\delta N = \operatorname{tr}(K^\mu \Lambda). \tag{3.3}$$

Its on-shell (2.24) divergence equals

$$\partial_\mu j^\mu = \operatorname{tr}(K_\mu\, \partial^\mu\Lambda) = \delta L. \tag{3.4}$$

If we have chosen $\Lambda(x)$ such that

$$\operatorname{tr}(K_\mu\, \partial^\mu\Lambda) = 0. \tag{3.5}$$

Then $j_\mu(x)$ is conserved

$$\partial_\mu J^\mu(x) = 0. \tag{3.6}$$

For example, let $\Lambda(x) \equiv T$, where $T$ is a constant element in $g$. Then

$$J_\mu = \operatorname{tr}(K_\mu\, T) \tag{3.7}$$

is a conserved current.

### B. Dual transformed current

Heuristically, in the dual transformed functional space with canonical variable $N(x;\gamma)$, let $L(x;\gamma) = \frac{1}{8}\operatorname{tr}(N\langle\gamma\rangle N\langle\gamma\rangle)$; take $\delta N\langle\gamma\rangle = -[N\langle\gamma\rangle, T]$. We get $J_\mu\langle\gamma\rangle = \operatorname{tr}(K_\mu\langle\gamma\rangle T)$, (3.7$\gamma$), which is conserved because of (1.24$\gamma$). Expanding $J_\mu\langle\gamma\rangle$ into series of $\gamma$; we get an infinite series of conserved nonlocal currents.

### C. Dual transformed generator

Now, return to the original functional space. Tentatively, neglecting the dependence of $U^{-1}(x;\gamma)$ on $T\delta\epsilon$ via $K_\mu(x;\gamma)$, assume

$$\begin{aligned}\delta N(x) &= U^{-1}(x;\gamma)\delta N(x;\gamma)U(x;\gamma)\\ &= -[N(x), U^{-1}(x;\gamma)TU(x;\gamma)];\end{aligned} \tag{3.8}$$

subsequently, $j_\mu(x;\gamma) = \operatorname{tr}(K_\mu(x)U^{-1}(x;\gamma)TU(x;\gamma))$, but it occurs to us that now its on shell divergence

$$\partial_\mu j^\mu(x;\gamma) = \delta L = \operatorname{tr}(K_\mu\, \partial^\mu(U^{-1}TU)) \neq 0. \tag{3.9}$$

However, using (2.29), (2.26), and (2.17), one may show that

$$\begin{aligned}\delta L &= -\sinh\phi\, \operatorname{tr}(\epsilon_{\mu\nu}\, K^\mu K^\nu T)\\ &= -\tanh\phi\, \partial_\mu\, \operatorname{tr}(\epsilon^{\mu\nu} K_{\mu\nu}\, U^{-1}TU)\\ &\equiv \partial_\mu\, i^\mu(x;\gamma).\end{aligned} \tag{3.10}$$

Put (3.10) together with (3.9); we regain the conserved current (3.7$\gamma$) with some coefficient, i.e.,

$$\begin{aligned}J_\mu(x;\gamma) &\equiv j_\mu(x;\gamma) + i_\mu(x;\gamma) = \operatorname{sech}\phi\, \operatorname{tr}(\tilde{K}_\mu\, U^{-1}TU)\\ &= \operatorname{sech}\phi\, \operatorname{tr}(K_\mu(\gamma)T),\end{aligned} \tag{3.11}$$

where (2.31) has been used.

Thus, we see that just as $j\mu$ (3.7) is related to the symmetry of rotation $\delta\epsilon$ around the fixed $T$ axis, $J_\mu\langle\gamma\rangle$ (3.7$\gamma$) or (3.11) is related to the rotation $\delta\epsilon$ around the transformed axis $U^{-1}TU$.

The variation (3.8) satisfies the condition for invariance of EL equation (2.23) under $\delta N$

$$D_\mu D^\mu[N,\delta N] = [K_\mu,[K^\mu,[N,\delta N]]] \qquad (3.12)$$

or, using (3.1),

$$D_\mu D^\mu A - [K_\mu,[K^\mu,A]] - N(D_\mu D^\mu A - [K_\mu,[K^\mu,A]])N = 0. \qquad (3.13)$$

At last, we emphasize that $K_\mu$ does not conserve invariantly with respect to local gauge transformation, as a covariant quantity; it conserves only covariantly (2.25) in general gauge. The true invariantly conserved currents are always gauge-invariant quantities such as projections of $K_\mu$ on $T$ or $\widetilde{K}_\mu$ on $U^{-1}TU$, etc., i.e., $\mathrm{tr}(K_\mu\, T)$ or $\mathrm{tr}(\widetilde{K}_\mu\, U^{-1}TU)$, etc. (cf. later sections).

## IV. INFINITESIMAL DUAL TRANSFORMATION

### A. Finite DT

Under finite DT (2.23), the finite variation of $L$ equals zero

$$\Delta L = \tfrac{1}{8}\,\mathrm{tr}(K_\mu\langle\gamma\rangle K^\mu\langle\gamma\rangle - K_\mu K^\mu) = \tfrac{1}{8}\,\mathrm{tr}(\widetilde{K}_\mu \widetilde{K}^\mu - K_\mu K^\mu) = 0. \qquad (4.1)$$

### B. Infinitesimal DT

But in order to find out the corresponding conserved currents, we must use the infinitesimal DT operator $u(x)$:

$$u(x) \equiv \left(\gamma \frac{dU(x;\gamma)}{d\gamma} U^{-1}(x;\gamma)\right)\Bigg|_{\gamma=1} = -\int_{-\infty}^{x_1} K_0(x_0, x_1')\, dx^1. \qquad (4.2)$$

It satisfies

$$\partial_\mu u(x) = \epsilon_{\mu\nu} K^\nu(x) \qquad (4.3)$$

from (4.2) and (2.24). The covariant form of (4.3) is

$$D_\mu u = -[K_\mu, u] + \epsilon_{\mu\nu} K^\nu. \qquad (4.4)$$

Now let

$$\delta N = -[N, u]; \qquad (4.5)$$

we have

$$\delta L = \mathrm{tr}(K_\mu\, \partial^\mu u) = \mathrm{tr}(\epsilon^{\mu\nu} K_\mu K_\nu) = 0. \qquad (4.6)$$

(Really, $K_\mu$ in $L$ has been changed into its dual $\epsilon_{\mu\nu} K^\nu$.) Therefore,

$$J_\mu = \mathrm{tr}(K_\mu\, u) \text{ is a conserved current.} \qquad (4.7)$$

### C. Dual transformed infinitesimal DT

Let

$$\tilde{u}(x;\gamma) = U^{-1}(x;\gamma) u(x;\gamma) U(x;\gamma), \qquad (4.\tilde{2})$$

where

$$u(x;\gamma) = -\int_{-\infty}^{x_1} K_0(x_0, x_1';\gamma)\, dx_1' = \gamma \frac{dU}{d\gamma} U^{-1}; \qquad (4.2\gamma)$$

they satisfy

$$\partial_\mu u(x;\gamma) = \epsilon_{\mu\nu} K^\nu(x;\gamma), \qquad (4.3\gamma)$$

$$D_\mu\langle\gamma\rangle u\langle r\rangle = -[K_\mu\langle\gamma\rangle, u\langle\gamma\rangle] + \epsilon_{\mu\nu} K^\nu\langle\gamma\rangle, \qquad (4.4\gamma)$$

$$D_\mu \tilde{u} = -[\widetilde{K}_\mu, \tilde{u}] + \epsilon_{\mu\nu} \widetilde{K}^\nu. \qquad (4.\tilde{4})$$

Let

$$\delta N = -[N, \tilde{u}]. \qquad (4.\tilde{5})$$

Then

$$j_\mu(x;\gamma) = \mathrm{tr}(K_\mu(x)\tilde{u}(x;\gamma)), \qquad (4.8)$$

$$\partial_\mu j^\mu(x;\gamma) = \delta L = \mathrm{tr}(K_\mu\, \partial^\mu \tilde{u}) = \sinh\phi\, \mathrm{tr}(-\epsilon^{\mu\nu}[K_\mu, K_\nu]\tilde{u} + K_\mu K^\mu) = -\tanh\phi\, \partial_\mu\, \mathrm{tr}(\epsilon^{\mu\nu} K_\nu\, \tilde{u}) \equiv -\partial_\mu\, i^\mu(x;\gamma). \qquad (4.9)$$

$$J_\mu(x;\gamma) \equiv j_\mu(x;\gamma) + i_\mu(x;\gamma) = \mathrm{sech}\,\gamma\, \mathrm{tr}(\widetilde{K}_\mu\, \tilde{u}) = \mathrm{sech}\,\phi\, \mathrm{tr}(K_\mu\langle\gamma\rangle u\langle\gamma\rangle) \qquad (4.10)$$

is conserved.

It is easy to check that the EL equation is invariant under (4.5) by substituting it into (3.12).

In the two-dimensional Euclidean space with self-dual (anti-dual) solution $\partial_\mu N(x) = \pm\epsilon_{\mu\nu} N(x)\partial^\nu N(x)$, we get $u(x) = \pm N(x)$. All these currents are trivial.

## V. BÄCKLUND TRANSFORMATION

### A. Finite BT

It operates on solution $N(x)$ of (2.23); giving a new solution

$$N'(x|\gamma) = N(x)B(x|\gamma) = B^+(x|\gamma)N(x) \qquad (5.1)$$

when $N$, $N'$, $B$ satisfy

$$2K_\mu - 2K_\mu' \equiv N'\partial_\mu N' - N\partial_\mu N = \epsilon_{\mu\nu}\partial^\nu B, \qquad (5.2)$$

$$B(x|\gamma) + B^+(x|\gamma) = -2\tanh\phi. \qquad (5.3)$$

Let

$$\widetilde{R} \equiv \tfrac{1}{2}\cosh\phi\,(B(x|\gamma) - B^+(x|\gamma)); \qquad (5.4)$$

then from (5.1)–(5.3)

$$D_\mu \widetilde{R} \equiv \partial_\mu \widetilde{R} + [H_\mu, \widetilde{R}] = \epsilon^{\mu\nu}(\widetilde{K}_\nu + \widetilde{R}\widetilde{K}_\nu \widetilde{R}). \qquad (5.5)$$

It is integrable from (2.16), (2.17), and (2.25). Conversely, from $R$, satisfying (5.5), let

$$B(x\gamma) = \sec\phi\widetilde{R} - \tanh\phi = \exp(2(\cot^{-1}\phi)\widetilde{R}); \qquad (5.6)$$

we obtain from (5.1) the new solution $N'$. The BT (5.1), (5.2) satisfies variational Bäcklund principle, i.e., $\delta L$ equals total divergence:

$$\Delta L = \tfrac{1}{8}\,\mathrm{tr}(K_\mu' K^{\mu\prime} - K_\mu K^\mu) = -\tfrac{1}{4}\,\mathrm{sech}^2\phi\, tr(K_\mu K^\mu + \widetilde{R}K^\mu \widetilde{R}K_\mu) = \tfrac{1}{8}\,\mathrm{sech}^3\phi\,\partial_\mu\, \mathrm{tr}(\epsilon^{\mu\nu}K_\nu \widetilde{R}) = \tfrac{1}{8}\,\mathrm{csch}\,\phi\, \mathrm{tr}\,\partial_\mu(\widetilde{K}^\mu \widetilde{R}). \qquad (5.7)$$

### B. Infinitesimal BT

Let

$$\delta N = \left[ N, B^{-1} \frac{dB}{d\gamma}\Big|_{\gamma=1} \right]$$
$$= 2\left[ N, \frac{-1}{(1+\gamma^2)} B + \cot^{-1}\gamma \frac{dR}{\alpha\gamma} \right]\Big|_{\gamma=1}. \quad (5.8)$$

The contribution of the second term in $\delta L$ equals

$$2\cot^{-1}\gamma\, \mathrm{tr}\left( K_\mu\, \partial^\mu \frac{dR}{d\gamma} \right)$$
$$= 2\gamma^2 \cot^{-1}\gamma\, \mathrm{tr}(K_\mu K^\mu + \widetilde{R}K^\mu \widetilde{R}K_\mu)/(1+\gamma^2), \quad (5.9)$$

which is a total divergence as the rhs of (5.7). Hence, we omit this term, keep the first only. Since we need dual transformed operator later, we replace $B(x|1)$ by

$$\widetilde{R}(x;1) = R(x;1) = B(x;1) \equiv R(x),$$

where

$$R(x;\gamma) = U(x;\gamma)\widetilde{R}(x;\gamma)U^{-1}(x;\gamma). \quad (5.10)$$

It satisfies

$$D_\mu \langle\gamma\rangle R(x;\gamma) = \epsilon_{\mu\nu}(K^\nu(x;\gamma) + R(x;\gamma)K^\nu(x;\gamma)R(x;\gamma)); \quad (5.5\gamma)$$

thus, we take

$$\delta N(x) = [N(x), B(x|1)] = [N(x), R(x)]. \quad (5.11)$$

Since now $\delta L = \mathrm{tr}(K_\mu(x)\partial^\mu R(x)) = 0$, the current $\mathrm{tr}(K_\mu(x)R(x))$ are conserved.

### C. Dual transformed BT

Let

$$\delta N(x;\gamma) = [N(x), \widetilde{R}(x;\gamma)], \quad (5.11\gamma)$$

we have

$$j_\mu(x;\gamma) = \mathrm{tr}(K_\mu \widetilde{R}), \quad (5.12)$$

$$\partial_\mu j^\mu(x;\gamma) = \delta L = \mathrm{tr}(K_\mu\, \partial^\mu \widetilde{R})$$
$$= -2\sinh\phi\, \mathrm{tr}(K_\mu K^\mu + K_\mu \widetilde{R}K^\mu \widetilde{R}),$$
$$= -\partial_\mu\, \mathrm{tr}(\epsilon^{\mu\nu}\widetilde{K}_\nu \widetilde{R})\tanh\phi \equiv -\partial_\mu\, i^\mu(x;\gamma). \quad (5.13)$$

Finally, we get the conserved current

$$J^\mu(x;\gamma) \equiv j^\mu(x;\gamma) + i^\mu(x;\gamma) = \mathrm{sech}\,\Phi\, \mathrm{tr}(\widetilde{K}_\mu \widetilde{R})$$
$$= \mathrm{sech}\,\phi\, \mathrm{tr}(K^\mu(x;\gamma)R(x;\gamma)). \quad (5.14)$$

The geometrical meaning are rotations around axis $\widetilde{R}$; $B(x|\gamma)$ are finite rotations with angle $\theta = 2\cot^{-1}\gamma$, while the $\delta N(x;|\gamma)$ are generated by rotation with infinitesimal constant angle $\delta\epsilon$.

