Beyond Gaussian integrals in Grassman algebra

S. M. de Souza and M. T. Thomaz

Instituto de Física, Universidade Federal Fluminense, Outeiro de São João Batista s/n^{o}, CEP 24020, Niterói, Rio de Janeiro, Brazil

(Received 27 October 1989; accepted for publication 31 January 1990)

Using the properties of determinants, the exact expressions for integrals of functions of the type $\exp[\eta \mathbb{A} \bar{\eta} + (\eta \mathbb{B} \bar{\eta})^{\ell}]$ and their moments, where η and $\bar{\eta}$ are Grassman generators with N components are derived.

I. INTRODUCTION

The quantum description of particles with semi-integer spin through path integrals is carried out by using anticommuting c numbers,[1] the same as occurs in supersymmetric quantum mechanics.[2]

The studies of quantum behavior of the theories through path integrals have been severely restricted by the fact that we are able to calculate only Gaussian integrals. This restriction imposes the introduction of auxiliary fields in the path integral formulation when we want to study models involving a four fermion-point interaction (e.g., Thirring model,[3] Gross–Neveu model,[4] etc.).

The aim of this article is to show how to calculate Grassman integrals of exponentials of powers beyond the quadratic one. We make no restriction on the dimension of the Grassman algebra. We will not consider explicitly the case where the matrices are the discrete version of continuous space-time derivatives.

In Sec. II we make a brief review of the Grassman algebra integral properties. In Sec. III we show a systematic way to calculate integrals of functions of the type $\exp[-\eta \mathbb{A} \bar{\eta} - (\eta \mathbb{B} \bar{\eta})^{\ell}]$, where η and $\bar{\eta}$ are generators of Grassman algebra with N components, and $\ell \leqslant N$. We also get the moments of these integrals. Finally, in Sec. IV, we discuss the results.

II. BRIEF REVIEW OF INTEGRALS IN GRASSMAN ALGEBRA

To make a brief review, we consider a Grassman algebra of dimension 2. Let η and $\bar{\eta}$ be the generators of the algebra. Therefore,

$$\{\eta, \bar{\eta}\} = 0 \tag{1a}$$

and

$$\{\eta, \eta\} = \{\bar{\eta}, \bar{\eta}\} = 0. \tag{1b}$$

As a consequence of the anticommutation (1b) we have $\eta^2 = 0$ and $\bar{\eta}^2 = 0$.

The integrals of the generators are defined as

$$\int d\eta \cdot 1 = 0, \quad \int d\bar{\eta} \cdot 1 = 0, \tag{2a}$$

$$\int d\eta \, \eta = 1, \quad \int d\bar{\eta} \, \bar{\eta} = 1, \tag{2b}$$

$$\int d\eta \, \bar{\eta} = 0, \quad \text{and} \quad \int d\bar{\eta} \, \eta = 0, \tag{2c}$$

and, if we write the most general element of this algebra, $g(\eta, \bar{\eta}) = g_{00} + g_{10}\eta + g_{01}\bar{\eta} + g_{11}\eta\bar{\eta}$, where $g_{ij} \in \mathbb{C}$, $i,j = 0,1$, then,

$$\int d\eta \, g(\eta, \bar{\eta}) = g_{10} + g_{11}\bar{\eta} \tag{3a}$$

and

$$\int d\bar{\eta} \, g(\eta, \bar{\eta}) = g_{01} - g_{11}\eta. \tag{3b}$$

To write down the results of the integrals (3a) and (3b) you need to bring η close to $d\eta$ and $\bar{\eta}$ close to $d\bar{\eta}$, respectively.

It is a well-known result[1,2] that

$$\int d\bar{\eta} \, d\eta \, e^{-a\eta\bar{\eta}} = -a. \tag{4}$$

This result is easily extended to the case of 2^N-dimensional algebra and amounts to

$$\int \prod_{i=1}^{N} d\bar{\eta}_i \prod_{i=1}^{N} d\eta_j \, e^{-\eta_i a_{ij} \bar{\eta}_j} = (-1)^N \det \mathbb{A}, \tag{5}$$

where a_{ij} are the elements of matrix $\mathbb{A}$.

III. EXACT RESULTS OF INTEGRALS INVOLVING EXPONENTIAL FUNCTIONS

Let us consider a Grassman algebra with dimension 2^N, whose generators are $\{\eta_1,...,\eta_N; \bar{\eta}_1,...,\bar{\eta}_N\}$ and who satisfy analogous anticommutation relations such as (1a) and (1b).

We want to calculate the integral,

$$I_2 \equiv \int \prod_{i=1}^{N} d\bar{\eta}_i \, d\eta_i \, \exp\{-[\eta \mathbb{A} \bar{\eta} + (\eta \mathbb{B} \bar{\eta})^2]\}, \tag{6}$$

where $\eta \mathbb{A} \bar{\eta} \equiv \eta_i a_{ij} \bar{\eta}_j$ and $a_{ij} \in \mathbb{C}$ are the elements of matrix $\mathbb{A}$, and $\eta \mathbb{B} \bar{\eta} \equiv \eta_i b_{ij} \bar{\eta}_j$, where $b_{ij} \in \mathbb{C}$ are the elements of matrix $\mathbb{B}$.

Using the anticommutation relations (1b) we have,

$$e^{-\eta \mathbb{A} \bar{\eta}} = \sum_{m=0}^{N} \frac{(-1)^m}{m!} \sum_{\substack{i_1, j_1 = 1 \\ \vdots \;\; \vdots \\ i_m, j_m = 1}}^{N} a_{i_1, j_1} \cdots a_{i_m, j_m} \times \eta_{i_1} \bar{\eta}_{j_1} \cdots \eta_{i_m} \bar{\eta}_{j_m}, \tag{7a}$$

and

$$e^{-(\eta \mathbb{B} \bar{\eta})^2} = \sum_{n=0}^{[N/2]} \frac{(-1)^n}{n!} \sum_{\substack{i_1, j_1 = 1 \\ \vdots \;\; \vdots \\ i_{2n}, j_{2n} = 1}}^{N} b_{i_1, j_1} \cdots b_{i_{2n}, j_{2n}} \times \eta_{i_1} \bar{\eta}_{j_1} \cdots \eta_{i_{2n}} \bar{\eta}_{j_{2n}}, \tag{7b}$$

where $[N/2]$ means the biggest integer smaller or equal to $N/2$.

The terms that can contribute to (6) are the ones that we get from the multiplication of (7a) by (7b) where we have N pairs of generators $\eta_i\bar{\eta}_j$,

$$\exp\{-[\eta\mathbb{A}\bar{\eta}+(\eta\mathbb{B}\bar{\eta})^2]\} = \sum_{\substack{m=0,1,\ldots,N\\ n=0,1,\ldots,[N/2]\\ m+2n=N}} \frac{(-1)^{n+m}}{n!m!} \sum_{\substack{i_1,j_1=1\\ \vdots\\ i_N,j_N=1}}^{N} a_{i_1,j_1}\cdots a_{i_m,j_m} \times b_{i_{m+1},j_{m+1}}\cdots b_{i_N,j_N}\eta_{i_1}\bar{\eta}_{j_1}\cdots\eta_{i_m}\bar{\eta}_{j_m}\cdots\eta_{i_N}\bar{\eta}_{j_N}. \quad (8)$$

However, the only integrals that give nonzero contribution to I_2 are

$$\int d\bar{\eta}_1\, d\eta_1\cdots d\bar{\eta}_N\, d\eta_N\; \eta_1\bar{\eta}_1\cdots\eta_N\bar{\eta}_N=1, \quad (9)$$

and all the integrands that are permutations from the basic configuration $\eta_1\bar{\eta}_1\cdots\eta_N\bar{\eta}_N$. Any configuration that has more than one generator η_i and/or $\bar{\eta}_j$ gives a null result.

From (8) we notice that if one permutes the position of any two elements a_{ij} there is no change in the sign of the element, as well as if one permutes the position of any two elements b_{ij}. Therefore, we have $m!(2n)!$ identical elements in (8), and it can be rewritten as

$$I_2=\int\prod_{i=1}^{N} d\bar{\eta}_i d\eta_i \sum_{\substack{m=0,1,\ldots,N\\ n=0,1,\ldots,[N/2]\\ m+2n=N}} (-1)^{n+m} \times(-1)^{(N/2)(N-1)}\eta_1\eta_2\cdots\eta_N\,\frac{(2n)!}{n!}\sum_{\sigma(q,\eta)}\sigma(q,\eta) \times\sum_{\substack{j_1=1\\ \vdots\\ j_N=1}}^{N} a_{q(1),j_1}\cdots a_{q(m),j_m} b_{q(m+1),j_{m+1}}\cdots b_{q(N),p_N} \times\bar{\eta}_{j_1}\bar{\eta}_{j_2}\cdots\bar{\eta}_{j_N}, \quad (10)$$

where $\sigma(q,\eta)$ are the inequivalent permutations of $\eta_{i_1}\cdots\eta_{i_N}$ to get the basic configurations $\eta_1\eta_2\cdots\eta_N$. We should pay attention to the fact that for a given configuration the indices are chosen such that $i_1\cdots i_m$ are in increasing order. We have $\sigma(q,p)=1(-1)$ if the permutation is even (odd).

Let $\sigma(p,\bar{\eta})$ be the permutation of the configuration $\bar{\eta}_{j_1}\cdots\bar{\eta}_{j_N}$ that becomes the basic configuration $\bar{\eta}_1\bar{\eta}_2\cdots\bar{\eta}_N$. As before, $\sigma(p,\bar{\eta})=1(-1)$ if the permutation is even (odd). The sum over $j_1\cdots j_N$ in expression (10) can be written as,

$$\sum_{\sigma(p,\bar{\eta})}\sigma(p,\bar{\eta})a_{q(1),p(1)}\cdots a_{q(m),p(m)} \times b_{q(m+1),p(m+1)}\cdots b_{q(N),p(N)}\bar{\eta}_1\bar{\eta}_2\cdots\bar{\eta}_N \equiv\det\mathbb{O}_q\bar{\eta}_1\bar{\eta}_2\cdots\bar{\eta}_N, \quad (11)$$

where Eq. (11) used the definition and properties of the determinant of any matrix.[5] From the definition (11) we see that the lines of matrix $\mathbb{O}_q$ are gotten from the lines of matrices $\mathbb{A}$ and $\mathbb{B}$. The order that the lines appear in $\mathbb{O}_q$ depends on the permutation $q=(q(1),q(2),\ldots,q(N))$.

Therefore, I_2 can be finally written as,

$$I_2=\sum_{\substack{m=0,1,\ldots,N\\ n=0,1,\ldots,[N/2]\\ m+2n=N}}(-1)^{n+m}\frac{(2n)!}{n!}\sum_{\sigma(q,\eta)}\sigma(q,\eta)\det\mathbb{O}_q. \quad (12)$$

The only restriction to expression (12) is that $N\geqslant 2$.

Let us consider the special cases $N=2$ and $N=3$ of the general expression (12):

(i) $N=2$:

$$I_2=\det\mathbb{A}-2\det\mathbb{B}. \quad (13)$$

For dimension 2, I_2 gives a null result if $\det\mathbb{A}=2\det\mathbb{B}$.

(ii) $N=3$:

$$I_2=-\det\begin{vmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{vmatrix}+2\det\begin{vmatrix}a_{11}&a_{12}&a_{13}\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{vmatrix} +2\det\begin{vmatrix}b_{11}&b_{12}&b_{13}\\a_{21}&a_{22}&a_{23}\\b_{31}&b_{32}&b_{33}\end{vmatrix}+2\det\begin{vmatrix}b_{11}&b_{12}&b_{13}\\b_{21}&b_{22}&b_{23}\\a_{31}&a_{32}&a_{33}\end{vmatrix}. \quad (14)$$

It is very easy to extend the result (12) for integrals I_ℓ, where ℓ is a positive integer, that is,

$$I_\ell\equiv\int\prod_{i=1}^{N} d\bar{\eta}_i\, d\eta_i\,\exp\{-[\eta\mathbb{A}\bar{\eta}+(\eta\mathbb{B}\bar{\eta})^\ell]\}, \quad (15)$$

where $2N$ is the number of generators of the Grassman algebra and $N\geqslant\ell$.

Proceeding in an analogous way as for the case I_2, we get

$$I_\ell=\sum_{\substack{m=0,1,\ldots,N\\ n=0,1,\ldots,[N/\ell]\\ m+n\ell=N}}(-1)^{n+m}\frac{(\ell n)!}{n!} \times\sum_{\sigma(q,\eta)}\sigma(q,\eta)\det\mathbb{O}_q, \quad (16)$$

where $[N/\ell]$ is the biggest integer smaller or equal to N/ℓ, $\sigma(q,\eta)$ has the same definition as the one given previously, and $\mathbb{O}_q$ is such that

$$\sum_{\sigma(p,\bar{\eta})}\sigma(p,\bar{\eta})a_{q(1),p(1)}\cdots a_{q(m),p(m)} \times b_{q(m+1),p(m+1)}\cdots b_{q(N),p(N)}=\det\mathbb{O}_q, \quad (17)$$

and now we have $\ell\cdot n$ elements b_{ij} in (17).