## VI. GENERAL CASE

Generally, we must find $\Lambda(x)$ such that

$$\mathrm{tr}(\widetilde{K}_\mu D^\mu \Lambda) = 0. \quad (6.1)$$

The most general equation for $\Lambda$ is

$$D_\mu \Lambda = \alpha\epsilon_{\mu\nu}\widetilde{K}^\nu + \beta\epsilon_{\mu\nu}\Lambda\widetilde{K}^\nu\Lambda$$
$$+ a[\Lambda, \widetilde{K}_\mu] + s\epsilon_{\mu\nu}\{\Lambda, \widetilde{K}^\nu\}; \quad (6.2)$$

it is integrable if $a^2 - s^2 + \alpha\beta = 1$. Let

$$\delta N = [N, \Lambda]. \quad (6.3)$$

Then,

$$j_\mu(x;\gamma) = \mathrm{tr}(K_\mu \Lambda), \quad (6.4)$$

$$\partial_\mu j^\mu(x;\gamma)$$
$$= \delta L = \mathrm{tr}(K_\mu\, \partial^\mu \Lambda) = -\tanh\phi\, \mathrm{tr}(\epsilon^{\mu\nu}K_\nu D_\mu \Lambda)$$
$$= -\tanh\phi\, \mathrm{tr}\, \partial_\mu(\epsilon^{\mu\nu}K_\mu \Lambda) \equiv -\partial_\mu i^\mu, \quad (6.5)$$

so

$$J_\mu \equiv j_\mu(x;\gamma) + i_\mu(x;\gamma) = \mathrm{sech}\,\phi\, \mathrm{tr}(\widetilde{K}_\mu \Lambda)$$

are conserved. This includes all currents discussed above:

if $a = 1$, $\alpha = \beta = s = 0$, $\Lambda = U^{-1}(x;\gamma)TU(x;\gamma)$ in Sec. III;
if $a = \alpha = 1$, $\beta = s = 0$, $\Lambda = \tilde{u}$ in Sec. IV;
if $\alpha = \beta = 1$, $a = s = 0$, $\Lambda = \widetilde{R}$ in Sec. V;

more generally, if $\mathrm{tr}[K^\mu, D_\mu \Lambda] = \partial_\mu \tilde{l}^\mu$, then let $\delta N = [N, \Lambda]$; we get the conserved current

$$J_\mu = \mathrm{sech}\,\phi\, (\mathrm{tr}(\widetilde{K}_\mu \Lambda) - \tilde{l}_\mu). \quad (6.6)$$

For example, under infinitesimal translation, $\delta N(x;\gamma) = \partial_\nu N(x;\gamma)$. Let

$$\delta N(x) = U^{-1}\delta N(x;\gamma)U = U^{-1}\partial_\nu N(x;\gamma)U$$
$$= -U^{-1}[N(x;\gamma), K_\nu(x;\gamma)]U$$
$$= -[N(x), \widetilde{K}_\nu(x;\gamma)], \quad (6.7)$$

i.e.,

$$\Lambda(x,\gamma) = \widetilde{K}_\nu(x;\gamma). \quad (6.8)$$

Then

$$\mathrm{tr}(\widetilde{K}_\mu D^\mu \widetilde{K}_\nu) = -\tfrac{1}{2}\partial_\mu\, \mathrm{tr}(\widetilde{K}_\mu \widetilde{K}^\mu) \equiv -\partial_\mu \tilde{l}_\mu. \quad (6.9)$$

The current (6.6) becomes energy momentum density $M_{\mu\nu}$

$$J_\mu = \mathrm{sech}\,\phi\, \mathrm{tr}(K_\mu K_\nu - \tfrac{1}{2}g_{\mu\nu}K_\lambda K^\lambda).$$

## VII. DISCUSSION

Thus, we formulate a general way to get infinitely many Noëther currents from any given Noëther current.

If we expand the generator $U^{-1}\langle\gamma\rangle TU\langle\gamma\rangle$ in series of the parameter $\lambda \equiv (\gamma-1)/(\gamma+1)$, we would obtain the series of generators of the so-called Kac–Moody algebra.[21] Meanwhile, to get the recurrence formulas for each order, one may simply use $\partial_\mu(U^{-1}TU) = \lambda\epsilon_{\mu\nu}(\partial^\mu U^{-1}TU + [H^\nu - K^\nu, U^{-1}TU])$. But the form $U^{-1}\Lambda U$ shows more apparently the origin of symmetry—dual transformed isotopic symmetry $T$, etc.; and the related current is constructed explicitly from the dual transformed solution $N\langle\gamma\rangle$ in the same way as the original current from original $N$. All our currents are related to a given symmetry of the action. Almost all of them (except the infinitesimal BT) keep the equation of motion invariant, while each elements of the Kac–Moody algebra (except the zero-order one) does not generate the symmetry of the original equation.

We have found a lot of new Noëther currents and related generators. It is interesting to point out that the infinitesimal generator $u(x)$ of dual transformation (which is the Lie transformation for the related sine–Gordon equation[22]) is just the position vector[23] of the so-called soliton surface,[24] in the case of the O(3) $\sigma$-model; it is the well-known pseudospherical surface with $N(x)$ as its normal and $\partial_\xi N$, $\partial_\eta N$ as its asymptotic directions. Then Eq. (5.2) becomes $2du-2du' = \cosh\phi\, dR$, we can identify the Riccati function $R\langle\gamma\rangle \cosh\phi$ as the common tangent of two pseudospherical surfaces.[23] Using the covariance of our formulation, we can show that $\mathrm{tr}(K_\mu\langle\gamma\rangle R\langle\gamma\rangle)$ gives the series of local conservation current in the ordinary soliton theory and is related to a total geodesic differential along the common tangent direction.

Our formulation is easy to generalize to supersymmetric cases.[25] Then, from the dual similar of the supersymmetric generator, we get infinitely many supersymmetric currents correspondingly obtaining Kac–Moody algebra with both anticommutators and commutators.

## ACKNOWLEDGMENTS

The author would like to thank Professor C. N. Yang for his hospitality and for the support of the ITP, SUNY at Stony Brook. This work was supported in part by NSF Grant #PHY-81-09110-A-01.

[1]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).
[2]A. A. Belavin and A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. **22**, xxx (1975) [JETP Lett. **22**, 245 (1975)].
[3]H. Eichenherr, Nucl. Phys. B **146**, 215 (1978); B **115**, 544 (1979).
[4]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979); B **164**, 528 (1980); Commun. Math. Phys. **82**, 227 (1981).
[5]V. E. Zakharov and A. V. Mikhailov, Zh. Eksp. Teor. Fiz. **74**, 1953 (1978) [Sov. Phys. JETP **47**, 1017 (1978)].
[6]M. Lüscher and K. Pohlmeyer, Nucl. Phys. B **137**, 46 (1978).
[7]D. Gross, Nucl. Phys. B **132**, 439 (1978); F. Gürsey and X. C. Tze, Ann. Phys. (N.Y.) **128**, 29 (1980); Bo-yu Hou, Y. B. Wang, Bo-yuan Hou, and P. Wang, Phys. Lett. B **93**, 415 (1980).
[8]A. T. Ogielski, M. K. Prasad, A. Sinha, and L. L. Chau Wang, Phys. Lett. B **91**, 387 (1980); M. K. Prasad, A. Sinha, and L. L. Chau Wang, Phys. Lett. B **87**, 237 (1979). L. L. Chau and T. Koikawa, Phys. Lett. B **123**, 47 (1983).
[9]K. Scheler, Phys. Lett. B **93**, 331 (1980); Z. Phys. C **6**, 365 (1980).
[10]K. C. Chou and X. C. Song, Sci. Sin. A **25**, 716 (1982).
[11]B. Brezin, C. Itzykson, J. Zinn-Justin, and J. B. Zuber, Phys. Lett. B **82**, 442 (1979).
[12]H. J. de Vega, Phys. Lett. B **87**, 233 (1979).
[13]A. T. Ogielski, Phys. Rev. D **21**, 406 (1980).
[14]L. Dolan and A. Roos, Phys. Rev. D **22**, 2018 (1980).
[15]E. Corrigan and C. K. Zachos, Phys. Lett. B **88**, 273 (1979); T. L. Cutright and C. K. Zachos, Phys. Rev. D **21**, 441 (1980).
[16]C. K. Zachos, Phys. Rev. D **21**, 3462 (1980).
[17]R. P. Zaikov, Teor. Math. Phys. **53**, 238 (1982).
[18]B. Y. Hou, Comm. Theor. Phys. **1**, 333 (1982); B. Y. Hou, M. L. Ge, and Y. S. Wu, Phys. Rev. D **24**, 2238 (1981); C. Devchand and D. B. Fairlie, Nucl. Phys. B **194**, 232 (1982).
[19]I. V. Cherednik, Theor. Math. Phys. **23**, 120 (1979); **41**, 997 (1979).
[20]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1969), Vol. 1.
[21]L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981); M. L. Ge and Y. S. Wu, Phys. Lett. B **108**, 411 (1982); K. Ueno and Y. Nakamura, Phys. Lett. B **109**, 273 (1982).
[22]S. S. Chern and C. L. Terng, Rocky Mountain J. Math. **10**, 105 (1980).
[23]A. Sym. Lett. Nuov Cimento **33**, 394 (1982).
[24]B. Y. Hou, B. Y. Hou, and P. Wang, preprint, University of Inner Mongolia, 1982.
[25]E. Witten, Phys. Rev. D **16**, 2991 (1977); P. D. Vecchia and S. Ferrara, Nucl. Phys. B **130**, 93 (1977); E. Cremmer and J. Scherk, Phys. Lett. B **74**, 341 (1978); E. Napolitano and S. Sciuto, Nuov. Cimento **64**, 406 (1981); Commun. Math. Phys. **84**, 171 (1982); Phys. Lett. B **113**, 43 (1982); Z. Popowicz and L. L. Chau Wang, Phys. Lett. B **98**, 253 (1981); L. L. Chau, M. L. Ge, and Y. S. Wu, Phys. Rev. D **25**, 1080 (1982); J. F. Shonfeld, Nucl. Phys. B **109**, 49 (1980).

# On the integrability of nonlinear Dirac equations

W-H. Steeb and W. Oevel
*Theoretische Physik, Universität Paderborn, D-4790 Paderborn, West Germany*

W. Strampp
*FB Mathematik, Gesamthochschule Kassel, D-3500 Kassel, West Germany*

(Received 19 April 1983; accepted for publication 24 February 1984)

The integrability of nonlinear Dirac equations is discussed applying recent results in soliton theory. Using the Lie point transformation groups of the nonlinear Dirac equations we reduce these partial differential equations to systems of ordinary differential equations and study whether these systems are integrable. We also discuss whether Lie–Bäcklund vector fields exist. We conclude that the nonlinear Dirac equations are not integrable.

PACS numbers: 11.30. – j, 11.10.Lm, 02.20. + b

## I. INTRODUCTION

Evolution equations which can be solved by the inverse scattering transform (IST) are usually called soliton equations. Soliton equations have several properties in common: (I) the initial value problem can be solved exactly with the help of the IST; (II) they have an infinite number of conservation laws; (III) they have auto Bäcklund transformations; (IV) besides Lie point vector fields they admit Lie–Bäcklund (LB) vector fields; (V) they describe pseudospherical surfaces, i.e., surfaces of constant negative Gaussian curvature; and (VI) they can be written as covariant exterior derivative of Lie algebra valued differential forms. It is conjectured that if property (I) holds, then the properties (II)–(VI) also hold. If one of these conditions is satisfied for an evolution equation, then this equation is usually called integrable.

Recently several authors[1–7] have investigated the connection between nonlinear evolution equations and the Painlevé property. The following conjecture has been made: "Every nonlinear ordinary differential equation (ode) resulting from a group theoretical reduction of a nonlinear partial equation (pde) which can be solved by the IST has the Painlevé property." Under the Painlevé property of an ode (considered in the complex domain) we understand the following: The only movable singularities of all its solutions are poles. We notice that a solution of an ode can have poles, essential singularities, and branch points. Consequently, for an ode to have the Painlevé property we must require that there are no movable essential singularities or movable branch points. It is assumed that if an ode (or a system of ode's) has the Painlevé property, then this system is integrable. However, we cannot conclude that, in general, an integrable system has the Painlevé property.

In the present paper we investigate the integrability of nonlinear Dirac equations. So far efforts have not been successful in finding whether nonlinear Dirac equations satisfy one of the properties given above (even in one space dimension). First of all we give the Lie point symmetry groups for a class of nonlinear Dirac equations in three space dimensions. These groups will be used for reducing the system of pde's to systems of ode's, where we restrict ourselves to one space dimension. These systems will be investigated as to their integrability in order to decide whether the nonlinear Dirac equations are integrable or not. If the systems of ode's are not integrable, then we can conclude that the system of pde's is not integrable. On the other hand, if we find that the systems of ode's are integrable, then no conclusion can be made. Furthermore we discuss whether a certain nonlinear Dirac equation (in one space dimension) can be written as a covariant derivative of Lie algebra valued differential forms and whether LB vector fields exist.

We also consider the massive Thirring model, because it can be solved by IST.[8–10] We also give the Lie point symmetry groups and perform group theoretical reductions. We show that the massive Thirring model can be written as a covariant derivative of Lie algebra valued differential forms. Moreover we give a LB vector field of this model.

## II. SYMMETRY GROUPS OF NONLINEAR DIRAC EQUATIONS

Nonlinear Dirac equations for constructing models of extended particles have been investigated by various authors.[11–24] Various types of nonlinearity have been studied. In particular the interest has been focused on the scalar interaction, i.e., in the Lagrangian the interaction term is given by $(\bar{\psi}\psi)^2$ ($\psi$ is a four-component Dirac spinor). The Lie point symmetry vector fields for this interaction have been given in the papers cited above. Let us summarize the results.

Consider the nonlinear Dirac equations

$$\sum_{k=1}^{3} \frac{\partial}{\partial x_k}(\gamma_k \psi) - i\frac{\partial}{\partial x_4}(\gamma_4\psi) + l^2\psi(\bar{\psi}\psi) = 0, \quad (1)$$

and

$$\lambda \sum_{k=1}^{3} \frac{\partial}{\partial x_k}(\gamma_k \psi) - \lambda i\frac{\partial}{\partial x_4}(\gamma_4\psi) + \psi + \lambda^3 \epsilon\psi(\bar{\psi}\psi) = 0. \quad (2)$$

Equation (2) contains a mass term, whereas Eq. (1) does not. Both the quantities $l$ and $\lambda$ have the dimension of a length. Now we give the symmetry groups, i.e., the infinitesimal generators (symmetry vector fields). With the help of a Lie series we can find the symmetry group. The technique for finding the symmetry vector fields has been described by several authors (for example, in Ref. 25). In the following we use the notation given by Steeb *et al.*[17] In this notation we put $\psi_j = u_j + iv_j$, where $j = 1, \dots, 4$. Consequently, the quantities $u_j$ and $v_j$ are real fields. Thus both Eqs. (1) and (2) are a coupled system of eight nonlinear pde's.

  

**Theorem 1:** The nonlinear Dirac equation (1) is invariant under the Lie point symmetry groups which are generated by the infinitesimal generators

$$X_1 = \frac{\partial}{\partial x_1},\ X_2 = \frac{\partial}{\partial x_2},\ X_3 = \frac{\partial}{\partial x_3},\ X_4 = \frac{\partial}{\partial x_4},$$

$$R_{12} = x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2} - \frac{v_1}{2}\frac{\partial}{\partial u_1} - \frac{v_2}{2}\frac{\partial}{\partial u_2} - \frac{v_3}{2}\frac{\partial}{\partial u_3} + \frac{v_4}{2}\frac{\partial}{\partial u_4} + \frac{u_1}{2}\frac{\partial}{\partial v_1} - \frac{u_2}{2}\frac{\partial}{\partial v_2} + \frac{u_3}{2}\frac{\partial}{\partial v_3} - \frac{u_4}{2}\frac{\partial}{\partial v_4},$$

$$R_{13} = x_3\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_3} - \frac{u_2}{2}\frac{\partial}{\partial u_1} + \frac{u_1}{2}\frac{\partial}{\partial u_2} - \frac{u_4}{2}\frac{\partial}{\partial u_3} + \frac{u_3}{2}\frac{\partial}{\partial u_4} - \frac{v_2}{2}\frac{\partial}{\partial v_1} + \frac{v_1}{2}\frac{\partial}{\partial v_2} - \frac{v_4}{2}\frac{\partial}{\partial v_3} + \frac{v_3}{2}\frac{\partial}{\partial v_4}, \quad (3)$$

$$R_{23} = x_3\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_3} - \frac{v_2}{2}\frac{\partial}{\partial u_1} - \frac{v_1}{2}\frac{\partial}{\partial u_2} - \frac{v_4}{2}\frac{\partial}{\partial u_3} - \frac{v_3}{2}\frac{\partial}{\partial u_4} + \frac{u_2}{2}\frac{\partial}{\partial v_1} + \frac{u_1}{2}\frac{\partial}{\partial v_2} + \frac{u_4}{2}\frac{\partial}{\partial v_3} + \frac{u_3}{2}\frac{\partial}{\partial v_4},$$

$$L_{14} = x_4\frac{\partial}{\partial x_1} + x_1\frac{\partial}{\partial x_4} + \frac{u_4}{2}\frac{\partial}{\partial u_1} + \frac{u_3}{2}\frac{\partial}{\partial u_2} + \frac{u_2}{2}\frac{\partial}{\partial u_3} + \frac{u_1}{2}\frac{\partial}{\partial u_4} + \frac{v_4}{2}\frac{\partial}{\partial v_1} + \frac{v_3}{2}\frac{\partial}{\partial v_2} + \frac{v_2}{2}\frac{\partial}{\partial v_3} + \frac{v_1}{2}\frac{\partial}{\partial v_4},$$

$$L_{24} = x_4\frac{\partial}{\partial x_2} + x_2\frac{\partial}{\partial x_4} + \frac{v_4}{2}\frac{\partial}{\partial u_1} - \frac{v_3}{2}\frac{\partial}{\partial u_2} + \frac{v_2}{2}\frac{\partial}{\partial u_3} - \frac{v_1}{2}\frac{\partial}{\partial u_4} - \frac{u_4}{2}\frac{\partial}{\partial v_1} + \frac{u_3}{2}\frac{\partial}{\partial v_2} - \frac{u_2}{2}\frac{\partial}{\partial v_3} + \frac{u_1}{2}\frac{\partial}{\partial v_4},$$

$$L_{34} = x_4\frac{\partial}{\partial x_3} + x_3\frac{\partial}{\partial x_4} + \frac{u_3}{2}\frac{\partial}{\partial u_1} - \frac{u_4}{2}\frac{\partial}{\partial u_2} + \frac{u_1}{2}\frac{\partial}{\partial u_3} - \frac{u_2}{2}\frac{\partial}{\partial u_4} + \frac{v_3}{2}\frac{\partial}{\partial v_1} - \frac{v_4}{2}\frac{\partial}{\partial v_2} + \frac{v_1}{2}\frac{\partial}{\partial v_3} - \frac{v_2}{2}\frac{\partial}{\partial v_4},$$

$$J_0 = \sum_{j=1}^{4}\left(v_j\frac{\partial}{\partial u_j} - u_j\frac{\partial}{\partial v_j}\right),$$

$$J_1 = u_4\frac{\partial}{\partial u_1} - u_3\frac{\partial}{\partial u_2} - u_2\frac{\partial}{\partial u_3} + u_1\frac{\partial}{\partial u_4} - v_4\frac{\partial}{\partial v_1} + v_3\frac{\partial}{\partial v_2} + v_2\frac{\partial}{\partial v_3} - v_1\frac{\partial}{\partial v_4},$$

$$J_2 = v_4\frac{\partial}{\partial u_1} - v_3\frac{\partial}{\partial u_2} - v_2\frac{\partial}{\partial u_3} + v_1\frac{\partial}{\partial u_4} + u_4\frac{\partial}{\partial v_1} - u_3\frac{\partial}{\partial v_2} - u_2\frac{\partial}{\partial v_3} + u_1\frac{\partial}{\partial v_4},$$

$$S = \sum_{j=1}^{4}\left(x_j\frac{\partial}{\partial x_j} - \frac{u_j}{2}\frac{\partial}{\partial u_j} - \frac{v_j}{2}\frac{\partial}{\partial v_j}\right).$$

**Theorem 2:** The nonlinear Dirac equation (2) is invariant under the Lie point symmetry groups which are generated by the infinitesimal generators

$$X_1, X_2, X_3, X_4, R_{12}, R_{13}, R_{23}, L_{14}, L_{24}, L_{34}, J_0, J_1, J_2.$$

Consequently, if we introduce a mass term, then the invariance under the scale change $S$ ceases to exist.

In the following we consider a special case where $\psi_2 = 0$ and $\psi_3 = 0$. Moreover, we restrict ourselves to one space dimension. With this simplification Eq. (1) takes the form

$$\frac{\partial u_1}{\partial x_1} + \frac{\partial u_4}{\partial x_4} + \epsilon K v_4 = 0, \quad -\frac{\partial v_1}{\partial x_1} - \frac{\partial v_4}{\partial x_4} + \epsilon K u_4 = 0, \quad (4)$$

$$-\frac{\partial u_4}{\partial x_1} - \frac{\partial u_1}{\partial x_4} + \epsilon K v_1 = 0, \quad \frac{\partial v_4}{\partial x_1} + \frac{\partial v_1}{\partial x_4} + \epsilon K u_1 = 0,$$

where $K = u_1^2 + v_1^2 - u_4^2 - v_4^2$, $x_4 = ct$, and $\epsilon$ is a real parameter. Note that in one space dimension the quantity $l^2$ becomes a dimensionless parameter which we call $\epsilon$. The system of pde's (4) admits seven symmetry generators, namely $X_1$,$X_4$,$L_{14}$,$J_0$,$J_1$,$J_2$, and $S$ (restricted to the special case $\psi_2 = \psi_3 = 0$ and one space dimension). From Eq. (4) we find immediately the conservation law (charge)

$$\frac{\partial(u_1^2 + v_1^2 + u_4^2 + v_4^2)}{\partial x_4} + 2\frac{\partial(u_1u_4 + v_1v_4)}{\partial x_1} = 0. \quad (5)$$

## III. SYMMETRY GROUPS OF THE MASSIVE THIRRING MODEL

Let us now consider the one-dimensional massive Thirring model and Lie point symmetry groups. The massive Thirring model describes the relativistic two-dimensional massive spinor field with current–current interaction. Several authors[8–10] have studied the integrability of the massive Thirring model. They found that the massive Thirring model is integrable. This means, this system of pde's can be solved by IST. The Gelfand–Levitan integral equations appear with tedious nonlinearities. Let $u_1$,$u_2$,$v_1$, and $v_2$ be real fields. Then the massive Thirring model can be written as[8]

$$-\frac{\partial u_1}{\partial x_1} + \frac{\partial u_1}{\partial x_4} = 4v_2 - (u_2^2 + v_2^2)v_1, \quad (6a)$$

$$\frac{\partial v_1}{\partial x_1} - \frac{\partial v_1}{\partial x_4} = 4u_2 - (u_2^2 + v_2^2)u_1, \quad (6b)$$

$$\frac{\partial u_2}{\partial x_1} + \frac{\partial u_2}{\partial x_4} = 4v_1 - (u_1^2 + v_1^2)v_2, \tag{6c}$$

$$-\frac{\partial v_2}{\partial x_1} - \frac{\partial v_2}{\partial x_4} = 4u_1 - (u_1^2 + v_1^2)u_2. \tag{6d}$$

We mention that Eq. (6) can be derived from a Lagrangian. With simple algebraic manipulations we find from Eq. (6) the conservation law (charge)

$$\frac{\partial(u_1^2 + v_1^2 + u_2^2 + v_2^2)}{\partial x_4} + \frac{\partial(-u_1^2 - v_1^2\ u_2^2 + v_2^2)}{\partial x_1} = 0. \tag{7}$$

**Theorem 3:** The massive Thirring model (6) is invariant under the Lie point symmetry groups which are generated by the infinitesimal generators

$$X_1, X_4,$$

$$L_{14}^* = x_1\frac{\partial}{\partial x_4} + x_4\frac{\partial}{\partial x_1} - \frac{u_1}{2}\frac{\partial}{\partial u_1} + \frac{u_2}{2}\frac{\partial}{\partial u_2} - \frac{v_1}{2}\frac{\partial}{\partial v_1} + \frac{v_4}{2}\frac{\partial}{\partial v_4}, \tag{8}$$

$$J^* = v_1\frac{\partial}{\partial u_1} + v_2\frac{\partial}{\partial u_2} - u_1\frac{\partial}{\partial v_1} - u_2\frac{\partial}{\partial v_2}.$$

If the rest mass is equal to zero ($m = 0$), then Eq. (6) also admits the symmetry generator

$$S^* = x_1\frac{\partial}{\partial x_1} + x_4\frac{\partial}{\partial x_4} - \frac{u_1}{2}\frac{\partial}{\partial u_1} - \frac{u_2}{2}\frac{\partial}{\partial u_2} - \frac{v_1}{2}\frac{\partial}{\partial v_1} - \frac{v_2}{2}\frac{\partial}{\partial v_2}. \tag{9}$$

## IV. GROUP THEORETICAL REDUCTIONS

Given Lie point transformation groups which are admitted by a given system of pde's, there are standard procedures for finding the similarity ansatz and the system of ode's (see for example, Refs. 26–30).

Consider first the nonlinear Dirac equation (4). For reducing the system of pde's (4) we study three cases, namely reduction with the help of space-time translation $X_1 + X_4$, Lorentz transformation $L_{14}$, and scale change $S$.

The space-time translation leads to the similarity ansatz

$$u_1(x_1,x_4) = \bar{u}_1(\eta),\ldots, v_4(x_1,x_4) = \bar{v}_4(\eta), \tag{10}$$

where the similarity variable $\eta$ is given by $\eta = x_1 + x_4$. The resulting system of ode's is completely integerable. There is a sufficiently large number of first integrals.

The reduction with the Lorentz transformation $L_{14}$ leads to the similarity ansatz

$$\begin{aligned} u_1(x_1,x_4) &= [\cosh(\epsilon/2)]\bar{u}_1(\eta) + [\sinh(\epsilon/2)]\bar{u}_4(\eta),\\ u_4(x_1,x_4) &= [\cosh(\epsilon/2)]\bar{u}_4(\eta) + [\sinh(\epsilon/2)]\bar{u}_1(\eta),\\ v_1(x_1,x_4) &= [\cosh(\epsilon/2)]\bar{v}_1(\eta) + [\sinh(\epsilon/2)]\bar{v}_4(\eta),\\ v_4(x_1,x_4) &= [\cosh(\epsilon/2)]\bar{v}_4(\eta) + [\sinh(\epsilon/2)]\bar{v}_1(\eta), \end{aligned} \tag{11}$$

where

$$\epsilon = \operatorname{arctanh}(x_4/x_1), \tag{12}$$

and

$$\eta^2 = x_1^2 - x_4^2. \tag{13}$$

With this ansatz we obtain

$$\begin{aligned} &\bar{u}_1' + \bar{u}_1/(2\eta) + \epsilon K(\bar{u}_1, \ldots, \bar{v}_4) = 0,\\ &\bar{v}_1' + \bar{v}_1/(2\eta) - \epsilon K(\bar{u}_1, \ldots, \bar{v}_4) = 0,\\ &\bar{u}_4' + \bar{u}_4/(2\eta) - \epsilon K(\bar{u}_1, \ldots, \bar{v}_4) = 0,\\ &\bar{v}_4' + \bar{v}_4/(2\eta) + \epsilon K(\bar{u}_1, \ldots, \bar{v}_4) = 0, \end{aligned} \tag{14}$$

where $' = d/d\eta$. For this nonautonomous system of ode's we can give at once two first integrals, namely,

$$\begin{aligned} h_1(\eta,\bar{u}_1,\ldots, \bar{v}_4) &= \eta\bar{u}_1^2 - \eta\bar{v}_4^2,\\ h_2(\eta,\bar{u}_1, \ldots, \bar{v}_4) &= \eta\bar{v}_1^2 - \eta\bar{u}_4^2. \end{aligned} \tag{15}$$

As third example, we consider the reduction with the help of the scale change $S$. We find the similarity ansatz

$$u_1(x_1,x_4) = x_4^{-1/2}\,\bar{u}_1(\eta), \ldots, v_4(x_1,x_4) = x_4^{-1/2}\bar{v}_4(\eta), \tag{16}$$

where $\eta = x_1/x_4$. By straightforward calculation we find that

$$\begin{aligned} &\bar{v}_4' - \bar{v}_1/2 - \eta\bar{v}_2' + \epsilon K(\bar{u}_1, \ldots, \bar{v}_4)\bar{u}_1 = 0,\\ &-\bar{u}_4' + \bar{u}_1/2 + \eta\bar{u}_1' + \epsilon K(\bar{u}_1, \ldots, \bar{v}_4)\bar{v}_1 = 0,\\ &-\bar{v}_1' + \bar{v}_4/2 + \eta\bar{v}_4' + \epsilon K(\bar{u}_1, \ldots, \bar{v}_4)\bar{u}_4 = 0,\\ &\bar{u}_1' - \bar{u}_2/2 - \eta\bar{u}_4' + \epsilon K(\bar{u}_1, \ldots, \bar{v}_4)\bar{v}_4 = 0. \end{aligned} \tag{17}$$

Two first integrals can be given, namely

$$\begin{aligned} h_1(\bar{u}_1, \ldots, \bar{v}_4) &= K(\bar{u}_1, \ldots, \bar{v}_4) \equiv \bar{u}_1^2 + \bar{v}_1^2 - \bar{u}_4^2 - \bar{v}_4^2,\\ h_2(\eta,\bar{u}_1, \ldots, \bar{v}_4) &= \bar{u}_1^2 - \bar{v}_1^2 - \bar{u}_4^2 + \bar{v}_4^2 - 2\eta\bar{u}_1\bar{u}_4 + 2\eta\bar{v}_1\bar{v}_4. \end{aligned} \tag{18}$$

To summarize, we find that the group theoretical reduction leads to systems of ode's which are integrable. Therefore the result cannot help us to decide whether the nonlinear Dirac equation (4) is integrable or not.