Finally, we want to derive the moments of integral I_ℓ, that is,

$$G_\ell^{(n,m)}(\eta,\bar{\eta}) \equiv\int\prod_{i=1}^{N} d\bar{\eta}_i\, d\eta_i\,\eta^m\bar{\eta}^n\exp\{-[\eta\mathbb{A}\bar{\eta}+(\eta\mathbb{B}\bar{\eta})^\ell]\}, \quad (18)$$

where $\eta^m\equiv\eta_1^{m_1}\eta_1^{m_2}\cdots\eta_N^{m_N}$, $m_i=0$ or 1; and, $\bar{\eta}^n\equiv\bar{\eta}_1^{n_1}\cdots\bar{\eta}_N^{n_N}$, $n_i=0$ or 1. We will assume from now on that $I_1,I_2,\ldots,I_c$ and $J_1,J_2,\ldots,J_c$ are the indices of the generators whose powers m_i and n_i, respectively, are not zero. In this case the integrand of (18) becomes,

$$\eta^m\bar{\eta}^n \exp\{-[\eta A\bar{\eta} + (\eta B\bar{\eta})^\ell]\}$$
$$= \sum_{\substack{r=0,1,\ldots,N \\ s=0,1,\ldots,[N/\ell] \\ r+s\ell+c=N}} \frac{(-1)^{s+r}(-1)^{c(c-1)/2}}{r!s!} a_{i_1,j_1}\cdots a_{i_r,j_r}$$
$$\times b_{i_{r+1},j_{r+1}}\cdots b_{i_{r+s\ell},j_{r+s\ell}}\eta_{I_1}\bar{\eta}_{J_1}\cdots\eta_{I_c}\bar{\eta}_{J_c}\eta_{i_1}$$
$$\times\bar{\eta}_{j_1}\cdots\eta_{i_{r+s\ell}}\bar{\eta}_{j_{r+s\ell}}. \tag{19}$$

We want to rewrite (19) in such a way that it resembles expression (11). For doing this, we use

$$\eta_{I_1}\bar{\eta}_{J_1}\cdots\eta_{I_c}\bar{\eta}_{J_c}$$
$$= \sum_{\substack{i_{r+s\ell},\, j_{r+s\ell}=1 \\ \vdots \quad \vdots \\ i_N,\quad j_N=1}} \delta_{I_1,i_{r+s\ell+1}}\delta_{J_1,j_{r+s\ell+1}}\cdots\delta_{I_c i_N}\delta_{J_c j_N}$$
$$\times\eta_{i_{r+s\ell+1}}\bar{\eta}_{j_{r+s\ell+1}}\cdots\eta_{i_N}\bar{\eta}_{j_N}, \tag{20}$$

and define the following matrix elements:

$$C_{i_{r+s\ell+1},j_{r+s\ell+1}} = \delta_{I_1,i_{r+s\ell+1}}\cdot\delta_{J_1,j_{r+s\ell+1}},$$
$$\vdots \tag{21}$$
$$C_{i_N,j_N} = \delta_{I_c i_N}\cdot\delta_{J_c j_N}.$$

Making use of the previous results, we get

$$G_\ell^{(n,m)}(\eta,\bar{\eta}) = \sum_{\substack{r=0,1,\ldots,N \\ s=0,1,\ldots,[N/\ell] \\ r+s\ell+c=N}} (-1)^{s+r}\frac{(\ell s)!}{s!}$$
$$\times(-1)^{c(c-1)/2}\sum_{\sigma(q,\eta)}\sigma(q,\eta)\det \mathbf{O}_q, \tag{22}$$

with $\sigma(q,\eta)$ as previously defined, and

$$\det \mathbf{O}_q \equiv \sum_{\sigma(p,\bar{\eta})}\sigma(p,\bar{\eta})a_{q(1),p(1)}\cdots a_{q(r),p(r)}$$
$$\times b_{q(r+1),p(r+1)}\cdots b_{q(r+s\ell),p(r+s\ell)}$$
$$\times C_{q(r+s\ell+1),p(r+s\ell+1)}\cdots C_{q(N),p(N)}. \tag{23}$$

IV. CONCLUSIONS

We have shown how to derive in a simple way the integrals of exponential functions of the type $(\eta B\bar{\eta})^\ell$. These integrals can be written as a sum of determinants of matrices, and these matrices are composed of lines of matrix **A** and **B**. Only when N is even does det **B** contribute to the final result. These features can be seen from the examples $N=2$ and $N=3$.

The extension to the case where we have a differential operator is under study. It is still necessary to write the results (12), (16), and (22) in more manageable form for physical applications.

ACKNOWLEDGMENTS

One of us (S.M.S.) would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support, and (M.T.T.) would like to thank CNPq and Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ) for partial financial support.

[1]C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980); L. D. Faddeev and A. A. Slavnov, *Gauge Fields: Introduction to Quantum Theory* (Benjamin/Cummings, New York, 1980).
[2]P. Salomonson and J. W. Van Holten, Nucl. Phys. **196**, 509 (1982); L. É. Gendenshtein and I. V. Krive, Sov. Phys. Usp. **28**, 645 (1985).
[3]W. Thirring, Ann. Phys. (N.Y.) **3**, 91 (1958).
[4]D. J. Gross and A. Neveu, Phys. Rev. D **10**, 3235 (1974).
[5]We used the basic properties of determinants to write down expression (11). For a review of such properties, see, for example, D. L. Krider *et al.*, *An Introduction to Linear Analysis* (Addison–Wesley, Reading, 1966) Appendix III.

Minimization of Landau potentials invariant under O(3).II

Y. Brihaye, F. Delaunois, and J. Nuyts
Physique Théorique et Mathématique, Université de Mons-Hainaut, B-7000 Mons, Belgium

(Received 21 November 1989; accepted for publication 7 January 1990)

In a previous paper, the absolute minima for Landau potentials were computed for the irreducible representations of SO(3) or O(3) of spin up to four. Here, this analysis is extended to the case of spin five and six. Some novel properties of the extrema are pointed out.

I. INTRODUCTION

A couple of years ago, an analysis of the minimization of SO(3) invariant Landau potentials was reported[1] for irreducible representations of integer spin up to $s=4$ and the results were discussed in relation with Michel's conjecture.[2] In this article we extend the analysis of Ref. 1 to the case $s=5$ and partially to the case $s=6$. Our purpose is to stress several new features that appear for $s>4$. The method used and the notations and conventions are those of Ref. 1. For the subgroups of O(3) we have adopted the notation of Schönflies.

II. THE CASE $s=5$

Among all invariants constructed from the symmetric traceless tensor S_{abcde} $(a,b,...=1,2,3)$ corresponding to the spin 5 (dimension 11) irreducible representation of SO(3), only two, chosen to be

$$Q=S_{abcde}S_{abcde}\,, \tag{1a}$$

$$K=S_{abcdp}S_{abcdq}S_{ijklp}S_{ijklq}\,, \tag{1b}$$

are polynomially independent and of degree less than four. The most general Landau potential invariant under SO(3) then reads

$$V=2\mu Q+qQ^2+kK\,. \tag{2}$$

It is in fact invariant under O(3). A fixed vector S can be invariant under any of the following subgroups of SO(3) (see Refs. 2 and 3 for the notations):

$$\mathrm{SO}(3),\ C_\infty,\ D_5,\ C_5,\ D_4,\ C_4,\ D_3,\ C_3,\ D_2,\ C_2,\ 1, \tag{3}$$

among which C_∞, D_5, D_4, and D_3 are maximal subgroups. The stability group of S then consists of one of the groups above eventually extended by a reflexion Z_2 of determinant -1.

The search of the extremal configurations of V, i.e., of the solutions of the 11 equations

$$\frac{\partial V}{\partial S}=0\,, \tag{4}$$

has been carried out for all the nontrivial subgroups of SO(3) appearing in the list (3). Table I collects the canonical directions $\hat{S}_g$ of the solutions of Eq. (4) that are invariant under a subgroup g. These directions are obtained by choosing some axes of symmetry of the group g along the coordinate axes: the x axis as the principal axis of any symmetry group C_n or D_n. In the case D_n, one extra C_2 axis is along the y axis. All the solutions are generated by applying rotations on these particular configurations. The other components of $\hat{S}_g$ can be obtained from the ones given in the table by using either the symmetry or the tracelessness of the tensor.

Remarkably enough, the directions $\hat{S}_g$ (normalized arbitrarily) are independent of the parameters of the potential (i.e., of μ, q, and k). Therefore, all the relevant information is contained into the ratio

$$\alpha(g)\equiv(Q^2/K)(\hat{S}_g)\,. \tag{5}$$

Indeed, denoting the extremum as

$$S_g=\lambda\hat{S}_g\,, \tag{6}$$

the virial theorem ($\partial_\lambda V=0$) can be used to fix the factor λ

$$\lambda^2=-\mu/(\hat{Q}\cdot L(g)),\quad L(g)\equiv q+k/\alpha(g)\,, \tag{7}$$

where $\hat{Q}$, the length of $\hat{S}$, is also given in Table I. The extremal value of the potential then reads

$$V(S_g)=-\mu^2/L(g)\,. \tag{8}$$

In order to satisfy the asymptotic condition of the potential (i.e., that V be increasing when S goes to infinity in all directions), it is necessary for all quantities $L(g)$ to be positive; in other words, we have to limit our analysis to the regions delimited by

$$q+k/3>0,\quad \text{if } k>0, \tag{9a}$$

$$q+k/2>0,\quad \text{if } k<0\,. \tag{9b}$$

We have checked numerically that these conditions are also sufficient.

One observes that (i) for every g in Table I, the value of $\alpha(g)$ is fixed (C_3 is not allowed as an extremum).

(ii) For every $g\neq D_2(II)$, all the components of $S(g)$ are fixed univokely with the axes of symmetry of g. As a consequence, the set of extrema invariant under g ($g\neq D_2(II)$) assemble into one orbit that is three dimensional except for the case $C\infty$ for which it is two dimensional.

The configurations invariant under $D_2(II)$ are of different nature since the Eqs. (4) do not fix the relative values of all components of S. Let

$$b_a\equiv S_{123aa}\quad (\text{no sum over } a)\,, \tag{10}$$

$$b_1+b_2+b_3=0\,. \tag{11}$$

The arbitrary normalization is

$$b_1^2+b_2^2+b_3^2=2\,, \tag{12}$$

with the restriction

$$b_1^2\neq b_2^2\neq b_3^2\neq b_1^2\,. \tag{13}$$

The union of the corresponding three-dimensional orbits forms a four-dimensional manifold, a strata.

 0022-2488/90/061300-04$03.00

TABLE I. The values of the independent components of the extremal directions $\hat{S}_{abcde}(g)$ $(abcde = 22233,...)$ are given in canonical position for all possible SO(3) subgroups [sgr. SO(3)] in the case $s = 5$. (The C_3 subgroup is never extremal.) The symmetry axes of g are described in the text. The table contains also $\hat{Q}$ (the length of $\hat{S}$), α [defined in Eq. (5)] and the stability group of $\hat{S}$ [sgr.O (3)] [as stability group of O(3), see Ref. 3 for notations]. For $D_2(II)$ one has $b_1^2 \neq b_2^2 \neq b_3^2 \neq b_1^2$. The cases have been classified with increasing values of α.

sgr. SO(3)	D_5	C_∞	D_4	$D_2(I)$	$D_2(II)$	D_3	C_5	C_4	C_2
22233	1	0	0	0	0	1	0	0	0
22333	0	0	0	0	0	0	$\pm\sqrt{21}$	0	0
11333	0	0	0	0	0	0	0	0	0
11133	0	-4	0	0	0	0	-4	-4	6
11122	0	-4	0	0	0	0	-4	-4	6
11222	0	0	0	0	0	4	0	0	0
12233	0	1	0	0	0	0	1	1	9
11233	0	0	0	0	0	-4	0	0	0
11223	0	0	0	0	0	0	0	0	0
12223	0	0	1	1	b_2	0	0	$\pm\sqrt{21}$	$\pm\sqrt{21}$
12333	0	0	-1	1	b_3	0	0	$\mp\sqrt{21}$	$\pm\sqrt{21}$
$\hat{Q}$	16	504	40	40	40	720	840	1344	8037
α	2	$\frac{27}{11}$	$\frac{25}{9}$	$\frac{25}{9}$	$\frac{25}{9}$	$\frac{1350}{451}$	3	3	3
sgr. O(3)	D_{5h}	$C_{\infty v}$	D_{4d}	D_{2d}	D_2	D_{3h}	C_{5v}	C_{4v}	C_{2v}

The triple degeneracy of $\alpha(g)$ for $g = C_5, C_4, C_2$ or for $g = D_2(I), D_2(II), D_4$ is intriguing. We will see similar features in the next section.

As it was the case for $s = 3$ (Ref. 1), the absolute minimum for the $s = 5$ potential is always a discrete subgroup of SO(3). It is the group D_5 (maximal subgroup) in the case $k < 0$; and, on the same footing, either C_5 or C_4 or C_2 when $k > 0$; none of these being maximal. The above statements were checked numerically for numerous values of q and k. The numerical analysis confirms also that the conditions (9) are indeed necessary and sufficient.

For completeness we have also given in Table I the stability group of the configuration as a subgroup of O(3) the true stability group of the potential (the notations for groups are those of Ref. 3).

Finally let us stress that the configurations of stability group reduced to the identity were not solved analytically but that they never appeared as minima of the potential in our detailed numerical minimizations for numerous values of the parameters.

III. THE CASE $s=6$

There exists in this case four (polynomially) independent invariants of degree $\leqslant 4$ that can be constructed from the six indices (symmetric and traceless) tensor S_{abcdef} corresponding to the spin 6 irreducible representation of SO(3) (dimension 13): a quadratic one Q, a cubic one P, and two quartic ones K and K_1. They can be chosen as follows:

$$\begin{aligned} Q &= S_{abcdef} S_{abcdef}, \\ P &= S_{abcdef} S_{abcghi} S_{defghi}, \\ K &= S_{abcdep} S_{abcdeq} S_{ijklmp} S_{ijklmq}, \\ K_1 &= S_{abcdpq} S_{abcdrs} S_{ijklpr} S_{ijklqs}. \end{aligned} \tag{14}$$

The most general Landau potential reads

$$V = 2\mu Q + qQ^2 + pP + kK + k_1 K_1, \tag{15}$$

but we will consider only the case $p = 0$. A fixed vector S can be invariant under a rich variety of subgroups of SO(3) (invariance under parity is in this case trivial as is the invariance under $S \to -S$):

$$\mathrm{SO}(3), D_\infty, Y, O, T, D_6, D_5, D_4, D_3, D_2, C_3, C_2, 1 . \tag{16}$$

For the first time, the three groups Y, O, T (the symmetry groups, respectively of the icosahedron, of the cube, and of the tetrahedron) appear as possible stability groups of the tensor S. For technical reasons, we solved Eq. (4) for configurations invariant under g running from D_∞ to D_4 in the list (16); in all other cases the equations are really cumbersome. However, by a numerical exploration of the absolute minima of (15) for various values of the parameters, we failed to find any minimal configuration giving a value lower than the ones obtained analytically.

As in the previous case, the canonical directions, $\hat{S}_g$, of extrema invariant under a group g are collected in Table II. For convenience, we define the parameters d_a and r so that

$$S_{112233} = -2r, \tag{17a}$$

$$\begin{aligned} S_{111122} &= r + d_3, \quad S_{222233} = r + d_1, \\ S_{333311} &= r + d_2 . \end{aligned} \tag{17b}$$

The other components of S are obtainable from the ones displayed in the table by exploiting the properties of symmetry and of tracelessness of the tensor. The conventions for the axes of C_n and D_n are the same as previously; the group Y is oriented so that the coordinate axis ox is a symmetry axis of order 5 and oy of order 2. For the cube O (resp. the thedrahedron T), the three axes of order 4 (resp. 2) were chosen as the coordinate axes.