When we consider the Thirring model (6) and group theoretical reduction with the help of the symmetry generators given by Eq. (8), we find the same result. In this case the result coincides with the fact that the Thirring model can be solved with the IST.

## V. COVARIANT EXTERIOR DERIVATIVE AND LIE BÄCKLUND VECTOR FIELDS

Now let us discuss the integrability of the nonlinear Dirac equation (4) and the massive Thirring model (6) from another point of view. As mentioned above the Thirring model can be solved with the help of IST, and a Bäcklund transformation and an infinite number of conservation laws have also been given. In the following we describe that the massive Thirring model can be written as covariant derivative of a Lie algebra valued differential form, and we also give a LB vector field. Motivated by this we discuss whether the nonlinear Dirac equation (4) can be written as covariant derivative and whether LB vector fields exist.

It is well known that the soliton equations like Korteweg–de Vries, sine–Gordon, modified Korteweg–de Vries, nonlinear Schrödinger, and Liouville can be written as covariant derivatives of Lie algebra valued differential forms, where the underlying Lie algebra is given by sl(2, $\mathbb{R}$). Notice that dim sl (2,$\mathbb{R}$) = 3. Consequently, the Thirring model cannot be represented within this Lie algebra. In order to represent the Thirring model we are forced to extend the Lie algebra sl(2,$\mathbb{R}$) to sl(2, $\mathbb{C}$), where dim sl(2, $\mathbb{C}$) = 6. A convenient choice of the basis of sl(2, $\mathbb{C}$) is given by

$$X_1=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad X_2=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix},\quad X_3=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix},$$
$$Y_1=\begin{pmatrix}i & 0\\ 0 & -i\end{pmatrix},\quad Y_2=\begin{pmatrix}0 & i\\ 0 & 0\end{pmatrix},\quad Y_3=\begin{pmatrix}0 & 0\\ i & 0\end{pmatrix}. \tag{19}$$

Consider the Lie algebra valued differential one-form

$$\Gamma=\sum_{i=1}^{3}(\alpha_i\otimes X_i+\beta_i\otimes Y_i), \tag{20}$$

where

$$\alpha_i=a_i(x_1,x_4)dx_1+A_i(x_1,x_4)dx_4,$$
$$\beta_i=b_i(x_1,x_4)dx_1+\beta_i(x_1,x_4)dx_4. \tag{21}$$

The covariant derivative of $\Gamma$ with respect to $\Gamma$ is given by $D_\Gamma\Gamma=d\Gamma+\frac{1}{2}[\Gamma,\Gamma]$. From the condition that $D_\Gamma\Gamma=0$ we find the system of pde's

$$-\frac{\partial a_1}{\partial x_4}+\frac{\partial A_1}{\partial x_1}+a_2A_3-a_3A_2-b_2B_3+b_3B_2=0, \tag{22a}$$

$$-\frac{\partial a_2}{\partial x_4}+\frac{\partial A_2}{\partial x_1}+2(a_1A_2-a_2A_1)-2(b_1B_2-b_2B_1)=0, \tag{22b}$$

$$-\frac{\partial a_3}{\partial x_4}+\frac{\partial A_3}{\partial x_1}-2(a_1A_3-a_3A_1)+2(b_1B_3-b_3B_1)=0, \tag{22c}$$

$$-\frac{\partial b_1}{\partial x_4}+\frac{\partial B_1}{\partial x_1}+a_2B_3-b_3A_2-a_3B_2+b_2A_3=0, \tag{22d}$$

$$-\frac{\partial b_2}{\partial x_4}+\frac{\partial B_2}{\partial x_1}+2(a_1B_2-b_2A_1)-2(a_2B_1-b_1A_2)=0, \tag{22e}$$

$$-\frac{\partial b_3}{\partial x_4}+\frac{\partial B_3}{\partial x_1}-2(a_1B_3-b_3A_1)+2(a_3B_1-b_1A_3)=0. \tag{22f}$$

By suitable choice of $a_1,\ldots,B_3$ we obtain Eq. (6). We choose

$$\begin{aligned}
&a_1=A_1=0,\\
&a_2=\lambda u_1+\lambda^{-1}u_2,\quad A_2=\lambda u_1-\lambda^{-1}u_2,\\
&b_2=\lambda v_1+\lambda^{-1}v_2,\quad B_2=\lambda v_1-\lambda^{-1}v_2,\\
&a_3=-\lambda u_1-\lambda^{-1}u_2,\quad A_3=-\lambda u_1+\lambda^{-1}u_2,\\
&b_3=\lambda v_1+\lambda^{-1}v_2,\quad B_3=\lambda v_1-\lambda^{-1}v_2,
\end{aligned} \tag{23}$$

and

$$b_1=\lambda^2-\lambda^{-2}-\tfrac{1}{4}(u_1^2+v_1^2-u_2^2-v_2^2),$$
$$B_1=\lambda^2+\lambda^{-2}-\tfrac{1}{4}(u_1^2+v_1^2+u_2^2+v_2^2). \tag{24}$$

Equation (22a) is satisfied identically and Eqs. (22b), (22c), (22e), and (22f) describe the Thirring model (6). Equation (22d) is given by

$$\frac{\partial(u_1^2+v_1^2-u_2^2-v_2^2)}{\partial x_4}-\frac{\partial(u_1^2+v_1^2+u_2^2+v_2^2)}{\partial x_1}-16(u_1v_2+u_2v_1)=0. \tag{25}$$

This equation can be obtained from Eq. (6) as follows. We multiply Eq. (6a) by $u_1$ and Eq. (6b) by $v_1$ and subtract. It follows that

$$\frac{\partial(u_1^2+v_1^2)}{\partial x_4}-\frac{\partial(u_1^2+v_1^2)}{\partial x_1}+8(-u_1v_2+u_2v_1)=0. \tag{26}$$

From Eq. (6c) and Eq. (6d) we obtain

$$\frac{\partial(u_2^2+v_2^2)}{\partial x_4}+\frac{\partial(u_2^2+v_2^2)}{\partial x_1}+8(u_1v_2-v_1u_2)=0. \tag{27}$$

When we add Eq. (24) and Eq. (25) we obtain the conservation law given by Eq. (7). When we subtract Eq. (25) from Eq. (24), Eq. (23) results.

From the above we are motivated to look for a possible choice of $a_1,\ldots,B_3$ in order to satisfy the nonlinear Dirac equation (4). For example, inserting the ansatz

$$\begin{aligned}
&a_1=c_{11}\lambda u_1+c_{12}\lambda^{-1}v_1+c_{13}\lambda u_4+c_{14}\lambda^{-1}v_4,\\
&A_1=c_{21}\lambda u_1+c_{22}\lambda^{-1}v_1+c_{23}\lambda u_4+c_{24}\lambda^{-1}v_4,
\end{aligned} \tag{28}$$

$$\begin{aligned}
&\cdot\\ &\cdot\\ &\cdot\\
&A_3=c_{61}\lambda u_1+c_{62}\lambda^{-1}v_1+c_{63}\lambda u_4+c_{64}\lambda^{-1}v_4,\\
&b_2=c_{71}\lambda u_1+c_{72}\lambda^{-1}v_1+c_{73}\lambda u_4+c_{74}\lambda^{-1}v_4,\\
&\cdot\\ &\cdot\\ &\cdot\\
&B_3=c_{101}\lambda u_1+c_{102}\lambda^{-1}v_1+c_{103}\lambda u_4+c_{104}\lambda^{-1}v_4,
\end{aligned} \tag{29}$$

and

$$b_1=k_1\lambda^2+k_2\lambda^{-2}+k_3(u_1u_4+v_1v_4),$$
$$B_1=k_4\lambda^2+k_5\lambda^{-2}+k_6(u_1^2+v_1^2+u_4^2+v_4^2)$$

into Eq. (20) we find that the nonlinear Dirac equation cannot be represented. The equations for the coefficients $c_{11},\ldots,k_6$ cannot be satisfied.

Let us now discuss the existence of LB vector fields for the Thirring model (6) and the nonlinear Dirac equation (4). We adopt the jet bundle technique.[31] Within this approach we consider the local coordinates $(x,t,u_1,\ldots,v_2,u_{11},u_{14},\ldots)$ and instead of Eq. (6) the submanifolds

$$\begin{aligned}
&F_1\equiv-u_{11}+u_{14}-4v_2+(u_2^2+v_2^2)v_1=0,\\
&F_2\equiv v_{11}-v_{14}-4u_2+(u_2^2+v_2^2)u_1=0,\\
&F_3\equiv u_{21}+u_{24}-4v_1+(u_1^2+v_1^2)v_2=0,\\
&F_4\equiv-v_{21}-v_{24}-4u_1+(u_1^2+v_1^2)u_2=0,
\end{aligned} \tag{30}$$

and all its differential consequences with respect to the space coordinate $x$. Let

$$V=f_1(u_1,\ldots,v_{211})\frac{\partial}{\partial u_1}+f_2(u_1,\ldots,v_{211})\frac{\partial}{\partial v_1}+f_3(u_1,\ldots,v_{211})\frac{\partial}{\partial u_2}+f_4(u_1,\ldots,v_{211})\frac{\partial}{\partial v_2} \tag{31}$$

be a LB vector field. Due to the structure of Eq. (6) we can simplify without loss of generality the vector field $V$, namely

$$f_i(u_1, ..., v_{211}) = f_{i1}(u_{11}, ..., v_{21}) + f_{i2}(u_1^3 u_{11}, u_1^2 v_1 u_{11}, ..., v_2^3 v_{21}) + f_{i3}(u_{111}, ..., v_{211}),$$

where $f_{i1}$ and $f_{i3}$ are linear functions. The function $f_{i2}$ is linear with respect to the arguments $u_1^3 u_{11}$, $u_1^2 u_2 u_{11}, ...,$ $v_2^3 v_{21}$. From the requirement that $L_{\bar{V}} F_i \hat{=} 0$, where $\bar{V}$ is the extended vector field of $V$, $L_{\bar{V}}(\cdot)$ denotes the Lie derivative and $\hat{=}$ stands for the restriction to solutions of Eq. (6), we find the vector field where $f_{i3} \neq 0$ (for further details of this technique see, for example, Ref. 32). Thus the Thirring model a LB vector field exists. Furthermore, there is a hierarchy of LB vector fields. This coincides with the fact that the Thirring model can be solved within IST.

If we consider the vector field (31) and the nonlinear Dirac equation (4) (substitute $u_2 \rightarrow u_4$, $v_2 \rightarrow v_4$), then we find that the Dirac equation does not admit a LB vector field of the form given by Eq. (31).

## VI. CONCLUSION

The group theoretical reduction of the nonlinear Dirac equation does not give a decision whether or not Eq. (4) is integrable, since the resulting ode's are integrable. Also the group theoretical reduction of the Thirring model leads to integrable ode's. From further investigations (existence of LB vector fields and representation as a covariant exterior derivative) we conclude that the nonlinear Dirac equation is not integrable. Alvarez and Carreras[21] studied Eq. (4) numerically including a mass term. They observed different types of interactions and bound state formations and conclude that this system is not integrable.

Recently, Weiss *et al.*[33] have introduced what is called the Painlevé property for pde's. Meanwhile various[34–38] authors have applied this approach. It would be interesting to study the pde's given above from this point of view.