Apart from the case $D_6(II)$ the set of extremal configurations invariant under g forms a unique orbit characterized by the two ratios

$$\alpha(g) = (Q^2/K)(\hat{S}_g), \quad \beta(g) = (Q^2/K_1)(\hat{S}_g), \tag{18}$$

also given in Table II. The normalization λ (i.e., such that $S \equiv \lambda \hat{S}$) is given by

TABLE II. The values of the independent components of the extremal directions $\hat{S}(g)$ are given for the main subgroups of SO(3) in the case $s = 6$. The symmetry axes of g are described in the text. The notation is the same as in Table I, α and β are defined in Eq. (18). For $D_6(II)$ the details are given in Eqs. (23)–(27).

sgr. SO(3)	D_∞	Y	O	T	D_4	D_5	$D_6(I)$	$D_6(II)$
222223	0	0	0	0	0	0	0	0
233333	0	0	0	0	0	0	0	0
133333	0	-7	0	0	0	0	0	0
111113	0	0	0	0	0	0	0	0
111112	0	0	0	0	0	0	0	0
122222	0	0	0	0	0	1	0	0
222333	0	0	0	0	0	0	0	0
111333	0	0	0	0	0	0	0	0
111222	0	0	0	0	0	0	0	0
d_1	0	0	0	1	0	0	1	d_1
d_2	-7	-7	0	1	± 1	0	0	-7
d_3	7	7	0	1	± 1	0	0	7
r	1	1	1	0	1	0	0	1
$\hat{Q}$	3696	8400	462	90	528	96	32	see
α	$\frac{242}{97}$	3	3	3	$\frac{242}{81}$	$\frac{8}{3}$	2	Eqs.
β	$\frac{1089}{355}$	5	$\frac{726}{155}$	$\frac{54}{11}$	$\frac{11616}{2465}$	$\frac{18}{5}$	2	(23)–(27)
sgr. O(3)	$D_{\infty h}$	Y_h	O_h	T_h	D_{4h}	D_{5d}	D_{6h}	D_{6h}

$$\lambda^2 = \frac{-\mu}{\hat{Q}\cdot L(g)}, \quad L(g) \equiv q + \frac{k}{\alpha(g)} + \frac{k_1}{\beta(g)}. \tag{19}$$

The extremal value of the potential reads again

$$V(S_g) = -\mu^2/L(g). \tag{20}$$

Of course, we should impose the positivity of the quartic part of V in general and in particular for all the invariant directions of S.

It can be checked analytically that the positive region is bounded by the planes $L(g) = 0$ corresponding to $g = Y, O$, and one of the D_6, namely, $D_6(I)$. Defining the variables P_1, P_2, and P_3 by

$$P_1 = q + k/2 + k_1/2,$$
$$P_2 = q + k/3 + k_1/5, \tag{21a}$$
$$P_3 = q + k/3 + 155k_1/726,$$

the three conditions

$$P_i > 0, \tag{21b}$$

guarantee the positivity of all the $L(g)$ of Table II. For completeness let us give the inverse of the transformation

$$q = -2P_1 - \frac{1305P_2}{49} + \frac{1452P_3}{49},$$
$$k = 6P_1 + \frac{6240P_2}{49} - \frac{6534P_3}{49}, \tag{22}$$
$$k_1 = -\frac{3630P_2}{49} + \frac{3630P_3}{49}.$$

For all the g in Table II except the second D_6, $D_6(II)$, it is easy to show that the planes $L(g)$, once written in terms of the variables P_1, P_2 and P_3 have positive coefficients which proves the result. For $D_6(II)$ also the conditions (21) are sufficient for the positivity but the proof is more subtle and will be discussed shortly. For the configurations invariant under D_3, D_2, C_3, C_2, and 1, we have not been able to prove analytically that the conditions (21) are sufficient. However a detailed numerical exploration of arbitrary directions indicates that these conditions are indeed necessary and sufficient.

The solutions invariant under the subgroup D_6 are peculiar in many respects, one observes in the table that they assemble into the following two different categories:

(i) a first one completely analog to the cases discussed above; we call it $D_6(I)$,

(ii) a second one, say $D_6(II)$, more complicated in view of the fact that the extremal directions and the ratios α and β depend explicitly on the parameters q, k, and k_1.

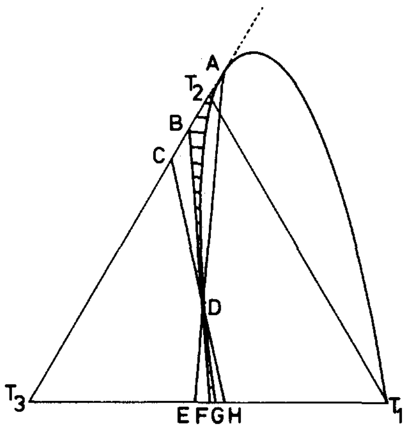


FIG. 1. The space of the variables P_1, P_2, P_3 defined in (21) is cut by the plane $P_1 + P_2 + P_3 = 1$. Inside the triangle $T_1T_2T_3$ the positivity conditions (21) are satisfied. Within this triangle, the hatched region where M of (27) is negative lies between the straight line BDF [where the denominator of (27) is zero] and the ellipse T_1ADG [where the numerator of (27) is zero]. The ellipse is tangent to BDF at D and to T_2T_3 at A. This region of negative M is entirely within the unallowed region (24) for $D_6(II)$ bounded by the straight lines CDH and ADE. The coordinates of the points are $T_1 = (1,0,0)$, $T_2 = (0,1,0)$, $T_3 = (0,0,1)$, $A = (0,\frac{121}{115},\frac{-5}{116})$, $B = (0,\frac{7139}{7474},\frac{335}{7474})$, $C = (0,\frac{1573}{1824},\frac{251}{1874})$, $D = (\frac{1}{3},\frac{1}{3},\frac{1}{3})$, $E = (\frac{121}{247},0,\frac{126}{247})$, $F = (\frac{7139}{13943},0,\frac{6804}{13943})$, $G = (\frac{363}{706},0,\frac{343}{706})$, $H = (\frac{1573}{2896},0,\frac{1323}{2896})$.

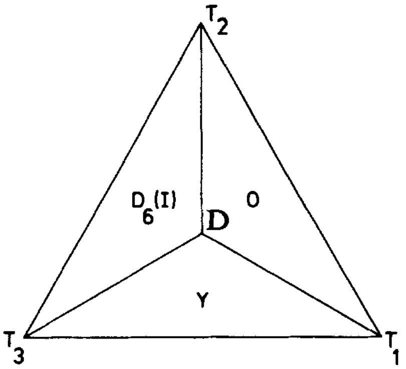


FIG. 2. In the variables P_1, P_2, P_3 defined in (21) cut by the plane $P_1+P_2+P_3=1$, the regions of the parameters for which the absolute minimum are $D_6(I)$, O and Y are, respectively, the triangle (T_2T_3D), (T_3T_1D), and (T_1T_2D). The coordinates of T_i and D are given in Fig. 1. On the line DT_1 the absolute minima coincide not only for O and Y but also for $D_6(II)$. At the point D the four minima of $Y, O, D_6(I)$ and $D_6(II)$ all coincide.

Indeed, in case (ii) the component d_1 defined above [Eq. (17)] is such that (remember $d_3=-d_2=7$ and $r=1$)

$$d_1^2=49[(27k+46k_1)/(18k+31k_1)]\,, \tag{23}$$

and obviously the solution does not exist in the region delimited by

$$\frac{-27}{46}\leqslant\frac{k_1}{k}\leqslant\frac{-18}{31}\,. \tag{24}$$

Here we write only the values of $\hat{Q}$ and of V that read

$$\hat{Q}=16(231+2d_1^2)\,, \tag{25}$$

$$V(D_6(II))=-\mu^2/M\,, \tag{26}$$

$$M\equiv q+\frac{648k^2+1523kk_1+706k_1^2}{2(972k+1667k_1)}\,. \tag{27}$$

The condition $M>0$ (necessary for asymptotic positivity) does not affect the region defined in Eq. (21). Indeed, once written in terms of the variables P_i, the numerator and denominator of M [see Eq. (27)] represent, respectively, an elliptical cone and a plane. The plane is tangent to the cone exactly on the diagonal line $P_1=P_2=P_3$. Within region (21), the subregion (hatched on Fig. 1) between the cone and the plane is the only region where M is negative. This region is entirely inside the region defined by the two planes (24) (which again cross exactly on the diagonal line) where $D_6(II)$ is not defined and thus harmless. The statements are summarized in Fig. 1 drawn in the plane $P_1+P_2+P_3=1$. Hence the result.

Direct comparison between the different values (20) and (26) allows the classification of the solutions corresponding to the absolute minimum for $D_6(I)$, O, and Y as a function of the parameters P_i and hence by (22) of the parameters q, k, and k_1 of the potential. The plot of the regions is particularly simple in the variables P_i as can be seen in Fig. 2.

TABLE III. The table contains all the possible stability groups for the potentials containing even powers of the field S and for $1\leqslant s\leqslant 6$.

Spin	Stability group in O(3)		
1	O(2)		
2	Accidental symmetry O(5)		
3	D_{3h}	T_d	
4	D_{4h}	O_h	
5	D_{5h}	C_{5v}, C_{4v}, C_{2v}	
6	D_{6h}	O_h	Y_h

As can be checked analytically, $D_6(II)$ occurs as an absolute minimum only on the line DT_1 of Fig. 2 where it coincides with Y and O leading to a triple degeneracy of the minimum. The point D, where in fact $k=k_1=0$, is even more singular as all the four minima coincide.

To end our analysis, we have explored numerically the region of positivity (21) and computed the absolute minimum for many arbitrary choices of the parameters. We have done this very systematically. In all instances the minimization program has produced the potential corresponding to one of the three predicted analytically in Fig. 2. Hence we feel that we can safely conclude, though an analytical proof is still lacking, that all the other possible stability groups do not produce absolute minima.

IV. CONCLUSIONS

The investigation reported here might seem very particular, since we specialize only in two representations and incomplete, since the complexity of the equations does not allow for a full classification of the extrema of the potentials considered. However, there are several points that should be stressed because they could play a role in more general investigations about Landau potentials.

First, the observation that the manifold of certain extrema (for instance of set of D_2 invariant solutions in the case $s=5$) can be a strata, not only an orbit.

Second, the existence of extremal configurations invariant under completely different subgroups and corresponding to identical values of the potential.

Third, the explicit construction of two quite different orbits of extrema invariant under the same subgroup of SO(3).

If we specialize into the stability groups of the absolute minima of the potentials, we see for $s=4, 5, 6$ that they are not necessarily maximal subgroups in SO(3) or O(3). For the cases of potentials involving only even powers of the fields, the results are summarized in Table III. The full stability groups are given [i.e., as part of O(3), see the notations in Ref. 3] and one can observe some regularity from these few cases.

[1]Y. Brihaye and J. Nuyts, J. Math. Phys. **28**, 1901 (1987).
[2]L. Michel, in *Regards sur la Physique Contemporaine* (CNRS, Paris, 1980).
[3]L. Michel, Rev. Mod. Phys. **52**, 638 (1980).

Similarity transformations of irreducible corepresentations in Wigner canonical form

P. Kasperkovitz
Institut für Theoretische Physik, TU Vienna, Wiedner Hauptstr. 8-10, A-1040 Vienna, Austria
M. K. Peev
Department of Solid State Physics, Sofia University, A. Ivanov Blvd. 5, BG-1126 Sofia, Bulgaria

(Received 29 March 1989; accepted for publication 24 January 1990)

The general form of the matrices that transform an irreducible corepresentation (coirrep) into an equivalent one, where both representations are assumed to be in the form proposed by Wigner, is analyzed for the three types of coirreps. In addition, the relation of these transformations to inner automorphisms of the corresponding group algebras is clarified.

I. INTRODUCTION

Recently, a method has been presented that systematizes and simplifies the calculation of reducing matrices,[1,2] especially Clebsch–Gordan[3] and subduction matrices.[4] The scheme proposed in Refs. 1–4 is based on a group of operations that transform irreducible representations (irreps), or corepresentations (coirreps), within certain sets and makes explicit use of the similarity transformations involved. In an extension of this work to isoscalar matrices[5] it became clear that a more detailed knowledge of the similarity transformations for coirreps (inherent ambiguity, possible standard forms, etc.) was needed. Since such an analysis does not seem to exist in literature, it is given in the present paper.

Corepresentations are matrix representations of ("magnetic") groups $G(H)$ that possess subgroups H of index 2. The multiplication law of these matrices reads

$$D(g_1g_2) = D(g_1)D(g_2)^{(g_1)}, \tag{1}$$

where the matrices $A^{(g)}$ are related to the matrices A according to the following rule:

$$A^{(g)} = \begin{cases} A, & \text{for } g\in H, \\ A^*, & \text{for } g\in G\setminus H. \end{cases} \tag{2}$$

In quantum mechanical problems, the matrices $D(h)$, $h\in H$, and $D(a)$, $a\in G\setminus H$, represent linear and antilinear operators, respectively; accordingly the corresponding group elements are often denoted as "unitary" and "antiunitary." It was Wigner[6] who first observed that every coirrep of a magnetic group $G(H)$ belongs to one of three classes. The type of a given coirrep is uniquely determined by the algebraic structure of the corresponding group algebra or its commutator algebra (see Ref. 7 and Sec. II below). Wigner[6] also proposed for each of the three types a form of the matrices that emphasizes the distinction between unitary and antiunitary elements. Every coirrep can be transformed into such a "canonical" form[8] and this form is always chosen in applications of the theory (see, e.g., Ref. 9). The three forms are given in the following list, where $h\in H$, $a\in G\setminus H$, Γ is a unitary irrep of H, A^T is the transpose of A, and $0'$ is the null matrix of dimension $\frac{1}{2}\dim D$:

$$\text{type I: } D(h) = \Gamma(h),\quad D(a) = Z(a),$$
$$Z(a^{-1}) = Z(a)^T,$$
$$Z(a_1)\Gamma(h)^*Z(a_2)^* = \Gamma(a_1ha_2).$$

$$\text{type II: } D(h) = \begin{pmatrix} \Gamma(h) & 0' \\ 0' & \Gamma(h) \end{pmatrix}, \tag{3}$$
$$D(a) = \begin{pmatrix} 0' & Z(a) \\ -Z(a) & 0' \end{pmatrix},$$
$$Z(a^{-1}) = -Z(a)^T,$$
$$Z(a_1)\Gamma(h)^*Z(a_2)^* = -\Gamma(a_1ha_2).$$

$$\text{type III: } D(h) = \begin{pmatrix} \Gamma(h) & 0' \\ 0' & \Gamma(a_0^{-1}ha_0)^* \end{pmatrix}, \tag{4}$$

$$D(a) = \begin{pmatrix} 0' & \Gamma(aa_0) \\ \Gamma(a_0^{-1}a)^* & 0' \end{pmatrix}. \tag{5}$$

In coirreps of types I and II all antiunitary elements have the the same status because matrices $Z(a)$ with the desired properties have to be found for all $a\in G\setminus H$. For coirreps of type III, on the other hand, the element a_0, selected from $G\setminus H$ by a convention, plays a special role since it fixes the form of the matrices $D(a)$.

In the following, we discuss the general form of similarity transformations $D(g)\to S^\dagger D(g)S^{(g)}$ induced by unitary matrices S. To this end we introduce in Sec. II the group algebras of the coirreps, i.e., we consider not only the matrices $D(g)$ but also complex linear combinations of these matrices. The structure of these algebras and their commutator algebras has been clarified by Dyson in a fundamental paper[7] but there all relations were derived in terms of real matrices of larger dimension. In Sec. II, we reformulate these results in terms of complex matrices as they occur in quantum mechanical problems. In Sec. III, similarity transformations are discussed as automorphisms of the group algebra and related to the inner automorphisms of this algebra. The conclusions from this discussion are summarized in Sec. IV.

II. COMPLEX GROUP ALGEBRAS

We consider a fixed coirrep D of some compact group $G(H)$ and assume that it has one of the canonical forms (3)–

 0022-2488/90/061304-06$03.00

(5). Starting from this we define three sets of matrices:

M^{lin} = set of complex linear combinations of the matrices

$$D(h), \quad h\in H, \tag{6}$$

M^{anti} = set of complex linear combinations of the matrices

$$D(a), \quad a\in G\setminus H, \tag{7}$$

M^{com} = commutator algebra of D. (8)

The matrices in the third set are characterized by the following relation:

$$T\in M^{\text{com}} \Leftrightarrow \begin{cases} TA = AT, & \text{for all } A\in M^{\text{lin}}, \\ TB = BT^*, & \text{for all } B\in M^{\text{anti}}. \end{cases} \tag{9}$$

Note that (9) and the canonical form of the matrices $A\in M^{\text{lin}}$ and $B\in M^{\text{anti}}$ fix the canonical form of the matrices $T\in M^{\text{com}}$ given in Eqs. (23), (27), and (30) below, and vice versa. The sets M^{lin} and M^{anti} are vector spaces over $\mathscr{C}$ (field of complex numbers), whereas M^{com} is a vector space over $\mathscr{R}$ (field of real numbers). Moreover

$$\begin{aligned} &A\in M^{\text{lin}}, B\in M^{\text{lin}} \Rightarrow AB\in M^{\text{lin}}, \\ &A\in M^{\text{lin}}, B\in M^{\text{anti}} \Rightarrow AB\in M^{\text{anti}}, \\ &A\in M^{\text{anti}}, B\in M^{\text{lin}} \Rightarrow AB^*\in M^{\text{anti}}, \\ &A\in M^{\text{anti}}, B\in M^{\text{anti}} \Rightarrow AB^*\in M^{\text{lin}}. \end{aligned} \tag{10}$$

The set underlying the definition of the group algebra **A** consists of all *ordered pairs* (A,B), $A\in M^{\text{lin}}$, $B\in M^{\text{anti}}$.

Next, we define the relations needed for **A** to become an algebra. *Addition* is defined by

$$(A,B) + (C,D) = (A+B, C+D) \tag{11}$$

and *multiplication* by

$$(A,B)(C,D) = (AC + BD^*, AD + BC^*). \tag{12}$$

Addition is associative and commutative, mulitplication is associative, and the distributive law holds. The *real multiples* of elements of **A** are defined by

$$r(A,B) = (rA, rB), \quad \text{for all } r\in\mathscr{R}; \tag{13}$$

this entails

$$(A,B)[r(C,D)] = [r(A,B)](C,D) = r[(A,B)(C,D)]. \tag{14}$$

Because of (11) and (12) **A** is a ring; because of (11) and (13) **A** is a vector space over $\mathscr{R}$; and because of (11)–(14) **A** is an algebra over $\mathscr{R}$.

The *involution* $(A,B)\to(A,B)^{\#}$, where

$$(A,B)^{\#} = (A^{\dagger}, B^{T}) \tag{15}$$

is an antiautomorphism of **A** (order of products reversed). This antiautomorphism can be used to prove that **A** is semisimple. That **A** is even a simple algebra follows from the irreducibility of the coirrep D.

According to Wedeerburn's structure theorem[10] every simple algebra **A** that is a finite-dimensional vector space over $\mathscr{R}$ is isomorphic to a full matrix algebra over a skewfield $\mathscr{F}$ which contains $\mathscr{R}$ as a subfield. That is, every element (A,B) of the group algebra **A** may be uniquely related to a finite matrix with elements taken from a skewfield $\mathscr{F}$. The skewfield is fixed by the type of D and isomorphic to its commutator algebra,

$$\mathscr{F} \cong M^{\text{com}}. \tag{16}$$

The dimension of the matrices depends both on the type of the coirrep D and on its dimension:

$$n = \dim D. \tag{17}$$

A full matrix algebra over a field $\mathscr{F}$ ($\supseteq\mathscr{R}$) has a natural basis from which all elements may be obtained by forming linear combinations over $\mathscr{R}$. It consists of "matrix units" $E_{s,t}$, that are matrices with the element 1 in position s, t, and 0 elsewhere, and a few "number units" chosen in such a way that each number in $\mathscr{F}$ may be represented as a real linear combination of these selected numbers (e.g., $a + ib\in\mathscr{C}$). Although this choice of a basis is the most obvious one, there exist infinitely many equivalent ones: the matrix units $E_{s,t}$ may be transformed with a fixed nonsingular matrix and the number units may be chosen differently [e.g., $a + (-i)b\in\mathscr{C}$]. Because fo the isomorphism, the group algebra **A** must possess similar bases. Each of them consists of "matrix units" $\mathbf{e}_{x,y}$ satisfying

$$\mathbf{e}_{x,y}\mathbf{e}_{u,v} = \delta_{y,u}\mathbf{e}_{x,v}, \tag{18}$$

$$\mathbf{e}^{\#}_{x,y} = \mathbf{e}_{y,x}, \tag{19}$$

and "number units" $\mathbf{f}_\sigma$ that satisfy

$$\mathbf{f}_0 = \mathbf{e} = \sum_x \mathbf{e}_{x,x} = (E,0) \quad (1\text{-element of } \mathbf{A}), \tag{20}$$

$$\mathbf{f}_\sigma^2 = -\mathbf{e} \text{ and } \mathbf{f}^{\#}_\sigma = -\mathbf{f}_\sigma \text{ for } \sigma\neq 0. \tag{21}$$

Moreover,

$$\mathbf{f}_\sigma\mathbf{e}_{x,y} = \mathbf{e}_{x,y}\mathbf{f}_\sigma. \tag{22}$$

Before giving one basis for each of the three classes we want to recall the corresponding structure theorem for ordinary representations. There the matrix basis consists of the matrices mentioned before (one element equal to 1, the other ones equal to 0) and the dimension of these matrices coincides with that of the irrep. That these matrices may be represented as complex linear combinations of the matrices $D(g)$, $g\in G$, follows from the orthogonality relations for irreps.[8] The field $\mathscr{F}$ from which the matrix elements are taken is always $\mathscr{C}$ if the set underlying the definition of the group algebra consists of the *complex* linear combinations of the matrices $D(g)$. This holds true irrespectively of the algebraic type of the irrep that refers to the smallest extension of $\mathscr{R}$ over which the irrep is absolutely irreducible.[6,7,11]

In the following list, the matrices $E_{s,t}$ are matrix units of the irrep Γ contained in the correp D. The primed matrices occurring for types II and III coireps are submatrices of dimension $\dim\Gamma = n/2$.

Type I: $\mathscr{F} = \mathscr{R}$, dimension over $\mathscr{R} = 4n^2$,

$$C\in M^{\text{com}} \Leftrightarrow C = aE, \quad a\in\mathscr{R}, \tag{23}$$

$$\mathbf{e}_{s\sigma,t\tau} = \mathbf{e}_{s,t}\breve{\mathbf{e}}_{\sigma,\tau} = \breve{\mathbf{e}}_{\sigma,\tau}\mathbf{e}_{s,t}, \tag{24}$$

$$s,t = 1,\ldots,n: \quad \mathbf{e}_{s,t} = (E_{s,t}, 0),$$

$$\breve{e}_{1,1} = \left(\frac{1}{2}E, \frac{1}{2}E\right), \quad \breve{e}_{1,2} = \left(-\frac{i}{2}E, \frac{i}{2}E\right), \tag{25}$$

$$\breve{e}_{2,1} = \left(\frac{i}{2}E, \frac{i}{2}E\right), \quad \breve{e}_{2,2} = \left(\frac{1}{2}E, -\frac{1}{2}E\right). \tag{26}$$

That the elements (26) can be obtained by forming suitable complex linear combinations of the matrices $D(h)$ and $D(a)$ can be seen as follows. Choose $(0,D(a))$, where a is an arbitrary but fixed element in $G\setminus H$. Since Γ is an irrep of H, $(X,0)\in\mathbf{A}$ for all complex matrices X. Let $X_1 = D(a)^{-1}$; then $(D(a)^{-1},0)(0,D(a)) = (0,E)\in\mathbf{A}$. The elements (26) are linear combinations of this element and $(X_2,0) = (E,0)$.

Type II: $\mathscr{F} = \mathscr{Q}$, dimension over $\mathscr{R} = n^2$,

$$C\in M^{\rm com}\Leftrightarrow C=\begin{pmatrix} aE' & bE' \\ -b^*E' & a^*E' \end{pmatrix},\quad a,b\in\mathscr{C}, \tag{27}$$

$$s,t=1,\ldots,n/2:\ e_{s,t} = (E'_{s,t}\oplus E'_{s,t},0), \tag{28}$$

$$\mathbf{f}_1 = \mathbf{i} = (iE,0),$$

$$\mathbf{f}_2 = \mathbf{j} = (0,J),\ \text{where } J=\begin{pmatrix} 0' & E' \\ -E' & 0' \end{pmatrix}, \tag{29}$$

$$\mathbf{f}_3 = \mathbf{k} = (0,iJ) = \mathbf{ij}.$$

The existence of the element $\mathbf{j}\in\mathbf{A}$ follows from similar arguments as the existence of the element $(0,E)$ for type I coirreps.

Type III: $\mathscr{F} = \mathscr{C}$, dimension over $\mathscr{R} = 2n^2$,

$$C\in M^{\rm com}\Leftrightarrow C = aE'\oplus a^*E',\quad a\in\mathscr{C}, \tag{30}$$

$$T=\begin{pmatrix} 0' & E' \\ E' & 0' \end{pmatrix}\in M^{\rm anti}, \tag{31}$$

$$\begin{aligned} s,t=1,\ldots,\frac{n}{2}:\ \mathbf{e}_{s1,t1} &= (E'_{s,t}\oplus 0',0),\\ \mathbf{e}_{s1,t2} &= (0,T[0'\oplus E'_{s,t}]),\\ \mathbf{e}_{s2,t1} &= (0,T[E'_{s,t}\oplus 0']),\\ \mathbf{e}_{s2,t2} &= (0'\oplus E'_{s,t},0), \end{aligned} \tag{32}$$

$$\mathbf{f}_1 = \mathbf{i} = (iE'\oplus(-i)E',0). \tag{33}$$

Existence of the elements $e_{s1,t1}$ and $e_{s2,t2}$ follows from the fact that the two irreps of H contained in $D\downarrow H$ are inequivalent.[8] This implies that all complex matrices of the form $A'\oplus B'$ belong to $M^{\rm lin}$. If we multiply $\Gamma(a_0^{-2})\oplus E'\in M^{\rm lin}$ with the matrix $D(a_0)\in M^{\rm anti}$ [cf. Eq. (5)] we obtain the matrix T, Eq. (31); this guarantees the existence of the units $\mathbf{e}_{s1,t2}$ and $\mathbf{e}_{s2,t1}$.

In the following discussion, the general form of the elements of $\mathbf{A}$ is of importance. Which matrices A and B can occur in an element $(A,B)\in\mathbf{A}$ follows from the canonical form of the coirrep, Eqs. (3)–(5), and is most easily seen from the bases given above. While for type I coirreps all pairs of complex matrices are admitted one finds for the other two types the following restrictions (A' etc. are arbitrary complex matrices of dimension $n/2$):

$$\text{type II:}\ (A,B)\in\mathbf{A}\Leftrightarrow A=\begin{pmatrix} A' & 0' \\ 0' & A' \end{pmatrix},\ B=\begin{pmatrix} 0' & B' \\ -B' & 0' \end{pmatrix}. \tag{34}$$

$$\text{type III:}\ (A,B)\in\mathbf{A}\Leftrightarrow A=\begin{pmatrix} A'_1 & 0' \\ 0' & A'_2 \end{pmatrix},\ B=\begin{pmatrix} 0' & B'_1 \\ B'_2 & 0' \end{pmatrix}. \tag{35}$$

As the set of complex matrices of dimension m has dimension $2m^2$ over $\mathscr{R}$, the dimension of $\mathbf{A}$ over $\mathscr{R}$ listed above is also evident from the form of the matrices A and B that are admitted in the elements $(A,B)\in\mathbf{A}$.

Before we start to discuss automorphisms of $\mathbf{A}$, especially those induced by similarity transformations, we introduce two subalgebras of $\mathbf{A}$ that are needed in this discussion. The first subalgebra is

$$\mathbf{A}^{\rm lin} = \{(A,0)\,|\,A\in M^{\rm lin}\}, \tag{36}$$

and the corresponding decomposition of $\mathbf{A}$ reads

$$\mathbf{A} = \mathbf{A}^{\rm lin} + \mathbf{A}^{\rm lin}\mathbf{b},\quad \mathbf{b}=(0,B),\quad B\ \text{nonsingular.} \tag{37}$$

A possible choice of coset representatives is the following:

$$\text{type I:}\quad B_0 = E, \tag{38}$$

$$\text{type II:}\quad B_0 = J, \tag{39}$$

$$\text{type III:}\quad B_0 = T. \tag{40}$$

For types I and II coirreps, $\mathbf{A}^{\rm lin}$ is isomorphic to a full complex matrix algebra, its dimension over $\mathscr{C}$ being n^2 (type I) or $(n/2)^2$ (type II). For type III coirreps $\mathbf{A}^{\rm lin}$ is isomorphic to a direct sum of two complex matrix algebras, each of dimension $(n/2)^2$, as is evident from (35).