[1] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[2] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 1006 (1980).
[3] R. S. Johnson, Phys. Lett. A **72**, 197 (1979).
[4] H. Flaschka, J. Math. Phys. **21**, 1016 (1980).
[5] A. Ramani, in *Fourth International Conference on Collective Phenomena*, (New York Academy of Sciences, New York, 1981), Vol. 373.
[6] R. Kaliappan and M. Lakshamanan, J. Phys. A: Math. Gen. **12**, L249 (1978).
[7] Y. F. Chang, M. Tabor, and J. Weiss, J. Math. Phys. **23**, 531 (1982).
[8] D. J. Kaup and A. C. Newell, Nuovo Cimento Lett. **20**, 325 (1977).
[9] T. Kawata, T. Morishima, and H. Inoue, J. Phys. Soc. Jpn. **47**, 1327 (1979).
[10] A. V. Mikhailov, JETP Lett. **23**, 320 (1976).
[11] S. Y. Lee, T. K. Kuo, and A. Gavrielides, Phys. Rev. D **12**, 2249 (1975).
[12] F. Gürsey, Nuovo Cimento **3**, 988 (1956).
[13] M. Soler, Phys. Rev. D **1**, 2766 (1970).
[14] T. Takahashi, J. Math. Phys. **20**, 1232 (1979).
[15] J. Werle, Phys. Lett. B **71**, 357 (1977).
[16] T. P. Branson and Y. Kosmann-Schwarzbach, Lett. Math. Phys. **7**, 63 (1983).
[17] W-H. Steeb, W. Erig, and W. Strampp, J. Math. Phys. **23**, 145 (1982).
[18] A. F. Rañada and L. Vazquez, Prog. Theor. Phys. **56**, 311 (1976).
[19] L. Garcia and A. F. Rañada, Prog. Theor. Phys. **64**, 671 (1980).
[20] P. Kersten, Memorandum No. 398, Twente University of Technology, Department of Applied Mathematics, Enschede, The Netherlands (1982); J. Math. Phys. **24**, 2374 (1983).
[21] A. Alvarez and B. Carreras, Phys. Lett. A **86**, 327 (1981).
[22] A. Galindo, Nuovo Cimento Lett. **20**, 210 (1977).
[23] L. Vazques, J. Phys. A: Math. Gen. **10**, 1361 (1977).
[24] A. F. Rañada, in *Old and New Questions in Physics, Cosmology, Philosophy and Theoretical Biology: Essays in Honor of W. Yourgrau*, edited by A. van der Merwe (Plenum, New York, 1983).
[25] W-H. Steeb, J. Math. Phys. **21**, 1656 (1980).
[26] G. Bluman and J. Cole, *Similarity Methods for Differential Equations* (Springer, New York, 1974).
[27] B. K. Harrison and F. B. Estabrook, J. Math. Phys. 12, 653 (1971).
[28] W. Strampp, W-H. Steeb, and W. Erig, Prog. Theor. Phys. **68**, 731 (1982).
[29] T. Branson and W-H. Steeb, J. Phys. A: Math. Gen. **16**, 469 (1983).
[30] W-H. Steeb and W. Oevel, Z. Naturforsch. **38a**, 86 (1983).
[31] W-H. Steeb and W. Strampp, Physica. A **114**, 95 (1982).
[32] W-H. Steeb and W. Oevel, Phys. Scr. **28**, 513 (1983).
[33] J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1982).
[34] J. Weiss, J. Math. Phys. **24**, 1405 (1983).
[35] W-H. Steeb, M. Kloke, B. M. Spieker, and W. Oevel, J. Phys. A: Math. Gen. **16**, L447 (1983).
[36] D. V. Chudnovsky, G. V. Chudnovsky, and M. Tabor, Phys. Lett. **97**, 268 (1983).
[37] W-H. Steeb, M. Kloke, and B. M. Spieker, Z. Naturforsch. **38a**, 1054 (1983).
[38] W-H. Steeb and W. Strampp, *XII International Colloquium on Groups Theoretical Methods in Physics; Lecture Notes in Physics* (Springer, Berlin, 1984).

# Critical properties of pseudospin Hamiltonians

R. Gilmore
*Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104*

(Received 18 January 1984; accepted for publication 2 March 1984)

The thermodynamic critical properties of a simple class of pseudospin model Hamiltonians are discussed. This class of models includes the spin van der Waals model and the Meshkov–Glick–Lipkin model as particular cases. Second-order thermodynamic phase transitions occur when the spin–spin interaction contributes negatively in a particular direction and the linear interaction term is orthogonal to the direction(s) of greatest energy gain through the spin–spin interaction.

PACS numbers: 75.10. – b, 75.10.Dg, 05.70.Jk

## I. INTRODUCTION

The critical properties of the spin van der Waals model have recently been studied by Lee.[1] In that analysis the $XY$-like regime and the Ising-like regime were studied separately.

In the present work a generalization of the spin van der Waals model is studied. A simple algorithm is applied to this model to determine both the ground state critical properties and the thermodynamic critical properties. We also study the effects of perturbations on the persistence of the second-order phase transition, if one is present before the perturbation is applied.

## II. THE MODEL

The spin van der Waals model is a mean-field model describing the interactions among a large number of identical particles. The $\alpha$th particle is assumed to have (pseudo) spin $S$ with components $S_i^{(\alpha)}(i = x, y, \text{or } z)$. The spin van der Waals model Hamiltonian can be expressed in terms of the total (pseudo) spin operators

$$J_i = \sum_{\alpha=1}^{N} S_i^{(\alpha)}. \tag{2.1}$$

A convenient generalization of the spin van der Waals model is defined by the Hamiltonian

$$\mathscr{H} = \frac{1}{N}\sum_{i,j} J_i Q_{ij} J_j, \tag{2.2}$$

where $Q$ is a real symmetric $3\times 3$ matrix. Specific choices of the matrix elements $Q_{ij}$ lead to the Ising-like and the $XY$-like regimes of the previously studied model.[2–4]

## III. GROUND STATE CRITICAL PROPERTIES

The Hamiltonian (2.2) does not exhibit a phase transition for finite $N^1$. We therefore consider the thermodynamic limit ($N\rightarrow\infty$) of (2.2). In this limit, the critical properties of $\mathscr{H}$ are determined by a simple algorithm.[5,6]

(1) Convert the Hamiltonian to "intensive" form:

$$\frac{\mathscr{H}}{N} = \sum_{i,j} Q_{ij}\left(\frac{J_i}{N}\right)\left(\frac{J_j}{N}\right).$$

(2) Replace the intensive operators $J_i/N$ by

$$J_1/N \rightarrow r \sin\theta\cos\phi,$$

$$J_2/N \rightarrow r \sin\theta\sin\phi, \quad 0\leqslant r\leqslant\tfrac{1}{2}$$

$$J_3/N \rightarrow r\cos\theta.$$

(3) Minimize the resulting function $\langle \mathscr{H}/N\rangle = h$ over the state variables $(r, \theta, \phi)$.

To apply this algorithm to the Hamiltonian (2.2), we let $\hat{\mathbf{n}}$ be a unit vector in the $(\theta, \phi)$ direction. Then according to the algorithm

$$\mathscr{H}/N \rightarrow h = r^2\, \hat{\mathbf{n}}\cdot Q\cdot\hat{\mathbf{n}}. \tag{3.1}$$

Let the eigenvalues $\lambda_i$ of the matrix $Q$ obey

$$\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3. \tag{3.2}$$

If $\lambda_1 > 0$, the minimum value of $h$ is attained for $r = 0$. If $\lambda_1 < 0$, the minimum value of $h$ is obtained for $r = \frac{1}{2}$ and $\hat{n}$ an eigenvector of $Q$ to eigenvalue $\lambda_1$:

$$\min_{(r\theta\phi)} h = (\tfrac{1}{2})^2\lambda_1. \tag{3.3}$$

The expectation values of the intensive operators $\mathbf{J}/N$ are given by

$$\langle \mathbf{J}/N\rangle = r\,\hat{\mathbf{n}}, \tag{3.4}$$

where $r = 0$ if $\lambda_1 > 0$ and $r = \frac{1}{2}$ if $\lambda_1 < 0$.

## IV. THERMODYMANIC CRITICAL PROPERTIES

The thermodynamic critical properties of (2.2) are also determined by a simple algorithm.[5,6]

(1) The free energy per particle is determined by subtracting the entropy term from the energy term

$$\langle F/N\rangle = \langle \mathscr{H}/N\rangle - kTs(r).$$

(2) The entropy term is an SU(2) multiplicity factor[7]

$$s(r) = -(\tfrac{1}{2}+r)\ln(\tfrac{1}{2}+r) - (\tfrac{1}{2}-r)\ln(\tfrac{1}{2}-r).$$

(3) Minimize the resulting function, $\langle F/N\rangle = f$, over the state variables $(r,\theta,\phi)$.

To apply this algorithm to the Hamiltonian (2.2), we again assume $\lambda_1$ is the minimum eigenvalue of $Q$. Then

$$\langle F/N\rangle = r^2\lambda_1 + kT\{(\tfrac{1}{2}+r)\ln(\tfrac{1}{2}+r) + (\tfrac{1}{2}-r)\ln(\tfrac{1}{2}-r)\}.$$

If $\lambda_1 > 0$ the minimum value of $\langle F/N\rangle$ occurs for $r = 0$ at all temperatures.

If $\lambda_1 < 0$ a second-order thermodynamic phase transition will occur. The relationship between the state variable $r(r = \langle J^2\rangle^{1/2}/N)$ and the temperature $T$ is determined through the minimization condition

$$\frac{d}{dr}\left\langle\frac{F}{N}\right\rangle = 2r\lambda_1 + kT\left\{\ln\left(\frac{1}{2}+r\right) - \ln\left(\frac{1}{2}-r\right)\right\}. \tag{4.1}$$

The critical temperature $T_c$ is determined by the vanishing of the second-degree Taylor series coefficient, as is usual for

 0022-2488/84/072336-03$02.50  

a second-order Ginzburg–Landau phase transition:

$$\frac{d^2}{dr^2}\left\langle\frac{F}{N}\right\rangle = 2\lambda_1 + kT\left\{\left(\frac{1}{2}+r\right)^{-1} + \left(\frac{1}{2}-r\right)^{-1}\right\}. \tag{4.2}$$

From (4.2) we determine

$$-\lambda_1/2 = kT_c. \tag{4.3}$$

This relation between the coupling strength (i.e., the ground state energy per particle $\lambda_1$) and the critical temperature can be used to write (4.1) in a simple scaled form

$$t = \frac{T}{T_c} = \frac{4r}{\ln[(1+2r)/(1-2r)]}, \quad 0 \leq t \leq 1, \quad \tfrac{1}{2} \geq r \geq 0, \tag{4.4}$$

and $r = 0$ for $T > T_c$. The relation between the reduced temperature $t$ and the rms expectation value of the (pseudo) angular momentum $r = \langle J^2\rangle^{1/2}/N$ is shown in Fig. 1. At any temperature the expectation values of the angular momentum operators are given by

$$\langle \mathbf{J}/N\rangle = r(T)\hat{\mathbf{n}}, \tag{4.5}$$

where $\hat{\mathbf{n}}$ is the unit eigenvector of $Q$ to minimum eigenvalue.

## V. PERTURBATIONS

If the model Hamiltonian (2.2) exhibits a second-order thermodynamic phase transition, then a perturbation may or may not destroy this phase transition. To determine the conditions under which the phase transition either persists or is unhinged, we consider perturbations which possess only linear and quadratic terms in the total (pseudo) angular momentum operators $\mathbf{J}$. The perturbed Hamiltonian has the form

$$\mathscr{H}_p = \mathbf{L}\cdot\mathbf{J} + (1/N)\mathbf{J}\cdot Q'\cdot\mathbf{J}. \tag{5.1}$$

The structural stability of the phase transition is determined by a simple algorithm.

(1) Choose as coordinate axes the eigenvectors $\hat{\mathbf{n}}_1$, $\hat{\mathbf{n}}_2$, $\hat{\mathbf{n}}_3$ of $Q'$, with eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$.

(2) Resolve the linear perturbation into components $L_1$, $L_2$, $L_3$ along the three coordinate directions.

(3) If $\mathbf{L}$ has a component in the subspace spanned by eigenvectors with minimum eigenvalue ($L_1 \neq 0$ if $\lambda_1 < \lambda_2$; $L_1$ or $L_2 \neq 0$ if $\lambda_1 = \lambda_2 < \lambda_3$; $L \neq 0$ if $\lambda_1 = \lambda_2 = \lambda_3$), the second-order phase transition is unhinged; otherwise it will persist if $\lambda_1 < 0$.

In the generic case[6] that the phase transition is unhinged, the state variables $\langle \mathbf{J}/N\rangle = r\hat{\mathbf{n}}$ are determined by minimizing the free energy expression

$$\langle F/N\rangle = r^2\hat{\mathbf{n}}\cdot Q'\cdot\hat{\mathbf{n}} + r\hat{\mathbf{n}}\cdot\mathbf{L} - kTs(r). \tag{5.2}$$

The unit vector $\hat{\mathbf{n}}$ is determined by introducing a Lagrange multiplier $\gamma$ through $-\gamma(\hat{\mathbf{n}}\cdot\hat{\mathbf{n}} - 1)$. We find

$$n_i(r) = rL_i/2[\gamma(r) - r^2\lambda_i], \tag{5.3}$$

where $\gamma(r)$ is determined by the constraint

$$\left(\frac{r}{2}\right)^2 \sum_{i=1}^{3} \frac{L_i^2}{[\gamma(r) - r^2\lambda_i]^2} = 1. \tag{5.4}$$

The smallest of the (up to) six values of $\gamma$ which satisfy (5.4) is used in (5.3). The state variable $r$ is related to the temperature $T$ through

$$kT = -2\gamma(r)/r \ln[(1+2r)/(1-2r)].$$

The ranges of $r$ and $T$ are related by $0 < r_{\min} \leq r \leq r_{\max} \leq \frac{1}{2}$ and $\infty \geq T \geq 0$.

The critical properties can readily be determined in the nongeneric case in which the second-order phase transition is not destroyed. To illustrate, we consider the case in which the minimum eigenvalue of $Q$ is nondegenerate ($\lambda_1 < \lambda_2 \leq \lambda_3$, $\lambda_1 < 0$) and choose $L = (0, L_2, L_3)$. The alternative possibility ($\lambda_1 = \lambda_2 < \lambda_3$) and $L = (0, 0, L_3)$ is treated similarly. An easy calculation shows that the components of the unit vector $\hat{\mathbf{n}}$ minimizing $\langle \mathscr{H}/N\rangle$ obey

$$n_j = -L_j/2r\Delta_j, \quad \Delta_j = \lambda_j - \lambda_1, \quad j = 2,3, \tag{5.5}$$

provided that $n_2^2 + n_3^2 < 1$. Since the maximum value of $r$ is $\frac{1}{2}$, we see that the second-order phase transition persists for $L$ sufficiently small,

$$(L_2/\Delta_2)^2 + (L_3/\Delta_3)^2 < 1, \tag{5.6}$$

but is destroyed by sufficiently strong linear "perturbations": $(L_2/\Delta_2)^2 + (L_3/\Delta_3)^2 > 1$.