The second subalgebra of $\mathbf{A}$ is

$$\mathbf{Z}^{\rm lin} = \text{center of } \mathbf{A}^{\rm lin}, \tag{41}$$

$$\text{types I and II: } \mathbf{Z}^{\rm lin} = \{(cE,0)\,|\,c\in\mathscr{C}\}, \tag{42}$$

$$\text{type III: } \mathbf{Z}^{\rm lin} = \{(c_1E'\oplus c_2E',0)\,|\,c_1,c_2\in\mathscr{C}\}. \tag{43}$$

III. AUTOMORPHISMS AND SIMILARITY TRANSFORMATIONS

An *automorphism* of the group algebra is a bijective mapping α: $\mathbf{A}\to\mathbf{A}$ that satisfies the following relations ($\mathbf{s},\mathbf{t},\ldots\in\mathbf{A}$, $r\in\mathscr{R}$):

$$\begin{aligned} \alpha(\mathbf{s}+\mathbf{t}) &= \alpha\mathbf{s}+\alpha\mathbf{t}, & \alpha(\mathbf{s},\mathbf{t}) &= \alpha\mathbf{s}\alpha\mathbf{t},\\ \alpha(r\mathbf{s}) &= r\alpha\mathbf{s}, & \alpha(\mathbf{s}^\#) &= (\alpha\mathbf{s})^\#. \end{aligned} \tag{44}$$

The mapping $(A,B)\to(A^*,B^*)$ is easily seen to be an automorphism [cf. Eqs. (11)–(15)]. Other examples of automorphisms are the *inner automorphisms* induced by norm-preserving elements $\mathbf{u}$. These are elements of $\mathbf{A}$ that satisfy

$$\mathbf{u}^\#\mathbf{u} = \mathbf{u}\mathbf{u}^\# = \mathbf{e}; \tag{45}$$

the corresponding automorphism is

$$\alpha_{\mathbf{u}}\mathbf{s} = \mathbf{u}^\#\mathbf{s}\mathbf{u}. \tag{46}$$

Because of Wedderburn's theorem, automorphisms of $\mathbf{A}$ are uniquely related to automorphisms of a full matrix algebra over a (skew)field $\mathscr{F}\supseteq\mathscr{R}$; this topic has been discussed in full detail in Ref. 11.

In this paper, we are interested in *automorphisms preserving the canonical form* of the elements of $\mathbf{A}$. These are those automorphisms that satisfy the following relation (note the implication):

$$\alpha\mathbf{A}^{\rm lin} = \mathbf{A}^{\rm lin}\quad(\Rightarrow\alpha\mathbf{Z}^{\rm lin} = \mathbf{Z}^{\rm lin}). \tag{47}$$

Although $\mathbf{Z}^{\rm lin}$ is invariant this does not imply that the elements of this subalgebra have to be invariant under the automorphisms (47). In fact there exists one nontrivial transformation for coirreps of types I and II, and three of them for coirreps of type III:

types I and II: (a) $\alpha(iE,0) = (iE,0)$, (b) $\alpha(iE,0) = (-iE,0)$; (48)

type III:

$$\begin{aligned}
&\text{(a)} \quad \alpha(E'\oplus 0',0) = (E'\oplus 0',0),\ \alpha(iE,0) = (iE,0),\\
&\text{(b)} \quad \alpha(E'\oplus 0',0) = (0'\oplus E',0),\ \alpha(iE,0) = (iE,0),\\
&\text{(c)} \quad \alpha(E'\oplus 0',0) = (E'\oplus 0',0),\ \alpha(iE,0) = (-iE,0),\\
&\text{(d)} \quad \alpha(E'\oplus 0',0) = (0'\oplus E',0),\ \alpha(iE,0) = (-iE,0).
\end{aligned} \tag{49}$$

If one requires, in addition to (47), also that

$$\alpha(iE,0) = (iE,0), \tag{50}$$

the second half of the possibilities listed above is ruled out. An example of an automorphism that satisfies (47) but not (50) is an inner automorphism induced by $\mathbf{b}_0$ [cf. Eqs. (37)–(40)]. On the other hand, all inner automorphisms induced by elements $\mathbf{u}\in\mathbf{A}^{\rm lin}$ satisfy both (47) and (50).

There exists a second class of automorphisms for which both relations (47) and (50) hold true. Let S be a *unitary matrix* of dimension $n(=\dim\ D)$ with the following properties:

$$\begin{aligned}
&A\in M^{\rm lin}\Rightarrow S^{\dagger}AS\in M^{\rm lin},\\
&B\in M^{\rm anti}\Rightarrow S^{\dagger}BS^{*}\in M^{\rm anti}.
\end{aligned} \tag{51}$$

The mapping $D(g)\rightarrow S\dagger D(g)S^{(g)}$ is then called a *similarity transformation* which *preserves the* (*Wigner*) *canonical form* of the coirrep D. It is extended to an automorphism of $\mathbf{A}$ by the following relation:

$$\alpha_S(A,B) = (S^{\dagger}AS,S^{\dagger}BS^{*}). \tag{52}$$

The mapping $S\rightarrow\alpha_S$ is not invertible because

$$\alpha_{S_1} = \alpha_{S_2}\Leftrightarrow S_1 = S_2C \text{ with } C(\text{unitary})\in M^{\rm com}. \tag{53}$$

This shows that the whole ambiguity of a similarity transformation is contained in the unitary elements of the commutator algebra.

The inner automorphisms induced by elements of $\mathbf{A}^{\rm lin}$ and the automorphisms induced by similarity transformations both belong to a class of automorphisms characterized by Eqs. (47) and (50). These automorphisms form a subgroup *aut* of *Aut*, the group of all automorphisms of $\mathbf{A}$. In the following, it will be shown how each $\alpha\in aut$ may be related to an inner automorphism $\alpha_{\mathbf{u}}$, $\mathbf{u}\in\mathbf{A}^{\rm lin}$. As $\alpha_S\in aut$ for all similarity transformations S these results also show to what extent a general similarity transformation S, that preserves the characteristic block structure of the coirrep, may be replaced by an equivalent similarity transformation $U\in M^{\rm lin}$.

A. Irreducible corepresentations of type I

Let $\alpha\in aut$; then $\alpha\mathbf{e}_{s,t}\in\mathbf{A}^{\rm lin}$ and these elements satisfy the same relations as the elements $\mathbf{e}_{s,t}$ given in Eq. (25). As has been shown in Secs. 4 and 6 B of Ref. 11 it is possible to construct, starting from the elements $\mathbf{e}_{s,t}$ and $\alpha\mathbf{e}_{s,t}$, an element $\bar{\mathbf{u}}\in\mathbf{A}^{\rm lin}$ such that

$$\alpha\mathbf{e}_{s,t} = \bar{\mathbf{u}}^{\#}\mathbf{e}_{s,t}\bar{\mathbf{u}}. \tag{54}$$

Under the automorphism α the real algebra $\breve{\mathbf{A}}$ spanned by the four elements $\breve{\mathbf{e}}_{\sigma,\tau}$ is mapped onto the algebra $\alpha\breve{\mathbf{A}}$ that consists of all elements of $\mathbf{A}$ that commute with all elements (54). Both quadruples $\alpha\breve{\mathbf{e}}_{\sigma,\tau}$ and $\bar{\mathbf{u}}^{\#}\breve{\mathbf{e}}_{\sigma,\tau}\bar{\mathbf{u}}$ form a basis of this algebra. It is therefore again possible to find an element $\breve{\mathbf{u}}\in\alpha\breve{\mathbf{A}}$ such that

$$\alpha\breve{\mathbf{e}}_{\sigma,\tau} = \breve{\mathbf{u}}^{\#}\bar{\mathbf{u}}^{\sigma,\tau}\bar{\mathbf{u}}\breve{\mathbf{u}}. \tag{55}$$

Since $\breve{\mathbf{u}}\in\alpha\breve{\mathbf{A}}$ this element has to be a real linear combination of the elements $\alpha(E,0)$, $\alpha(iE,0)$, $\alpha(0,E)$, and $\alpha(0,iE)$ [cf. Eqs. (26)]. We also know that

$$\bar{\mathbf{u}}^{\#}(iE,0)\bar{\mathbf{u}} = (iE,0) = \alpha(iE,0) = \breve{\mathbf{u}}^{\#}[\alpha(iE,0)]\breve{\mathbf{u}}. \tag{56}$$

The first of these equations follows from $\bar{\mathbf{u}}\in\mathbf{A}^{\rm lin}$ and the multiplication law (12); the second holds by definition ($\alpha\in aut$); and the third one follows from (55). It is easily verified that all elements $\breve{\mathbf{u}}\in\alpha\breve{\mathbf{A}}$ which commute with $\alpha(iE,0)$ have to be of the form $\alpha(cE,0) = (cE,0)$ $(c\in\mathscr{C})$ and are therefore also elements of $\mathbf{A}^{\rm lin}$.

Given an automorphism $\alpha\in aut$ it is therefore always possible to find an element $\bar{\mathbf{u}}\breve{\mathbf{u}} = \mathbf{u}\in\mathbf{A}^{\rm lin}$ such that $\alpha = \alpha_{\mathbf{u}}$. Since the matrix U in $\mathbf{u} = (U,0)$ is a similarity transformation that is unique up to the sign, the element $\mathbf{u}$ is also fixed by α up to a sign.

B. Irreducible corepresentations of type II

For $\alpha\in aut$ the units $\mathbf{e}_{s,t} = (E'_{s,t}\oplus E'_{s,t},0)$ are mapped onto the new units $\alpha\mathbf{e}_{s,t} = (\hat{E}'_{s,t}\oplus\hat{E}'_{s,t},0)$. Knowing the matrices $E'_{s,t}$ and $\hat{E}'_{s,t}$, one can construct a matrix U' such that

$$\hat{E}'_{s,t} = U'^{\dagger}E'_{s,t}U'. \tag{57}$$

Set $\mathbf{u} = (U'\oplus U',0)$; then $\mathbf{u}\in\mathbf{A}^{\rm lin}$ and $\alpha\mathbf{e}_{s,t} = \mathbf{u}^{\#}\mathbf{e}_{s,t}\mathbf{u}$.

The elements $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$, defined by

$$\hat{\mathbf{i}} = \mathbf{u}^{\#}\mathbf{i}_{s,t}\mathbf{u}, \text{ etc.}, \tag{58}$$

commute with the elements (57) and lie therefore in the subalgebra $\alpha\mathbf{Q}$ formed by the real linear combinations of e, α**i**, α**j**, α**k**. Since the elements $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ satisfy the same relations as the elements α**i**, α**j**, α**k**, and $\alpha\mathbf{Q}\cong\mathscr{Q}$ there exists an element $\mathbf{q}\in\alpha\mathbf{Q}$ such that

$$\begin{aligned}
&\mathbf{q}^{\#}\mathbf{e}\mathbf{q} = \mathbf{e}, \quad \mathbf{q}^{\#}\hat{\mathbf{i}}\mathbf{q} = \alpha\mathbf{i},\\
&\mathbf{q}^{\#}\hat{\mathbf{j}}\mathbf{q} = \alpha\mathbf{j}, \quad \mathbf{q}^{\#}\hat{\mathbf{k}}\mathbf{q} = \alpha\mathbf{k}.
\end{aligned} \tag{59}$$

Now $\hat{\mathbf{i}} = \mathbf{i}$, as follows from (29) and (12), and $\alpha\mathbf{i} = \mathbf{i}$ because $\alpha\in aut$. Accordingly, $\mathbf{q}$ has to be of the form

$$\mathbf{q} = (cE,0), \quad c\in\mathscr{C},|c| = 1. \tag{60}$$

Including this phase factor in the matrix U we see that for every automorphism $\alpha\in aut$ there exist an element $\mathbf{u}\in\mathbf{A}^{\rm lin}$ such that $\alpha = \alpha_{\mathbf{u}}$. As the matrix $U\in M^{\rm lin}$ induces a similarity transformation it is unique up to right-multiplication with unitary matrices in $M^{\rm lin}\cap M^{\rm com}$. Since this set consists of the matrices $\pm E$ the matrix U is uniquely determined by the automorphism α up to a sign.