For small linear perturbations the critical temperature $T_c$ is determined by

$$2r_c(-\lambda_1) = kT_c \ln[(1+2r_c)/(1-2r_c)], \tag{5.7}$$

where

$$2r_c = [(L_2/\Delta_2)^2 + (L_3/\Delta_3)^2]^{1/2}. \tag{5.8}$$

The condition defining the critical temperature (5.7) can be written in the more familiar gap-equation form

$$2r_c = \tanh \tfrac{1}{2}\beta_c(-\lambda_1)(2r_c). \tag{5.9}$$

For $T < T_c$, $r_c < r \leq \frac{1}{2}$, the values of $\langle \mathbf{J}/N\rangle = r(T)\hat{\mathbf{n}}$ are determined by (5.5), together with the relation $n_1 = \pm[1 - n_2^2 - n_3^2]^{1/2}$ and the condition defining $r$:

$$2r\lambda_1 + kT\ln[(1+2r)/(1-2r)] = 0. \tag{5.10}$$

For $r < r_c$, $T > T_c$, $n_1 = 0$, $n_j = \alpha^{-1}L_j/\Delta_j$, $\alpha = [(L_2/\Delta_2)^2 + (L_3/\Delta_3)^2]^{1/2}$, and $r$ and $T$ are related by

$$2r\alpha^{-2}\sum_{j=2}^{3}\lambda_j\left(\frac{L_j}{\Delta_j}\right)^2 - \alpha^{-1}\sum_{j=2}^{3}\left(\frac{L_j^2}{\Delta_j}\right) + kT\ln\frac{1+2r}{1-2r} = 0. \tag{5.11}$$

The particular pseudospin model Hamiltonian for which $Q$

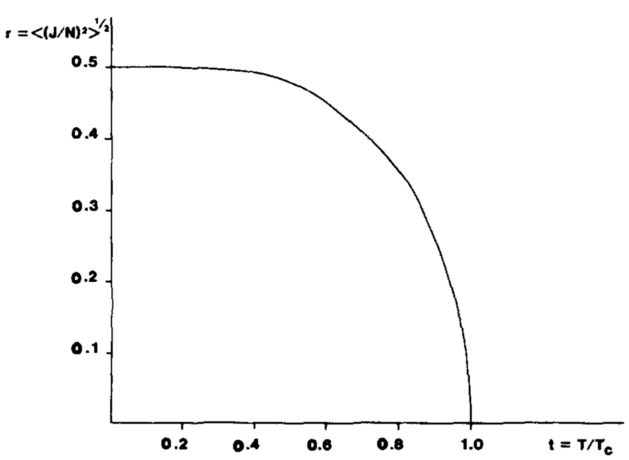


FIG. 1. The coupling constants and critical properties of the general spin van der Waals model are related by a simple scaled curve.

has eigenvalues $(-|V|, 0, 0)$ and $L = (0, 0, \epsilon)$ corresponds to the Meshkov–Glick–Lipkin model Hamiltonian[8]

$$\mathscr{H} = \epsilon J_z - (|V|/N)J_x^2 \qquad (5.12)$$

widely studied in nuclear physics.[6] For this Hamiltonian, many of the results derived above are well known.

The ground state energy phase transition for this model is usually studied as a function of increasing value of the normalized quadrupole interaction strength, $|V|/\epsilon$ (see Ref. 9). There is a second-order ground state energy phase transition at $|V|/\epsilon = 1$, by (5.6). For $|V| \leqslant \epsilon$, in the ground state $\langle \mathbf{J}/N \rangle = \frac{1}{2}(0, 0, -1)$ by (5.6). For $|V| \geqslant \epsilon$, we have $\langle \mathbf{J}/N \rangle = \frac{1}{2}(\sqrt{1-(\epsilon/|V|)^2},\ 0,\ -\epsilon/|V|)$ by (5.5).

For $|V|/\epsilon > 1$, this model exhibits a second-order thermodynamic phase transition[10] at $2r_c = \epsilon/|V|$ by (5.8). The corresponding critical temperature is determined from the gap equation $(|V|/\epsilon)\tanh\frac{1}{2}\beta_c\epsilon = 1$, which is a direct consequence of (5.9). In the ordered state below the phase transition we have $T < T_c$, $r > r_c$, $r$ and $T$ are related by $kT\ln[(1+2r)/(1-2r)] = 2r|V|$, and $\langle \mathbf{J}/N \rangle = r(\sqrt{1-(\epsilon/2rV)^2},\ 0,\ -\epsilon/2r|V|)$ by (5.10). In the disordered state above the phase transition, we have $T > T_c$, $r < r_c$, $r$ and $T$ are related by $kT\ln[(1+2r)/(1-2r)] = \epsilon$, and $\langle \mathbf{J}/N \rangle = r(0,0,-1)$ by (5.11). In addition, it is known[5,11] that the second-order phase transition (ground state energy or thermodynamic) is destroyed by addition of either $J_x$ or $J_z^2$ to the Hamiltonian (5.12). This result also follows directly from the algorithm presented in this section.

## VI. SUMMARY

Critical properties of pseudospin models which are generalizations of the spin van der Waals model and the Meshkov–Glick–Lipkin models have been studied. These models are general superpositions of terms linear and quadratic in the pseudospin operators. The conditions for the occurrence of a second-order thermodynamic phase transition have been determined. The structural stability of these transitions has also been discussed.


## ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant PHY 830-4868.



[1]M. H. Lee, J. Math Phys. **23**, 464 (1982).
[2]R. Dekeyser and M. H. Lee, Phys. Rev. B **19**, 265 (1979).
[3]I. M. Kim and M. H. Lee, Phys. Rev. B **24**, 3961 (1981).
[4]M. H. Lee, Phys. Rev. B **26**, 2547 (1982).
[5]R. Gilmore, J. Math. Phys. **18**, 17 (1977).
[6]R. Gilmore, *Catastrophe Theory for Scientists and Engineers* (Wiley, New York, 1981).
[7]F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[8]H. J. Lipkin, N. Meshkov, and A. J. Glick, Nucl. Phys. **62**, 188 (1965).
[9]R. Gilmore and D. H. Feng, Phys. Lett. B **76**, 26 (1978).
[10]R. Gilmore and D. H. Feng, Nucl. Phys. A **301**, 189 (1978).
[11]R. Gilmore, Phys. Lett. A **55**, 459 (1976).

# Crystal field effect in an Ising model

Ibha Chatterjee
*Saha Institute of Nuclear Physics, 92 Acharya Prafulla Chandra Road, Calcutta 700 009, India*

(Received 22 March 1983; accepted for publication 18 November 1983)

The crystal field effect has been studied in the linear Ising model which can be solved exactly. For spins $S = \frac{1}{2}, 1, \frac{3}{2}$ it is easy to diagonalize the transfer matrix analytically, but for spins $S > \frac{3}{2}$, the transfer matrices are diagonalized numerically. The numerical results are accurate to seven decimal places and can be treated as exact for all practical purposes. Crystal field has no effect on spin-$\frac{1}{2}$ systems, its effect has been studied in systems with spins $S > \frac{1}{2}$. The ferromagnetic as well as the antiferromagnetic sysceptibilities have been computed for the systems with and without the crystal field. It has been found that, for small crystal fields, the susceptibility behavior is not much different from that in the absence of crystal field. But for large crystal fields, not only the antiferromagnetic susceptibilities but the ferromagnetic susceptibilities also start showing maxima which appear for integer spins only.

PACS numbers: 75.10.Hk, 75.30.Cr, 71.70.Ch

## I. INTRODUCTION

The one-dimensional Ising model with spin $\frac{1}{2}$ in the absence of crystal field was solved by Ising[1] as early as 1925 using a combinatorial method. Later Kramers and Wannier[2] and Kubo[3] solved the same problem using a matrix method. The one-dimensional Ising model with general spin was solved by Suzuki, Tsujiyama, and Katsura[4] in the year 1967. First, they developed a perturbation method and then demonstrated an implicit differentiation method. They obtained exact solutions for $S = \frac{3}{2}$ and $S = 1$ and numerical results for these two spin systems were compared with the $S = \frac{1}{2}$ system. Finally, they concluded that both the perturbation and differentiation methods could be applied for the problem of general spin. In their works there was no mention of the crystal field which is important in the case of solids. Nobody has solved the Ising model in the presence of the crystal field until very recently when Lines[5] solved this problem for $S = 1$. Lines solved this problem exactly for the comparison of correlated effective field results with the exact results.

In the present paper the Ising model with general spin and in presence of a crystal field has been solved exactly. The crystal field effect has been studied on the susceptibility only. In order to calculate the susceptibility, first the transfer matrix is constructed and then diagonalized to obtain the eigenvalues and eigenfunctions of this matrix. Using these eigenvalues and eigenfunctions, correlation functions and susceptibilities (both ferromagnetic and antiferromagnetic) are calculated. The susceptibilities in the absence of crystal field are also calculated for comparison. The transfer matrices for $S = \frac{1}{2}, 1, \frac{3}{2}$ can easily be diagonalized analytically and those for $S > \frac{3}{2}$ cannot be diagonalized analytically very easily and, therefore, these are diagonalized numerically. The method of diagonalization is due to Jacobi. The exact susceptibility for any spin is calculated by writing a FORTRAN program where only the transfer matrix is supplied and the rest of the calculation is performed numerically.

## II. THEORY

The Hamiltonian for the one-dimensional Ising problem for an $N$-spin ring in the presence of an axial crystal field is given by

$$\mathscr{H} = \sum_{n=1}^{N} [D(S_n^Z)^2 - 2JS_n^Z S_{n+1}^Z]. \tag{1}$$

Since $S_{n+1}^Z \equiv S_1^Z$, regrouping the terms in the form

$$\mathscr{H} = \sum_{n=1}^{N} \{\tfrac{1}{2}D[(S_n^Z)^2 + (S_{n+1}^Z)^2] - 2JS_n^Z S_{n+1}^Z\}, \tag{2}$$

one notes that the partition function $Z$ can be expressed in $S^z$ representation as

$$Z = \mathrm{Tr} \prod_{n=1}^{N} T_n = \mathrm{Tr}\, T^N, \tag{3}$$

where all the transfer matrices $T_n$ have an identical Hermitian form. The transfer matrices are different for different spin systems. The first term in the Hamiltonian [Eq. (1)] introduces an axial crystal field anisotropy in the system, and for $S = \frac{1}{2}$ this term is a constant and therefore has no effect. The axial crystal field starts showing its effect for $S > \frac{1}{2}$. Since all the transfer matrices have an identical Hermitian form

$$T_n = T = e^{-\beta\mathscr{H}}, \tag{4}$$

where $\beta = 1/kT$ and the matrix elements are given by

$$\langle S|T|S'\rangle = e^{2\beta JSS' - (\beta D/2)[(S)^2 + (S')^2]}, \tag{5}$$

where $S$ and $S'$ are the projections of spin **S**.

As an example, the transfer matrix for $S = 1$ is given by

$$T = \begin{array}{c|ccc} S \backslash S' & 1 & 0 & -1 \\ \hline 1 & e^{-\beta(D-2J)} & e^{-\beta D/2} & e^{-\beta(D+2J)} \\ 0 & e^{-\beta D/2} & 1 & e^{-\beta D/2} \\ -1 & e^{-\beta(D+2J)} & e^{-\beta D/2} & e^{-\beta(D-2J)} \end{array}. \tag{6}$$

 0022-2488/84/072339-06$02.50  

Similarly transfer matrices for any spin can be constructed, and these matrices can be diagonalized analytically for $S=\frac{1}{2}, 1, \frac{3}{2}$. For $S>\frac{3}{2}$, diagonalization is performed numerically. Once eigenvalues and eigenfunctions of the transfer matrices are known, the susceptibility is calculated from the correlation function[6] as follows:

$$\chi=\frac{Ng^2\mu_B^2}{kT}\sum_{l=-\infty}^{+\infty}\langle S_K S_{K+l}\rangle, \tag{7}$$

where

$$\langle S_K S_{K+l}\rangle=\frac{1}{Z}\sum_{i,j=1}^{n}\lambda_i^l\lambda_j^{N-l}\langle\psi_i|\mathbf{S}|\psi_j\rangle\langle\psi_j|\mathbf{S}|\psi_i\rangle. \tag{8}$$

$n$ is the dimension of the transfer matrix. $\lambda_i, \lambda_j$ and $\psi_i, \psi_j$ are the eigenvalues and eigenfunctions of the transfer matrix. For $S=1$, **S** has the form

$$S=\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & -1\end{pmatrix}. \tag{9}$$

By solving the transfer matrix [Eq. (6)], eigenvalues and eigenfunctions are obtained as given below.

*Eigenvalues*:

$$\lambda_1=\tfrac{1}{2}[2e^{-\beta D}\cosh(2\beta J)+1+(\{2e^{-\beta D}\cosh(2\beta J)-1\}^2+8e^{-\beta D})^{1/2}],$$
$$\lambda_2=\tfrac{1}{2}[2e^{-\beta D}\cosh(2\beta J)+1-(\{2e^{-\beta D}\cosh(2\beta J)-1\}^2+8e^{-\beta D})^{1/2}], \tag{10}$$
$$\lambda_3=2e^{-\beta D}\sinh(2\beta J).$$

*Eigenvectors*:

$$|\psi_1\rangle=\frac{1}{(2+\alpha_1^2)^{1/2}}\begin{pmatrix}1\\ \alpha_1\\ 1\end{pmatrix},$$

where $\alpha_1=e^{\beta D}(\lambda_1-2e^{-\beta D}\cosh(2\beta J))$,

$$|\psi_2\rangle=\frac{1}{(2+\alpha_2^2)^{1/2}}\begin{pmatrix}1\\ \alpha_2\\ 1\end{pmatrix},$$

where $\alpha_2=e^{\beta D}(\lambda_2-2e^{-\beta D}\cosh(2\beta J))$, (11)

$$|\psi_3\rangle=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 0\\ -1\end{pmatrix}.$$

It is obvious that $\lambda_1>\lambda_2$. Again,

$$\begin{aligned}\lambda_1&>\tfrac{1}{2}[2e^{-\beta D}\cosh(2\beta J)+1+2e^{-\beta D}\cosh(2\beta J)-1]\\ &>2e^{-\beta D}\cosh(2\beta J)\\ &>2e^{-\beta D}\sinh(2\beta J)\\ &>\lambda_3.\end{aligned}$$

Thus we see $\lambda_1$ is the largest eigenvalue. The partition function is given by

$$Z=\mathrm{Tr}\,T^N=\lambda_1^N+\lambda_2^N+\lambda_3^N. \tag{12}$$

For large $N$, $Z\to\lambda_1^N$.

Let us find out the matrix elements of the spin operator **S**:

$$\begin{aligned}\langle\psi_1|\mathbf{S}|\psi_1\rangle&=\langle\psi_2|\mathbf{S}|\psi_2\rangle=\langle\psi_3|\mathbf{S}|\psi_3\rangle\\ &=\langle\psi_1|\mathbf{S}|\psi_2\rangle=\langle\psi_2|\mathbf{S}|\psi_1\rangle=0,\\ \langle\psi_3|\mathbf{S}|\psi_1\rangle&=\langle\psi_1|\mathbf{S}|\psi_3\rangle=[2/(2+\alpha_1^2)]^{1/2},\\ \langle\psi_3|\mathbf{S}|\psi_2\rangle&=\langle\psi_2|\mathbf{S}|\psi_3\rangle=[2/(2+\alpha_2^2)]^{1/2}.\end{aligned} \tag{13}$$

The correlation function is calculated as follows:

$$\begin{aligned}\langle S_K S_{K+l}\rangle&=\frac{1}{Z}\sum_{i,j=1,2,3}\lambda_i^l\lambda_j^{N-l}\langle\psi_i|\mathbf{S}|\psi_j\rangle\langle\psi_j|\mathbf{S}|\psi_i\rangle\\ &=\frac{1}{Z}\left[\lambda_1^l\lambda_3^{N-l}\left(\frac{2}{2+\alpha_1^2}\right)+\lambda_2^l\lambda_3^{N-l}\left(\frac{2}{2+\alpha_2^2}\right)+\lambda_3^l\lambda_1^{N-l}\left(\frac{2}{2+\alpha_1^2}\right)+\lambda_3^l\lambda_2^{N-l}\left(\frac{2}{2+\alpha_2^2}\right)\right]\\ &=\frac{[2/(2+\alpha_1^2)](\lambda_1^l\lambda_3^{N-l}+\lambda_3^l\lambda_1^{N-l})+[2/(2+\alpha_2^2)](\lambda_2^l\lambda_3^{N-l}+\lambda_3^l\lambda_2^{N-l})}{\lambda_1^N+\lambda_2^N+\lambda_3^N}.\end{aligned} \tag{14}$$

As $N\to\infty$

$$\langle S_K S_{K+l}\rangle\to\frac{2}{2+\alpha_1^2}\frac{\lambda_1^{N-l}\lambda_3^l}{\lambda_1^N}=\frac{2}{2+\alpha_1^2}\left(\frac{\lambda_3}{\lambda_1}\right)^l. \tag{15}$$

From this it follows that when $l\to\infty$, $\langle S_K S_{K+l}\rangle\to 0$, which means there is no spontaneous magnetization. Using the correlation function evaluated as above, the susceptibility is calculated as

$$\begin{aligned}\chi&=\frac{Ng^2\mu_B^2}{kT}\sum_{l=-\infty}^{+\infty}\langle S_K S_{K+l}\rangle=\frac{Ng^2\mu_B^2}{kT}\frac{2}{2+\alpha_1^2}\sum_{l=-\infty}^{+\infty}\left(\frac{\lambda_3}{\lambda_1}\right)^l\\ &=\frac{Ng^2\mu_B^2}{kT}\frac{2}{2+\alpha_1^2}\left[1+2\sum_{l=1}^{\infty}\left(\frac{\lambda_3}{\lambda_1}\right)^{|l|}\right]=\frac{Ng^2\mu_B^2}{kT}\frac{2}{2+\alpha_1^2}\left[1+\frac{2\lambda_3/\lambda_1}{1-\lambda_3/\lambda_1}\right]=\frac{Ng^2\mu_B^2}{kT}\frac{2}{2+\alpha_1^2}\left(\frac{\lambda_1+\lambda_3}{\lambda_1-\lambda_3}\right),\end{aligned} \tag{16}$$

where $g=2$. $\lambda_1$, $\lambda_3$, and $\alpha_1$ are obtained from Eqs. (10) and (11).

Using the same procedure, susceptibilities for $S=\frac{1}{2}$ and $S=\frac{3}{2}$ can be calculated. The analytical formula for the susceptibility for $S=\frac{1}{2}$ is given by

$$\chi = (N\mu_B^2/kT)e^{J/kT}, \tag{17}$$

using $g = 2$.

Starting from the Hamiltonian [Eq. (2)], the transfer matrix for $S = \frac{3}{2}$ is obtained in the same way as for $S = 1$ and is given by

$$T = \begin{array}{c|cccc} S \backslash S' & \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{3}{2} \\ \hline \frac{3}{2} & e^{9K-9\alpha} & e^{3K-5\alpha} & e^{-3K-5\alpha} & e^{-9K-9\alpha} \\ \frac{1}{2} & e^{3K-5\alpha} & e^{K-\alpha} & e^{-K-\alpha} & e^{-3K-5\alpha} \\ -\frac{1}{2} & e^{-3K-5\alpha} & e^{-K-\alpha} & e^{K-\alpha} & e^{3K-5\alpha} \\ -\frac{3}{2} & e^{-9K-9\alpha} & e^{-3K-5\alpha} & e^{3K-5\alpha} & e^{-9K-9\alpha} \end{array}, \tag{18}$$

where $K = J/2kT$ and $\alpha = D/4kT$. Diagonalizing this transfer matrix, eigenvalues and eigenvectors are obtained as follows:

*Eigenvalues*:

$$\begin{aligned} \lambda_1 &= e^{-9\alpha}\cosh 9K + e^{-\alpha}\cosh K + [(e^{-9\alpha}\cosh 9K - e^{-\alpha}/\cosh K)^2 + 4e^{-10\alpha}\cosh^2 3K]^{1/2}, \\ \lambda_2 &= e^{-9\alpha}\cosh 9K + e^{-\alpha}\cosh K - [(e^{-9\alpha}\cosh 9K - e^{-\alpha}\cosh K)^2 + 4e^{-10\alpha}\cosh^2 3K]^{1/2}, \\ \lambda_3 &= e^{-9\alpha}\sinh 9K + e^{-\alpha}\sinh K + [(e^{-9\alpha}\sinh 9K - e^{-\alpha}\sinh K)^2 + 4e^{-10\alpha}\sinh^2 3K]^{1/2}, \\ \lambda_4 &= e^{-9\alpha}\sinh 9K + e^{-\alpha}\sinh K - [(e^{-9\alpha}\sinh 9K - e^{-\alpha}\sinh K)^2 + 4e^{-10\alpha}\sinh^2 3K]^{1/2}. \end{aligned} \tag{19}$$

It is evident that $\lambda_1$ is the largest eigenvalue.

*Eivenvectors*:

$$|\psi_1\rangle = \frac{1}{[2(1+x_1^2)]^{1/2}}\begin{pmatrix}1\\x_1\\x_1\\1\end{pmatrix},$$

where $x_1 = (\lambda_1 - 2e^{-9\alpha}\cosh 9K)/(2e^{-5\alpha}\cosh 3K)$,

$$|\psi_2\rangle = \frac{1}{[2(1+x_2^2)]^{1/2}}\begin{pmatrix}1\\x_2\\x_2\\1\end{pmatrix},$$

where $x_2 = (\lambda_2 - 2e^{-9\alpha}\cosh 9K)/(2e^{-5\alpha}\cosh 3K)$,

$$|\psi_3\rangle = \frac{1}{[2(1+x_3^2)]^{1/2}}\begin{pmatrix}1\\x_3\\-x_3\\-1\end{pmatrix},$$

where $x_3 = (\lambda_3 - 2e^{-9\alpha}\sinh 9K)/(2e^{-5\alpha}\sinh 3K)$,

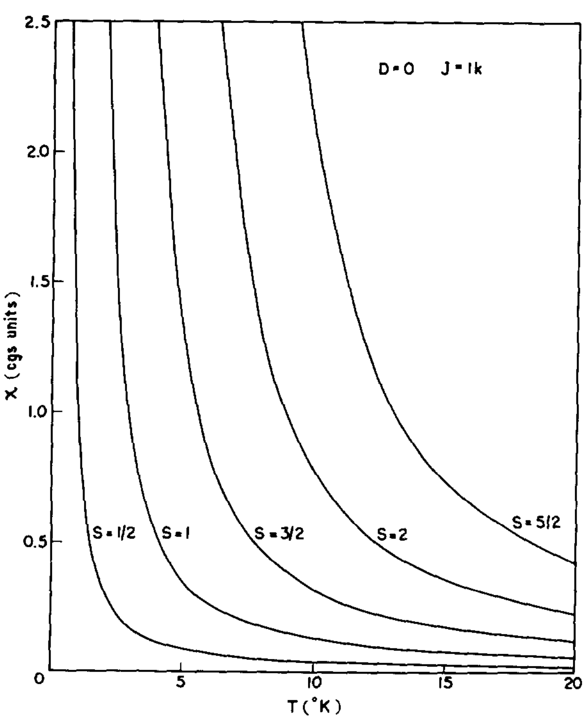


FIG. 1. Ferromagnetic susceptibilities in the absence of the crystal field.

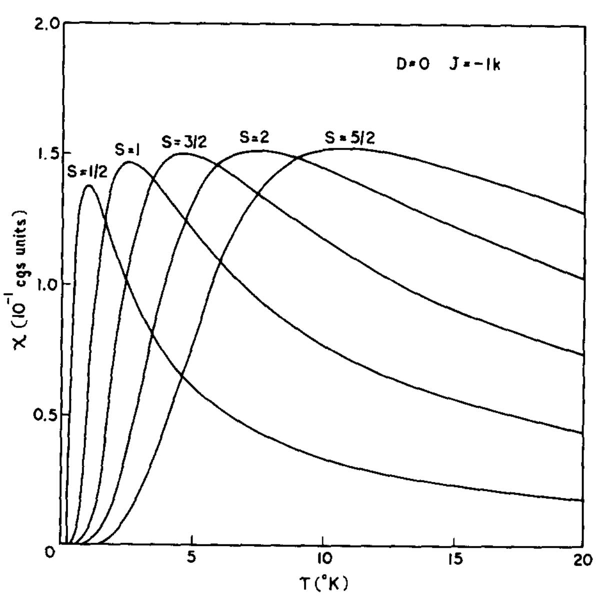


FIG. 2. Antiferromagnetic susceptibilities in the absence of the crystal field.

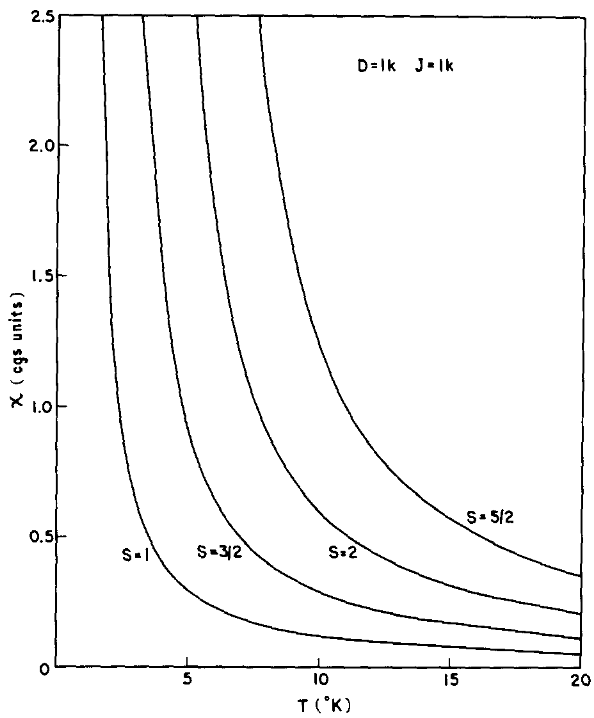


FIG. 3. Ferromagnetic susceptibilities in the presence of the small crystal field.

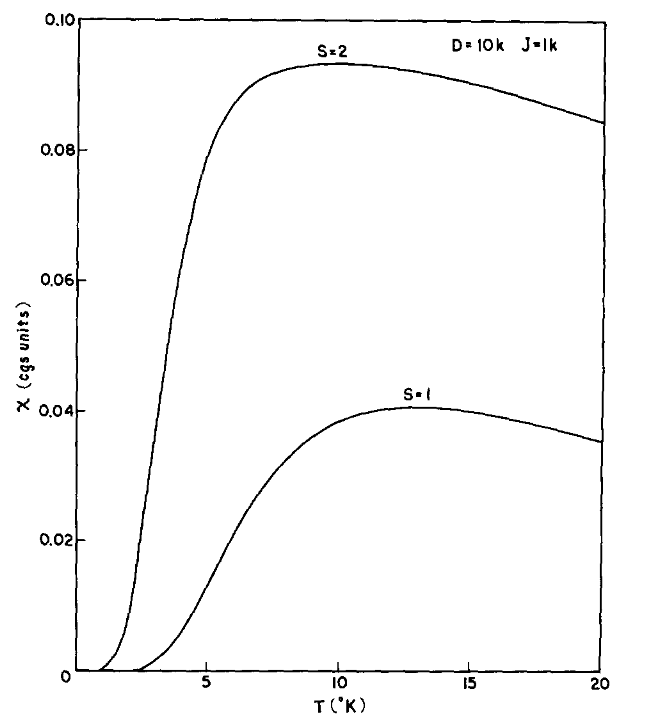


FIG. 5. Ferromagnetic susceptibilities for integral spins in the presence of the large crystal field.

$$|\psi_4\rangle = \frac{1}{[2(1+x_4^2)]^{1/2}}\begin{pmatrix} 1 \\ x_4 \\ -x_4 \\ -1 \end{pmatrix},$$

where $x_4 = (\lambda_4 - 2e^{-9\alpha}\sinh 9K)/(2e^{-5\alpha}\sinh 3K)$. (20)

Susceptibility is calculated in the same way as for $S = 1$ and is obtained as

$$\chi = \frac{N\mu_B^2}{kT(1+x_1^2)}\left\{\frac{(x_1x_3+3)^2}{1+x_3^2}\left(\frac{\lambda_1+\lambda_3}{\lambda_1-\lambda_3}\right) + \frac{(x_1x_4+3)^2}{1+x_4^2}\left(\frac{\lambda_1+\lambda_4}{\lambda_1-\lambda_4}\right)\right\} \quad (21)$$

using $g = 2$. $\lambda$ 's and $x$'s are obtained from Eqs. (19) and (20).

For any spin $S > \frac{3}{2}$, the calculation of susceptibility is performed numerically by writing a FORTRAN program which is very general. This program yields the same results for $S \leqslant \frac{3}{2}$ as obtained analytically.