C. Irreducible corepresentations of type III

According to (49), each automorphism $\alpha\in aut$ belongs to one of two classes depending on the transformation of the elements $\mathbf{e}_1,\mathbf{e}_2\in\mathbf{A}^{\rm lin}$,

$$\mathbf{e}_1 = (E'\oplus 0',0), \quad \mathbf{e}_2 = (0'\oplus E',0). \tag{61}$$

Since

$$\mathbf{i} = i\mathbf{e}_1 - i\mathbf{e}_2, \tag{62}$$

the transformation law of the elements (61) is uniquely related to that of $\mathbf{i}$. Under an automorphism, this element is either invariant or transformed into $-\mathbf{i}$ because these elements are the only solutions of the following equations:

$$\mathbf{i}\in\mathbf{A}^{\text{lin}},\ \mathbf{i}^{\#}=-\mathbf{i},\ \mathbf{i}^2=-\mathbf{e},$$
$$\mathbf{i}\mathbf{e}_{xs,xt}=\mathbf{e}_{xs,xt}\mathbf{i},\ \text{for}\ x=1,2\ \text{and all}\ s,t=1,\ldots,n/2. \tag{63}$$

The same remark holds for the elements $\pm\mathbf{b}_0$, the only solutions of the following equations:

$$\mathbf{b}_0\in\mathbf{A}\setminus\mathbf{A}^{\text{lin}},\quad \mathbf{b}_0^{\#}=\mathbf{b}_0,\quad \mathbf{b}_0^2=\mathbf{e},$$
$$\mathbf{b}_0\mathbf{e}_{1s,1t}=\mathbf{e}_{2s,2t}\mathbf{b}_0,\quad \text{for all}\ s,t=1,\ldots,n/2. \tag{64}$$

Class (a): $\alpha\mathbf{e}_1=\mathbf{e}_1$. Given the images $\alpha\mathbf{e}_{xs,xt}$ of the units $\mathbf{e}_{xs,xt}$ for both $x=1$ and $x=2$ it is possible to construct unitary matrices U'_x such that

$$\alpha\mathbf{e}_{xs,xt}=\mathbf{u}^{\#}\mathbf{e}_{xs,xt}\mathbf{u}\ \text{for}\ x=1,2, \tag{65}$$

where

$$\mathbf{u}=(U,0),\quad U=U'_1\oplus U'_2,\quad U^{\dagger}U=UU^{\dagger}=E. \tag{66}$$

Each of the two matrices U'_1 and U'_2 is fixed by α up to a phase factor. Because of (65) the element $\mathbf{u}\alpha\mathbf{b}_0\mathbf{u}^{\#}$ is also a solution of Eqs. (64), i.e., $\mathbf{u}\alpha\mathbf{b}_0\mathbf{u}^{\#}=\sigma\mathbf{b}_0$, $\sigma=\pm 1$. If $\sigma=-1$ we replace the matrices U'_x by iU'_x which does not change relations (65) but the sign of σ. It is therefore always possible to choose the element $\mathbf{u}\in\mathbf{A}^{\text{lin}}$ in such a way that not only Eqs. (65) but also the following relation is satisfied:

$$\alpha\mathbf{b}_0=\mathbf{u}^{\#}\mathbf{b}_0\mathbf{u}. \tag{67}$$

If this element of $\mathbf{A}$ is known it can be used to obtain U'_1 from U'_2 and vice versa:

$$\alpha\mathbf{b}_0=\begin{pmatrix}0' & U_1'^{\dagger}U_2'^{*}\\ U_2'^{\dagger}U_1'^{*} & 0'\end{pmatrix}. \tag{68}$$

For each automorphism $\alpha\in aut$ the corresponding matrix U is therefore uniquely fixed up to the phase of the submatrix U'_1.

Class (b): $\alpha\mathbf{e}_1=\mathbf{e}_2$. The product of two automorphisms in class (b) is in class (a). It is therefore sufficient to fix one automorphism in class (b), say α_1 and to consider the products $\alpha_1\alpha_2$, $\alpha_2\in$class (a). A possible representative of class (b) is $\alpha_1=\alpha_R$, the automorphism induced by a similarity transformation of the following form:

$$R=\begin{pmatrix}0' & V'\\ W' & 0'\end{pmatrix},\quad R^{\dagger}R=RR^{\dagger}=E. \tag{69}$$

Note that such an automorphism is not an inner one because $R\in M^{\text{lin}}$ and $(0,R)$ induces an automorphism different from α_R. Which unitary matrices V' and W' are chosen for the representative automorphism α_R is a matter of convention. We chose $V'=W'=E'$, i.e., $R=T$, for the representative. Then $\alpha_T\alpha_U=\alpha_{TU}\in$class (b) if, and only if, U is of the form (66). As U varies over all matrices of the form (66), the matrix $TU=R$ varies over all matrices of the form (69). Since the image of $\mathbf{b}_0$ under the automorphism α_R,

$$\alpha_R\mathbf{b}_0=\begin{pmatrix}0' & W'^{\dagger}V'^{*}\\ V'^{\dagger}W'^{*} & 0'\end{pmatrix}, \tag{70}$$

relates the matrices V' and W', the matrix R is fixed by $\alpha\in$class (b) up to the phase of the submatrix V'.

The role of a_0. Up to now it has not been taken into account that the Wigner canonical form of coirreps of type III requires more than the block structure of the matrices $A\in M^{\text{lin}}$ and $B\in M^{\text{anti}}$ given in Eq. (35). If one is only interested in coirreps of the form (5) one has to specify a set of such coirreps that is contained in $\mathbf{A}$ in the following sense:

$$D_\kappa(h)\in M^{\text{lin}},\quad D_\kappa(a)\in M^{\text{anti}};$$
$$D_\kappa(g_1)D_\kappa(g_2)^{(g_1)}=D_\kappa(g_1g_2), \tag{71}$$
$$\text{trace}\ D_\kappa(h)=\chi(h),\ \text{for all}\ \kappa\in K. \tag{72}$$

There exists at least one set with these properties, namely the set $\{D\}$ containing only the coirrep D that has been used to define the group algebra $\mathbf{A}$ [cf. Eqs. (6)–(7)]; other sets will be discussed below. Assuming the automorphisms to be induced by similarity transformations, we now define a subgroup of aut by the following condition: For each $\alpha=\alpha_S\in aut_K$ and each $\kappa\in K$ there exists a $\kappa'\in K$ such that

$$S^{\dagger}D_\kappa(g)S^{(g)}=D_{\kappa'}(g). \tag{73}$$

This condition implies the following restriction for the submatrices U'_1 and V' that specify the automorphisms in class (a) and (b), respectively:

$$\text{case (a):}\ U_1'^{\dagger}\Gamma_\kappa(h)U'_1=\Gamma_{\kappa'}(h), \tag{74}$$
$$U_1'^{\dagger}U_2'^{*}=E'; \tag{75}$$
$$\text{case (b):}\ V'^{\dagger}\Gamma_\kappa(h)V'=\Gamma_{\kappa'}(a_0^{-1}ha_0)^{*}, \tag{76}$$
$$V'^{\dagger}\Gamma_\kappa(a_0^2)W'^{*}=E'. \tag{77}$$

Equation (75) shows that $U'_2=U_1'^{*}$ while (74) restricts the matrices U'_1 to a set that depends on the set $D(K)=\{D_\kappa|\kappa\in K\}$ under consideration. Equation (77) implies, first of all,

$$\Gamma_\kappa(a_0^2)=\Gamma(a_0^2),\quad \text{for all}\ \kappa,\kappa'\in K, \tag{78}$$

because it is assumed that the transformation R is applied to all members of the set. In addition, Eq. (77) shows that W' is related to V' through $W'=\Gamma(a_0^2)^TV'^{*}$. Equation (76) selects those transformations R that leave the set $D(K)$ invariant. We are therefore left with two possibilities: (i) $\Gamma_\kappa(a_0^2)\neq\Gamma_{\kappa'}(a_0^2)$ for some pair $\kappa,\kappa'\in K$. In this case it is impossible to find transformations of type (b) that leave this set $D(K)$ invariant; therefore $aut_K\subseteq$class (a). The maximal group of this kind is class (a) because all transformations $\alpha=\alpha_U\in$class (a) transform a coirrep of the form (5) into a coirrep of the same form. (ii) $\Gamma_\kappa(a_0^2)=\Gamma_{\kappa'}(a_0^2)$, for all $\kappa\in K$ or, equivalently, $D_\kappa(a_0)=D(a_0)$, for all $\kappa\in K$. In this case $aut_K\subseteq aut_0$ where aut_0 is defined by

$$aut_0\subseteq aut;\quad \alpha\in aut_0\Leftrightarrow\alpha(0,D(a_0))=(0,D(a_0)). \tag{79}$$

The automorphisms $\alpha\in aut_0$ are induced by those similarity transformations $S(=U$ or $R)$ for which $S^{\dagger}D(a_0)S^{*}=D(a_0)$. To what extent aut_0, the maximal group of this type, is restricted to a subgroup aut_K depends on the set $D(K)$.

IV. CONCLUSION

A similarity transformation of a coirrep D is a mapping $D(g)\rightarrow\hat{D}(g)=S^{\dagger}D(g)S^{(g)}$, where S is a unitary matrix and both D and $\hat{D}$ are of the canonical form proposed by Wigner [see Eqs. (3)–(5)]. Let D and $\hat{D}$ be given coirreps of the same dimension and $\mathrm{tr}\, D(h) = \mathrm{tr}\, \hat{D}(h)$ for all $h \in H$.

Type I. The matrix S is determined by the coirreps D and $\hat{D}$ up to a phase factor ± 1.

Type II. Any matrix S that transforms D into $\hat{D}$ may be factorized as $S = UC = CU$ where the unitary matrix U has the same block structure as the matrices $D(h)$ and the unitary matrix C is of the form (27). These two matrices are determined by S up to a common sign. As C belongs to the commutator algebra of D it is always possible to choose the similarity transformation in the form $S = U = U' \oplus U'$; this matrix is fixed by the pair D, $\hat{D}$ up to the sign.

Type III. For these type of coirreps, the similarity transformations fall into two classes depending on the relation between the matrices $D(h) = \Gamma_1(h) \oplus \Gamma_2(h)$ and $\hat{D}(h) = \hat{\Gamma}_1(h) \oplus \hat{\Gamma}_2(h)$. Class (a): $\Gamma_1 \sim \hat{\Gamma}_1$. The matrix S has the form $U' \oplus U'^*$ where the unitary submatrix U' is determined by D and $\hat{D}$ up to a phase factor. Class (b): $\Gamma_1 \sim \hat{\Gamma}_2$. These similarity transformations leave the "antiunitary" matrix $D(a_0)$ invariant. The matrix S has the form (69) where the unitary submatrix W' is related to the submatrix V' by (77).

The similarity transformations of a coirrep may be extended to the corresponding group algebra. Except for coirreps of type III and transformations of class (b) they may be identified with inner automorphisms induced by linear combinations of the matrices $D(h)$.

[1]R. Dirl, P. Kasperkovitz, M. I. Aroyo, J. N. Kotzev, and M. Angelova-Tjurkedjieva, J. Math Phys. **27**, 37 (1986).

[2]M. I. Aroyo, J. N. Kotzev, M. Angelova-Tjurkedjieva, R. Dirl, and P. Kasperkovitz, J. Math. Phys. **27**, 2236 (1986).

[3]R. Dirl, P. Kasperkovitz, M. I. Aroyo, J. N. Kotzev, and B. L. Davies, J. Math. Phys. **28**, 1947 (1987).

[4]J. N. Kotzev, M. K. Peev, M. I. Aroyo, R. Dirl, and P. Kasperkovitz, J. Math. Phys. **30**, 9 (1989).

[5]M. K. Peev, J. N. Kotzev, P. Kasperkovitz, R. Dirl, and M. I. Aroyo (unpublished).

[6]E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).

[7]F. J. Dyson, J. Math Phys. **3**, 1199 (1962).

[8]L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967).

[9]C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).

[10]C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Wiley, New York, 1962).

[11]P. Kasperkovitz, J. Math. Phys. **22**, 2417 (1981).

Shifted tableaux, Schur's Q functions, and Kronecker products of S_n spin irreps

M. A. Salam and B. G. Wybourne
Physics Department, University of Canterbury, Christchurch, New Zealand

(Received 30 November 1989; accepted for publication 7 January 1990)

Properties of shifted tableaux have been explored in order to improve the algorithm for the calculation of Q function outer products. A simple technique has been established for finding out the highest and lowest partitions in the expansion of Q function outer products. Using these techniques and Young's raising operators, we have completed the Kronecker product for S_n spin irreps.

I. INTRODUCTION

Sagan[1] and Worley[2] have developed a combinatorial theory of shifted tableaux. These tableaux play a similar role in the description of Schur's Q functions as do ordinary Young tableaux of S functions. Stembridge[3] has given a shifted analog of the Littlewood–Richardson rule appropriate to the evaluation of Q function outer products. He has also extended the theory of the resolution of the Kronecker products of a basic spin representation with the ordinary irreps of the symmetric group S_n.

In this paper we give a simplified presentation of the Stembridge algorithm of Q function outer products in which "dead partitions"[4] are automatically eliminated and the coefficients of the "live partitions" readily evaluated.

We also present a complete description for the evaluation of the inner products of Q functions and hence the Kronecker products of spin irreps of S_n. This both extends and simplifies the earlier work of Luan Dehuai and Wybourne.[5]

We follow closely the notations of Stembridge.[3] Sections II and III briefly review relevant properties of Q functions and shifted tableaux while Sec. IV establishes theorems relating to skew shifted tableaux and the requirements for a live partition. The inner product of Q functions is discussed in Sec. V and the resolution of Kronecker products of spin irreps of S_n in Sec. VI.

II. SHIFTED TABLEAUX

A shifted diagram is a diagonally adjusted Young diagram with the restriction that the $(i+1)$th row does not exceed the ith row. This condition ensures that partitions are restricted to those involving distinct parts. Let P' denote the ordered alphabet $\{1'<1<2'<2\cdots\}$. The letters $1',2'...$ we said to be *marked* and we denote an *unmarked* version of any $a\in P'$ by $|a|$. Let DP represent partitions into distrinct parts only, then for each $\lambda\in$ DP there is an associated shifted diagram defined as

$$D'_\lambda=\{(i,j)\in Z^2: i\leqslant j\leqslant\lambda_i+i-1,\ 1\leqslant i\leqslant\ell(\lambda)\}.$$

A shifted tableau T of shape λ is an assignment T: $D'_\lambda\to P'$ satisfying the following conditions:

(i) $T(i,j)\leqslant T(i+1,j),\quad T(i,j)\leqslant T(i,j+1);$

(ii) Each column has at most one $k(k=1,2,...)$; (1)

(iii) Each row has at most one $k'(k'=1',2',...)$.

The tableau T is said to have content

$$\gamma=(\gamma_1,\gamma_2,...)\text{ and }x^T=x^\gamma=x_1^{\gamma_1}x_2^{\gamma_2}\cdots,$$

where γ_i is the number of occurrences of $|i|$ in T.

We can define a generating function $Q_\lambda=Q_\lambda(x)$ in the variables $x_1, x_2, ...$ for each $\lambda\in$ DP such as

$$Q_\lambda(x)=\sum_{T:D'_\lambda\to P'}x^T, \tag{2}$$

where summation is over tableaux T.

Given a tableau T a *word* $w(T)=w_1w_2\cdots w_n$ is a sequence obtained by reading the rows of T from left to right, starting with the last row. Let $w^r=w_nw_{n-1}\cdots w_1$ denote the reverse of w and let $\hat{w}=\hat{w}_1\cdots\hat{w}_n$ denote the word obtained by inverting the marks of w, i.e., $\hat{2}=2'$ and $\hat{2}'=2$. Let $n_i(w,j)$ denote the number of occurrences of the letter i among $w_1\cdots w_j$ and $n_i(w,0)=0$. An extended word of T is the sequence defined by $e(T)=w^r\hat{w}$. The tableau S is said to satisfy the *shifted lattice property* if the extended word $e=e_1\cdots e_{2n}$ satisfies the following conditions for all $i>1$ and $0\leqslant j<2n$:

$$n_i(e,j)=n_{i-1}(e,j)\text{ implies }\begin{cases}e_{j+1}\neq i,i', & 0\leqslant j<n,\\ e_{j+1}\neq i,(i-1)', & n\leqslant j<2n\end{cases}. \tag{3}$$

III. SKEW Q FUNCTIONS AND OUTER PRODUCTS

The outer product of two Q functions Q_μ and Q_ν,[3] such that $\mu,\nu,\lambda\in$ DP can be written as

$$Q_\mu\cdot Q_\nu=\sum_\lambda 2^{[\ell(\mu)+\ell(\nu)-\ell(\lambda)]}f^\lambda_{\mu\nu}Q_\lambda, \tag{4}$$

where $\ell(\rho)$ are the *lengths* of the partitions (ρ) and the coefficients $f^\lambda_{\mu\nu}$ are positive integers. The same coefficients appear in the following expansion:

$$Q_{\lambda/\mu}=\sum_\nu f^\lambda_{\mu\nu}Q_\nu, \tag{5}$$

where the shifted skew diagram of shape λ/μ is a collection of boxes of the form $D'_{\lambda/\mu}:=D'_\lambda-D'_\mu$ provided $D'_\mu\subseteq D'_\lambda$. A shifted skew tableau S of shape λ/μ is an assignment $S: D'_{\lambda/\mu}\to P'$ satisfying the rules of shifted tableaux.

The coefficient $f^\lambda_{\mu\nu}$ is defined as the number of shifted

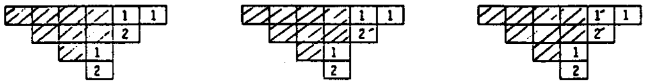
 0022-2488/90/061310-05$03.00

tableaux S of shape λ/μ and content ν such that

(i) S satisfies the shifted lattice property;

(ii) The leftmost i of $w(S)$ is unmarked $(1 \leqslant i \leqslant \ell(\nu))$. (6)

As an example, if $\lambda \equiv 6421$, $\mu \equiv 431$, and $\nu \equiv 32$, then we obtain the following three shifted tableaux of shape $Q_{6421}/_{431}$ and content 32 that satisfy (6):



Hence $f^{6421}_{431,32} = 3$.

IV. PROPERTIES OF SKEW SHIFTED TABLEAUX

In the previous section we observe that skew shifted tableaux play an important role in skew Q functions and Q-function outer products.

In this section we explore some of the important properties of skew shifted tableaux that will simplify the algorithms for the calculation of the coefficients $f^{\lambda}_{\mu\nu}$ and lead to algorithms for finding the highest and the lowest partitions in the expansion of a Q-function outer product.

A partition $\nu = (\nu_1\nu_2\cdots\nu_i)$ is *lower* than $\mu = (\mu_1\mu_2\cdots\mu_j)$ if for all $1 \leqslant k \leqslant j$, $\Sigma^k_{m=1}\mu_m \geqslant \Sigma^k_{m=1}\nu_m$ and $|\mu| = |\nu|$. Throughout this section we have $\nu, \mu, \lambda \in \mathrm{DP}$.

Theorem 1: In a skew shifted tableau of shape λ/μ and content ν no $|i|$ can be placed in the jth row such that $|i| > j$.

Proof: In the first part we prove that an entry $x > 1$ placed in the first row violates the shifted lattice property.

In order to satisfy (i) the largest entry $x > 1$ in the first row must be placed in the right-most position. This entry will ultimately appear at the first position of the extended work $e = e_1\cdots e_{2n}$. Noting (3) requires

$$n_x(e,0) = n_{x-1}(e,0) = 0,$$

but $e_1 = x$ which violates the shifted lattice property. Similarly if an entry $y > 2$ is placed in the second row we again obtain a violation of the shifted lattice property leading readily to the same conclusion for every row and hence Theorem 1.

Use of Theorem 1 makes it possible to eliminate most of the "dead tableaux" that do not satisfy the shifted lattice property.

Theorem 2: In a skew shifted tableau of shape λ/μ and content ν for all $1 \leqslant i \leqslant \ell(\lambda)$,

$$\sum_{k=1}^{i} \lambda_k \leqslant \sum_{k=1}^{i} (\mu_k + \nu_k). \tag{7}$$

Proof: If $\Sigma^i_{k=1}(\lambda_k - \mu_k)$ is greater than $\Sigma^i_{k=1}\nu_k$ for any value of i then we have to make $\Sigma^i_{k=1}(\lambda_k - \mu_k) - \Sigma^i_{k=1}\nu_k$ entries greater than i, which violates Theorem 1.

Corollary 1: The largest partition λ appearing in the expansion of the outer product of two Q functions Q_μ and Q_ν is given by

$$\lambda = \mu + \nu,$$

such that

$$\lambda_i = \mu_i + \nu_i,$$

for all values of i.

Proof: It is easily concluded from (7) that the maximum value of λ_i is obtained when only equality holds for all i.

Corollary 1 immediately gives the highest partition in the outer product of two Q functions. We now use the relationship between skew shifted tableaux and the shifted lattice property to establish the lowest live partition in a Q-function outer product.

Theorem 3: Let μ and ν be self-conjugate partitions, each with distinct parts. Then,

$$Q_\mu \cdot Q_\nu = 2^{\min[\ell(\mu),\,\ell(\nu)]} Q_\lambda, \tag{8}$$

where $\lambda = \mu + \nu$ and min $[\ell(\mu), \ell(\nu)]$ is the minimum length of the partitions (μ) and (ν).

Proof: Let $\nu = (\nu_1\nu_2\cdots\nu_{n-1}\nu_n)$ be the content of the skew shifted tableaux of shape λ/μ. It is clear that $\nu_n = 1$ and $\nu_{n-1} = 2$. Let x denote ν_{n-1} entries and y denote ν_n entries. Six possible types of extended words arise:

$$\begin{array}{llllllll} 1\cdots\cdots & j\cdots\cdots\cdots\cdots\cdots\cdots & & & k\cdots & 2n & & \\ & w^r & & & \hat{w} & & & \\ \cdots\; x'\cdots & y\cdots\cdots & x\cdots & x'\cdots\cdots & y'\cdots & x\cdots & & \text{(i)} \\ \cdots\; x\cdots & y\cdots\cdots & x\cdots & x'\cdots\cdots & y'\cdots & x' & & \text{(ii)} \\ \cdots\cdots\; y\cdots & x'\cdots & x\cdots & x'\cdots & x\cdots & y'\cdots & & \text{(iii)} \\ \cdots\cdots\; y\cdots & x\cdots & x\cdots & x'\cdots & x'\cdots & y'\cdots & & \text{(iv)} \\ x'\cdots\; x\cdots & y\cdots\cdots\cdots\cdots\cdots & & & y'\cdots & x'\cdots & x & \text{(v)} \\ x\cdots\; x\cdots & y\cdots\cdots\cdots\cdots\cdots & & & y'\cdots & x'\cdots & x'. & \text{(vi)} \end{array}$$

We can see that "y" and the right-most "x" of w^r are not marked because the left-most i in w is not allowed to be marked. Let "y" always appear at the jth position then (i), (iii), and (iv) do not satisfy the lattice property since

$$n_x(e,j-1) = n_y(e,j-1) = 0,$$

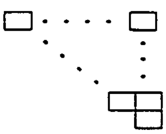
but $e_j = y$, when $y = x + 1$.

(v) does not satisfy the shifted lattice property because

$$n_x(e,k-1) = n_y(e,k-1) = 1, \text{ by } e_k = y.$$

The only surviving words are (ii) and (vi) both of which do not contain marked letters.

Successive applications of these arguments for other parts show that in such a situation we can not make any marked entry. As the partition μ is self-conjugate it forms a right-adjusted shifted diagram as shown below:



It should now be clear that if we are not permitted to make marked entries then all the 1's must be placed in the first row, all the 2's in the second row, etc. Thus $\lambda = \mu + \nu$. Since none of the entries can be marked or interchanged we conclude $f^{\lambda}_{\mu\nu} = 1$.

We have, for example,

$$Q_{4321} \cdot Q_{4321} = 16 Q_{8642}, \quad Q_{4321} \cdot Q_{54321} = 16 Q_{97531}.$$

We note that for the S-function outer product $\{4321\}\cdot\{4321\}$ involves 206 distinct Young tableaux and a multiplicity sum of 930.

Extending the above arguments for any pair μ and ν we establish the following algorithm to yield the lowest partition that arises in $Q_\mu \cdot Q_\nu$.

(i) Combine μ and ν in in nonascending order and denote by $\gamma = \gamma_1 \cdots \gamma_n$.

(ii) Find i for which we get a sequence such that $\gamma_i = \gamma_{i+1}$ or $\gamma_i = \gamma_{i+1} + 1$.

(iii) Find the numbers k of successive pairs of equal parts in this sequence provided $\gamma_j = \gamma_{j+1} = \gamma_{j+2}$ will be treated as one "pair."

(iv) Add k in γ_i and substract the same from γ_{i+1};

i.e., $\gamma = \gamma_1 \cdots \gamma_i + k, \quad \gamma_{i+1} - k, \ldots \gamma_n,$

and rearrange.

(v) Repeat steps (ii)–(iv) till the end of γ.

(vi) Repeat steps (ii)–(v) until a partition into distinct parts is obtained.

As an example of $\mu = 541$ and $\nu = 542$ then the lowest partition can be obtained by the use of the above algorithm:

$$\begin{aligned} \gamma &\equiv 554421, && \text{using (i)},\\ &\equiv 744321, && \text{using (ii)–(v)},\\ &\equiv 753321, && \text{using (ii)–(v)},\\ &\equiv 754221, && \text{using (ii)–(v)},\\ &\equiv 754311, && \text{using (ii)–(v)},\\ &\equiv 75432, && \text{using (ii)–(v)}. \end{aligned}$$

Making use of Corollary 1 the highest partition is 10 83. It follows that the only live partitions are those partitions ρ into distinct parts such that $10\,83 \geqslant \rho \geqslant 75432$. In the case of μ and ν being self-conjugate partitions, each into distinct parts, the highest and lowest live partitions coincide as expected from (8).

V. INNER PRODUCT OF Q FUNCTIONS

The inner product of two Q functions, $Q_\mu^* Q_\nu$ may be written as follows[6]:

$$Q_\mu^* Q_\nu = \sum_\lambda b_{\mu\nu}^\lambda S_\lambda, \tag{9}$$

where $|\mu| = |\nu| = |\lambda|$ and S_λ is an S function. The same coefficients $b_{\mu\nu}^\lambda$ appear in the expansion of the inner product of a Q function, Q_μ, and an S function, S_λ,[6] i.e.,

$$S_\lambda^* Q_\mu = \sum_\nu b_{\mu\nu}^\lambda 2^{-\ell(\nu)} Q_\nu. \tag{10}$$

There is no direct method of calculating the coefficients $b_{\mu\nu}^\lambda$. We can however calculate the coefficients $b_{\mu\nu}^\lambda$ in (10) by the following algorithm.

(i) Expand Q_μ in terms of S functions using the inverse raising operator

$$Q_\mu = \prod_{i<j} (1 - \delta_{ij} + \delta_{ij}^2 \cdots) S_\mu. \tag{11}$$

(ii) Compute the S-function inner product.

(iii) Convert the resulting list of S functions into Q functions using the raising operator

$$S_\lambda = \prod_{i<j} (1 + \delta_{ij}) Q_\lambda. \tag{12}$$

(iv) Multiply the coefficients of Q_ν by $2^{\ell(\nu)}$ to get $b_{\mu\nu}^\lambda$.

Instead of making the whole expansion of S function in terms of the Q function in step (iii) we use a direct method of computing the coefficient $b_{\mu\nu}^\lambda$ for any set of λ, μ, and ν. For this purpose we use the following analogy.

Stembridge[3] has defined the Kronecker product of basic spin representation φ^n and an ordinary irrep χ^μ of S_n as follows:

Let

$\lambda \in \mathrm{DP}$, and $|\mu| = n$,

then

$$\varphi^n * \chi^\mu = \sum_\lambda \frac{1}{\epsilon_\lambda \epsilon_n} 2^{[\ell(\lambda)-1]/2} g_{\lambda\mu} \varphi^\lambda,$$

where $g_{\lambda\mu}$ is the number of shifted tableaux S of unshifted shape λ and content μ such that (a) $w = w(s)$ satisfies the shifted lattice property; and (b) the left-most i of $|w|$ is unmarked in $w(1 \leqslant i \leqslant \ell(\lambda))$, and φ^λ is a spin representation.

The coefficients $g_{\lambda\mu}$ can easily be calculated using the techniques developed earlier.

VI. SPIN IRREPS OF S_n AND KRONECKER PRODUCT

Much has been written about spin irreps of the symmetry group,[3,5,7,8] but little is know about the Kronecker products of spin irreps of S_n. We will use the techniques developed in the previous section for this Kronecker product.

Let $\{\rho\}$ denote an ordinary irrep of S_n when $\rho = \rho_1 \rho_2 \cdots \rho_k$ is a partition of "n," and $[\Delta;\lambda]$ denote a spin irrep of S_n when Δ is basic spin representation. Here we are using reduced notation,[5] i.e., $\rho_1 = n - |\lambda|$. The inner product of an ordinary irrep $\{\lambda\}$ and a spin irrep $[\Delta;\mu]$ can easily be worked out using (10) and the algorithm given in the previous section. In order to convert the Q functions into representations, the right-hand side of (10) needs to be multiplied by $2^{[(\ell(\nu) - n(\mathrm{mod}\,2) + 1)/2]}$ and divided by $2^{[(\ell(\mu) - n(\mathrm{mod}\,2) + 1)/2]}$ when $[x]$ means only the integer part of x. We also observe that

$$[\Delta;\mu]_\pm^* \{\lambda\} = \tfrac{1}{2}([\Delta;\mu]^*\{\lambda\}). \tag{13}$$

If $n - k$ is odd then λ is called odd and spin irreps $[\Delta;\lambda]$ splits into an associate pair such as

$$[\Delta;\lambda] \equiv [\Delta;\lambda]_+ + [\Delta;\lambda]_-, \quad \text{for } n-k \text{ odd} \tag{14}$$

and

$$\Delta \equiv \Delta_+ + \Delta_-, \qquad \text{for } n \text{ even}. \tag{15}$$

Here we need difference characters for a complete resolution of the Kronecker product of the type $[\Delta;\lambda]_\pm^2$ or $[\Delta;\lambda]_\pm [\Delta;\lambda]_\mp$.