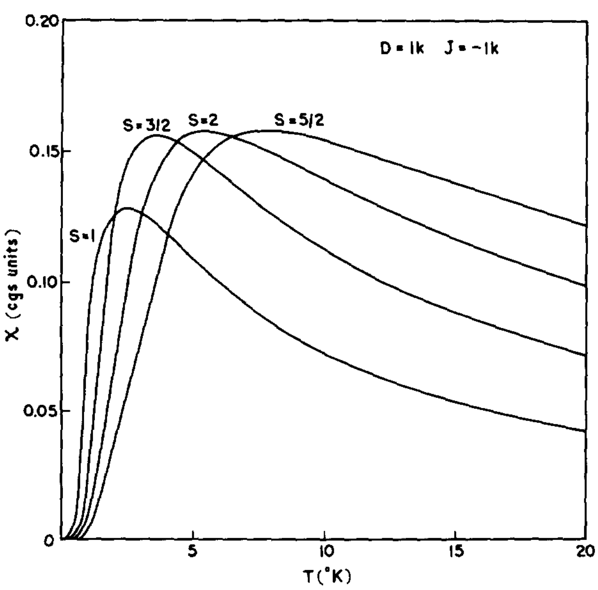


FIG. 4. Antiferromagnetic susceptibilities in the presence of the small crystal field.

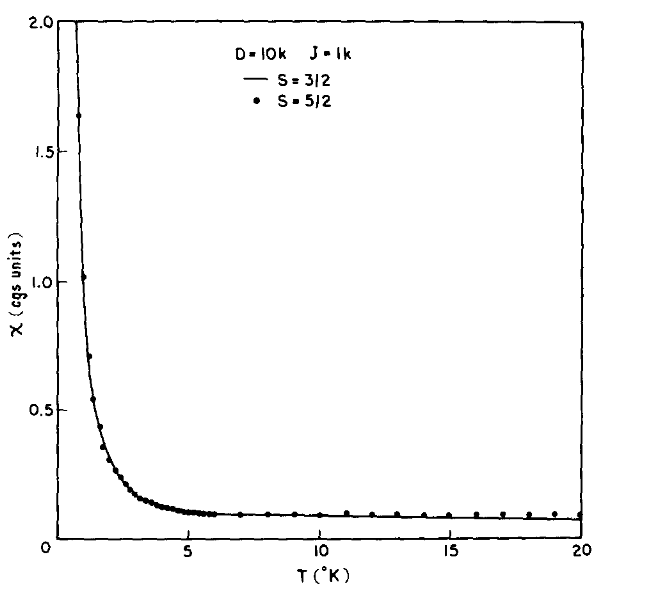


FIG. 6. Ferromagnetic susceptibilities for half-integral spins in the presence of the large crystal field.

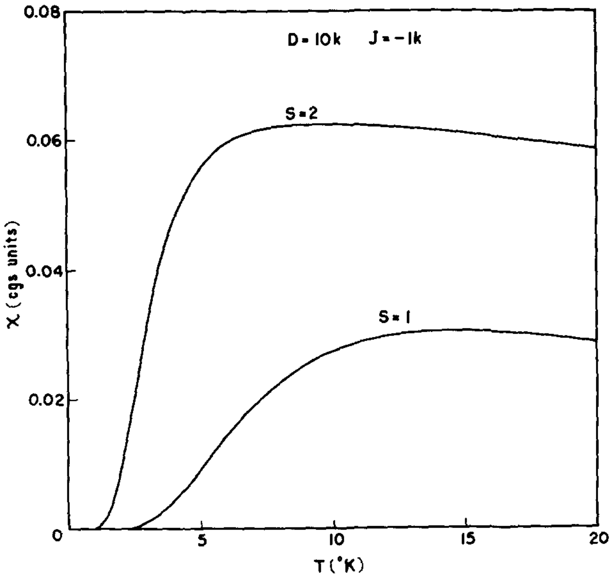


FIG. 7. Antiferromagnetic susceptibilities for integral spins in the presence of the large crystal field.

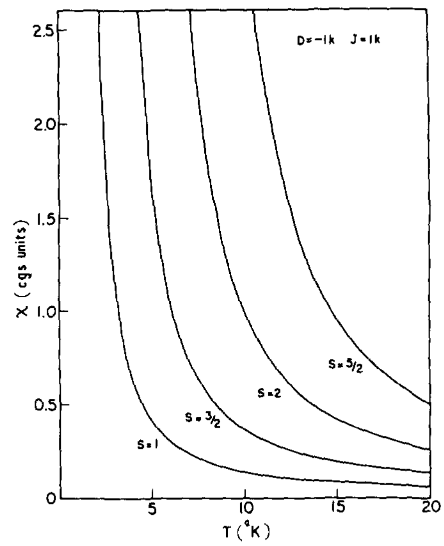


FIG. 9. Ferromagnetic susceptibilities for the low negative crystal field.

## III. RESULTS AND DISCUSSION

In order to study the crystal field effect in a linear Ising model, first the ferromagnetic as well as the antiferromagnetic susceptibilities are calculated in absence of crystal field ($D = 0$) and the results are shown in Figs. 1 and 2, respectively. The antiferromagnetic susceptibilities are calculated by reversing the sign of $J$ in the susceptibility formula given in Sec. II. Suzuki *et al.*[4] calculated these susceptibilities by using different procedures and claimed that the method could be applied for any spin, though they have shown only the results for $S = \frac{1}{2}$, 1, $\frac{3}{2}$. In the present calculation the results are shown up to $S = \frac{5}{2}$, though the method can be applied for general spin. The exact ferromagnetic susceptibilities (Fig. 1) show the usual behavior. The exact antiferromagnetic susceptibilities (Fig. 2) show maxima at certain temperatures which are higher for larger spin values. The maxima become broader as we go to higher spin values. As the axial crystal field is switched on ($D \neq 0$) and its value is small ($D = 1k$) the results do not differ much (from $D = 0$ case) as shown in Figs. 3 and 4. But when the crystal field is large ($D = 10k$), the ferromagnetic susceptibilities start showing broad maxima as indicated in Fig. 5. This happens in the case of integral spins only, and the temperatures at which these maxima occur are lower for higher spin values. This means there must be some critical value of $D$ above which these maxima occur. To calculate this critical value, let us examine the behavior of susceptibility near $T = 0$.

For integral spins since we have the analytical formula of susceptibility for $S = 1$, let us see how the susceptibility for this system behaves at $T = 0$. From Eqs. (10) and (11) we see in the limit of

$$T \to 0,$$

$$\lambda_1 \to 0, \quad \lambda_3 \to 0, \quad \text{and} \quad \alpha_1 \to e^{\beta D}.$$

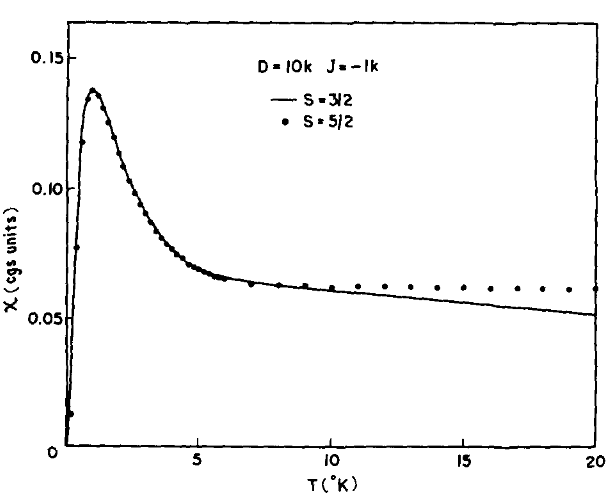


FIG. 8. Antiferromagnetic susceptibilities for half-integral spins in the presence of the large crystal field.

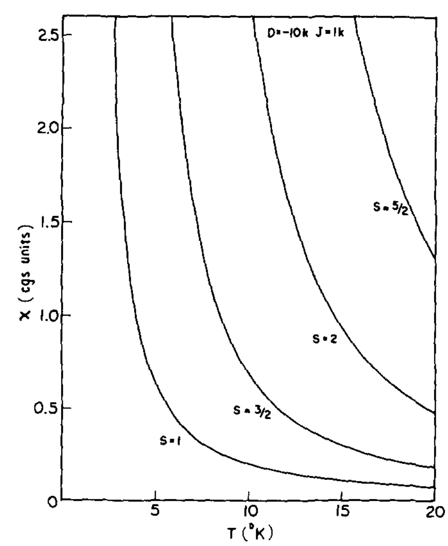


FIG. 10. Ferromagnetic susceptibilities for the high negative crystal field.

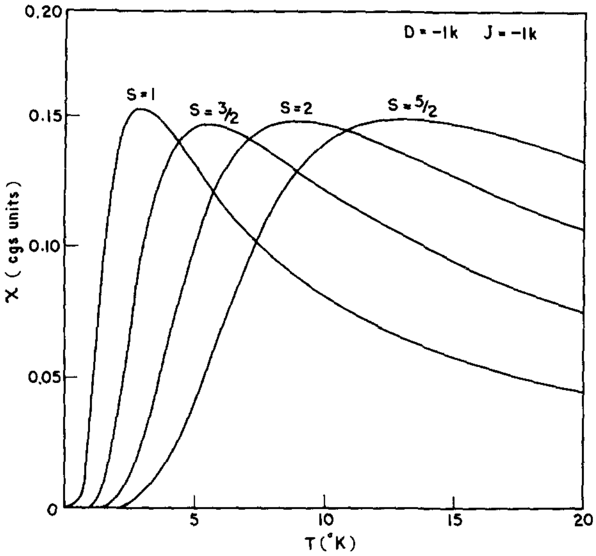


FIG. 11. Antiferromagnetic susceptibilities for the low negative crystal field.

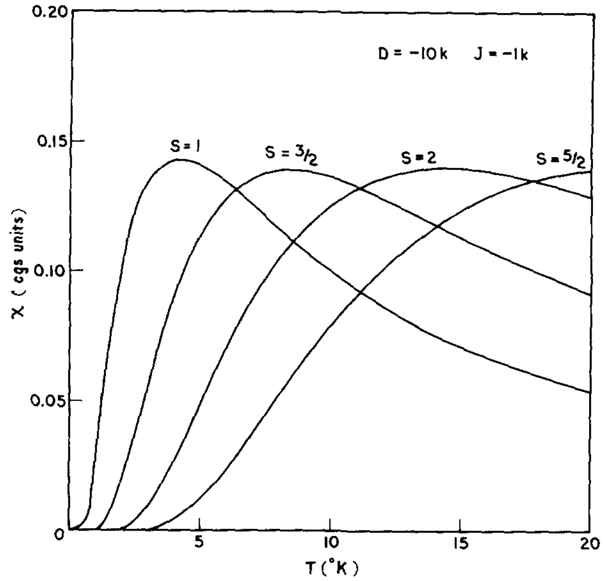


FIG. 12. Antiferromagnetic susceptibilities for the high negative crystal field.

This is true when $D > 2J$. Therefore, the susceptibility evaluated from Eq. (16) becomes

$$\chi = (Ng^2\mu_B^2/kT)e^{-2\beta D} \quad \text{near} \quad T = 0. \tag{22}$$

This shows $\chi \to 0$ at $T = 0$ for $D > 2J$. At high temperature also the susceptibility vanishes. The susceptibility maxima, therefore, appear for $D > 2J$. Since, for positive values of $D$, $D > 2J$ always favors $S_n^z = 0$ (minimum spin) state for integral spins, the critical value of $D$ ($D = 2J$) is same for all integral spins. This has been checked numerically.

On the other hand, half-integral spin susceptibilities show spin $\frac{1}{2}$ behavior at $T = 0$ as in these cases $S_n^z = \frac{1}{2}$ (minimum spin) state is favored. The results are shown in Fig. 6. When $D < 2J$ from Eqs. (10) and (11), we see $\lambda_1 \to 2\cosh(2\beta J)$, $\lambda_3 \to 2\sinh(2\beta J)$, and $\alpha_1 \to 0$ in the limit of $T \to 0$. Near $T = 0$ the susceptibility from Eq. (16) becomes

$$\chi = (Ng^2\mu_B^2/kT)e^{4J/kT}. \tag{23}$$

This is similar to behavior of spin $\frac{1}{2}$ susceptibility given by Eq. (17) and $J$ is replaced by $4J$.

When the crystal field is large ($D = 10k$), the antiferromagnetic susceptibilities, however, do not differ from the behavior shown in the presence of the small crystal field, but broader maxima appear for integral spins as shown in Fig. 7. These maxima are shifted towards the lower temperatures as one consider higher spins in contrast to the small crystal field effect. For half-integral spins the susceptibility maxima become sharper compared to the case of small crystal field and these results are shown in Fig. 8.

So far we have discussed the role of crystal field when $D > 0$. When $D < 0$, as it corresponds to the case of $D < 2J$, the conclusion regarding the ferromagnetic susceptibility is same as given by Eq. (23). This means this crystal field prefers an alignment $S_n^z = \pm S$ (maximum spin) and as $T \to 0$, the susceptibility approaches that for a spin-$\frac{1}{2}$ system with $J \to 4JS^2$. The results for low and high crystal fields are shown in Figs. 9 and 10. The antiferromagnetic susceptibilities for these crystal fields are shown in Figs. 11 and 12. As evident from the figures, the ferromagnetic and the antiferromagnetic susceptibilities show the usual behavior of the Ising model in the absence of the crystal field or in the presence of the small crystal field.

Therefore, from the study of magnetic susceptibilities in the presence of crystal fields of both postive and negative and also of high and low values, one can conclude that an axial crystal field plays an important role in the study of magnetic properties. Its effect can change the magnetic behavior drastically, especially in the case of ferromagnets with integer spins.

## ACKNOWLEDGMENT

Thanks are due to A. Chatterjee for many useful suggestions in the present investigation.

[1]E. Ising, Z. Physik **31**, 253 (1925).
[2]H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252, 263 (1941).
[3]R. Kubo, Busseiron-Kenkyu (in Japanese) No. 1, 1 (1943).
[4]M. Suzuki, B. Tsujiyama, and S. Katsura, J. Math. Phys. **8**, 124 (1967).
[5]M. E. Lines, Phys. Rev. Lett. **42**, 533 (1979).
[6]C. J. Thompson, *Phase Transitions and Critical Phenomena, Vol. 1*, edited by C. Domb and M. S. Green (Academic, New York, 1972), pp. 177.