The difference character $[\Delta;\lambda]'$ is defined as

$$[\Delta;\lambda]' = [\Delta;\lambda]_+ - [\Delta;\lambda]_-. \tag{16}$$

From (14) we get

$$\begin{aligned} [\Delta;\lambda]^2 &= [\Delta;\lambda]_+ [\Delta;\lambda]_+ + [\Delta;\lambda]_- [\Delta;\lambda]_- \\ &\quad + 2[\Delta;\lambda]_+ [\Delta;\lambda]_-, \end{aligned} \tag{17}$$

and from (16) we get

$$[\Delta;\lambda]'^2 = [\Delta;\lambda]_+[\Delta;\lambda]_+ + [\Delta;\lambda]_-[\Delta;\lambda]_- - 2[\Delta;\lambda]_+[\Delta;\lambda]_-. \quad (18)$$

Adding (17) and (18)

$$[\Delta;\lambda]_\pm [\Delta;\lambda]_\pm = \tfrac{1}{4}([\Delta;\lambda]^2 + [\Delta;\lambda]'^2), \quad (19)$$

and subtracting (18) from (17)

$$[\Delta;\lambda]_\pm [\Delta;\lambda]_\mp = \tfrac{1}{4}([\Delta;\lambda]^2 - [\Delta;\lambda]'^2), \quad (20)$$

when

$$[\Delta;\lambda]_+ [\Delta;\lambda]_+ = [\Delta;\lambda]_- [\Delta;\lambda]_-,$$

and

$$[\Delta;\lambda]_+ [\Delta;\lambda]_- = [\Delta;\lambda]_- [\Delta;\lambda]_+.$$

For two different spin irreps of S_n, say $[\Delta;\mu]$ and $[\Delta;\nu]$, the Eqs. (19) and (20) reduce to

$$[\Delta;\mu]_\pm [\Delta;\nu]_\pm = [\Delta;\mu]_\pm [\Delta;\nu]_\mp = \tfrac{1}{4}[\Delta;\mu][\Delta;\nu]. \quad (21)$$

For any μ and ν including $\mu = \nu$ we get the following:

$$[\Delta;\mu]_\pm [\Delta;\nu] = \tfrac{1}{2}[\Delta;\mu][\Delta;\nu]. \quad (22)$$

Here, $[\Delta;\mu]$ $[\Delta;\nu]$ or $[\Delta;\lambda]^2$ can be calculated using the algorithms given in the previous section. In order to convert the Q functions into spin irreps we need to multiply the coefficients $b^{\lambda}_{\mu\nu}$ by a factor $(\epsilon_\mu\epsilon_\nu)^{-1} 2^{-([\ell(\mu)+\ell(\nu)+2])/2}$ when

$$\epsilon_\mu = \begin{cases} \sqrt{2}, & \text{if } \mu \text{ is odd,} \\ 1, & \text{if } \mu \text{ is even.} \end{cases}$$

For the calculation of difference characters let us consider $[\Delta;\lambda]'^2$. This will have nonzero characters only for the classes of (λ) and $(\Delta;\lambda)$. We can expand $[\Delta;\lambda]'^2$ in terms of ordinary irreps ρ of S_n as follows:

$$[\Delta;\lambda]'^2 = g^{\rho}_{\lambda\lambda}\{\rho\},$$

where $g^{\rho}_{\lambda\lambda} = 2i^{n-k+1}\chi^{\rho}_{\lambda}$.

The character χ^{ρ}_{λ} can easily be calculated using Littlewood's theorem (see Ref. 9, p. 70).

Some typical examples of products for the symmetric group that illustrate the above algorithms are given below.

Group is $S(10)$
[Δ;32]*[Δ;31]→
4{9} + 16{82} + 16{811} + 32{73} + 68{721} + 36{7111} + 36{64}
+ 124{631} + 88{622} + 140{6211} + 52{61111} + 16{55} + 108{541}
+ 168{532} + 216{5311} + 200{5221} + 176{52111} + 52{511111} + 92{442}
+ 112{4411} + 76{433} + 284{4321} + 200{43111} + 112{4222} + 216{42211}
+ 140{421111} + 36{4111111} + 76{3331} + 92{3322} + 168{33211}
+ 88{331111} + 108{32221} + 124{322111} + 68{3211111} + 16{31111111}
+ 16{22222} + 36{222211} + 32{2221111} + 16{22111111} + 4{211111111}

dimension = 1382400
dim [Δ;32] = 864
dim [Δ;31] = 1600;

$[\Delta;32]_\pm$*[Δ;31]→
2{91} + 8{82} + 8{811} + 16{73} + 34{721} + 18{7111} + 18{64} + 62{631}
+ 44{622} + 70{6211} + 26{61111} + 8{55} + 54{541} + 84{532}
+ 108{5311} + 100{5221} + 88{52111} + 26{511111} + 46{442} + 56{4411}
+ 38{433} + 142{4321} + 100{43111} + 56{4222} + 108{42211} + 70{421111}
+ 18{4111111} + 38{3331} + 46{3322} + 84{33211} + 44{331111}
+ 54{32221} + 62{322111} + 34{3211111} + 8{31111111} + 8{22222}
+ 18{222211} + 16{2221111} + 8{22111111} + 2{211111111}

dimension = 691200;

$[\Delta;32]_\pm$*$[\Delta;32]_\pm$ →
{10} + {91} + 4{82} + 4{811} + 5{73} + 11{721} + 7{7111} + 4{64}
+ 18{631} + 13{622} + 21{6211} + 8{61111} + 3{55} + 13{541} + 22{532}
+ 29{5311} + 27{5221} + 25{52111} + 9{511111} + 12{442} + 14{4411}

$$+ 9\{433\} + 36\{4321\} + 27\{43111\} + 14\{4222\} + 29\{42211\} + 21\{421111\}$$
$$+ 6\{4111111\} + 10\{3331\} + 11\{3322\} + 22\{33211\} + 13\{331111\} + 14\{32221\}$$
$$+ 17\{322111\} + 12\{3211111\} + 4\{31111111\} + \{22222\} + 5\{222211\}$$
$$+ 5\{2221111\} + 3\{22111111\} + 2\{211111111\}$$

dimension = 186624;

$[\Delta;32]_{\pm} * [\Delta;32]_{\mp} \to$
$$2\{91\} + 3\{82\} + 4\{811\} + 5\{73\} + 12\{721\} + 6\{7111\} + 5\{64\} + 17\{631\}$$
$$+ 13\{622\} + 21\{6211\} + 9\{61111\} + \{55\} + 14\{541\} + 22\{532\} + 29\{5311\}$$
$$+ 27\{5221\} + 25\{52111\} + 8\{511111\} + 11\{442\} + 14\{4411\} + 10\{433\}$$
$$+ 36\{4321\} + 27\{43111\} + 14\{4222\} + 29\{42211\} + 21\{421111\} + 7\{4111111\}$$
$$+ 9\{3331\} + 12\{3322\} + 22\{33211\} + 13\{331111\} + 13\{32221\} + 18\{322111\}$$
$$+ 11\{3211111\} + 4\{31111111\} + 3\{22222\} + 4\{222211\} + 5\{2221111\}$$
$$+ 4\{22111111\} + \{211111111\} + \{1111111111\}$$

dimension = 186624;

$[\Delta;32]_{\pm} * \{532\} \to$
$$23[\Delta;41]_{+} + 23[\Delta;41]_{-} + 36[\Delta;4] + 5[\Delta;321] + 22[\Delta;32]_{+} + 22[\Delta;32]_{-}$$
$$+ 42[\Delta;31]_{+} + 42[\Delta;31]_{-} + 43[\Delta;3] + 22[\Delta;21]_{+} + 22[\Delta;21]_{-} + 26[\Delta;2]$$
$$+ 8[\Delta;1] + [\Delta;0]_{+} + [\Delta;0]$$

dimension = 194400

dim {532} = 450;

$[\Delta;32] * \{532\} \to$
$$46[\Delta;41]_{+} + 46[\Delta;41]_{-} + 72[\Delta;4] + 10[\Delta;321] + 44[\Delta;32]_{+} + 44[\Delta;32]_{-}$$
$$+ 84[\Delta;31]_{+} + 84[\Delta;31]_{-} + 86[\Delta;3] + 44[\Delta;21]_{+} + 44[\Delta;21]_{-} + 52[\Delta;2]$$
$$+ 16[\Delta;1] + 2[\Delta;0]_{+} + 2[\Delta;0]_{-}$$

dimension = 388800.

VII. CONCLUSION

We have improved the algorithm for the calculation of the Q function outer product and established a simple technique for finding out the highest and lowest partitions in the expansion of the Q function outer product. We have completed the Kronecker product of spin irreps of S_n.

ACKNOWLEDGMENTS

We are grateful to Dr. John R. Stembridge for very valuable suggestions on Q-function inner products. We are also very thankful to Dr. Barry Searle for useful comments during programing and to Mary Kinnaird for preparing this manuscript.

One of us (B. G. W.) derived great benefit from the CAP-NSERC 1989 Summer Institute in Theoretical Physics at Kingston, Ontario, with special thanks to Professor D. J. Rowe and Professor B. Castel who made the Institute possible.

[1] B. E. Sagan, J. Combin. Theory Ser. A, **45**, 62–103 (1987).
[2] D. R. Worley, Ph.D. thesis, M.I.T., 1989.
[3] J. R. Stembridge, Adv. Math. **74**, 87–134 (1989).
[4] Y. M. Chen, A. M. Garsia, and J. Remmel, Contemp. Math. **34**, 109–153 (1989).
[5] Luan Dehuai and B. G. Wybourne, J. Phys. A: Math. Nuc. Gen. **14**, 327–348 (1981).
[6] J. R. Stembridge (private communication).
[7] A. O. Morris, Proc. London Math. Soc. **12**, 55 (1962).
[8] A. O. Morris, J. London Math. Soc. **37**, 445 (1962).
[9] D. E. Littlewood, *The Theory of Group Characters and Matrix Representations of Groups* (Oxford U. P., London, 1958).

Finite geometries and Clifford algebras. II

R. Shaw and T. M. Jarvis
School of Mathematics, University of Hull, Hull HU6 7RX, United Kingdom

(Received 15 November 1989; accepted for publication 31 January 1990)

The Clifford algebra in dimension $d = 2^{m+1} - 1$, $m \geqslant 2$, is treated using the finite m-dimensional projective geometry PG(m, 2) over the field of order 2. Full details are given for the case $m = 4$, $d = 31$, generalizing previous work for $m = 2$ and 3. Details are given of some configurations that arise in PG(4, 2).

I. INTRODUCTION

In a previous paper Shaw[1] treated the Clifford algebra of dimension $d = 2^{m+1} - 1$, $m \geqslant 2$, using the incidence properties of the m-dimensional projective geometry PG(m, 2) over the field $\mathbb{F}_2$ with two elements. Shaw's paper, which we refer to as I, gives full details in the case $m = 3$. The present paper deals with $m = 4$ and provides an affirmative answer to conjectures A and B in this case.

We recall some notation and definitions from I. The linear operators $\Gamma_1, \ldots, \Gamma_d$ satisfy the relations

$$(\Gamma_i)^2 = -I, \quad \Gamma_i\Gamma_j = -\Gamma_j\Gamma_i, \quad i \neq j, \tag{1.1}$$

and

$$\Gamma_1\Gamma_2 \cdots \Gamma_d = I. \tag{1.2}$$

Consider the finite group G_0 generated by the operators Γ_i:

$$G_0 = \{\pm I, \pm\Gamma_i, \pm\Gamma_i\Gamma_j, \pm\Gamma_i\Gamma_j\Gamma_k, \ldots\}. \tag{1.3}$$

The center $\{I, -I\}$ is also the commutator subgroup of G_0, so we can obtain the Abelian group

$$C_0 = G_0/\{I, -I\}. \tag{1.4}$$

Let s_i be the image of Γ_i under the canonical projection π: $G_0 \to C_0$. The elements $s_1, \ldots, s_d$ generate C_0 subject to the relations

$$(s_i)^2 = 1, \quad s_is_j = s_js_i, \quad s_1 \ldots s_d = 1. \tag{1.5}$$

Consider the set $S = \{s_1, \ldots, s_d\}$. A subset $\alpha \subseteq S$ is said to be a small subset, or a figure, if $|\alpha| \leqslant \frac{1}{2}(d-1)$. We can identify C_0 with the set of all figures of S, the multiplication being given by

$$\alpha\beta = \begin{cases} \alpha\Delta\beta, & \text{if } \alpha\Delta\beta \text{ is small,} \\ (\alpha\Delta\beta)^c, & \text{if } \alpha\Delta\beta \text{ is large,} \end{cases}$$

where $\alpha\Delta\beta$ is the symmetric difference.

Now view the elements s_i of S as the points of the finite projective geometry PG(m, 2), $d = 2^{m+1} - 1$. For any collection $\mathscr{F}$ of figures, let $\langle\mathscr{F}\rangle$ be the subgroup of C_0 generated by $\mathscr{F}$. We are interested in the subgroups $C_1, C_2, \ldots, C_{m-1}$ generated, respectively, by lines, planes, ..., hyperplanes in PG(m, 2). In other words, if $S_r(m)$ is the set of r-spaces in PG(m, 2), $C_r(m) = \langle S_r(m)\rangle$. [Of course we write C_r for $C_r(m)$ when the m is understood.]

From I we know that we can identify $S_0(m) \cup \{0\}$ with a vector space $V(m+1)$ of dimension $m+1$ over $\mathbb{F}_2$. We also know that $C_{m-1}(m)$ and $\widehat{V}(m+1)$ [= dual of $V(m+1)$] are isomorphic. Now choose a simplex of reference $\mathscr{S}$ whose vertices $v_1, \ldots, v_{m+1}$ correspond to a basis of $V(m+1)$. It is clear that the $(m-1)$-dimensional faces of $\mathscr{S}$ will correspond to the dual basis of $\widehat{V}(m+1)$ and hence that $C_{m-1}(m)$ is generated by the set of $(m-1)$-dimensional faces of $\mathscr{S}$.

Let $\mathscr{F}_r^m$ denote the set of r-dimensional faces of the chosen simplex of reference $\mathscr{S}$ in PG(m, 2). There is good reason to believe that $C_{m-2}(m)$ is generated by $\mathscr{F}_{m-2}^m \cup \mathscr{F}_{m-1}^m$ and, in general, that $C_r(m)$ is generated by $\mathscr{F}_r^m \cup \cdots \cup \mathscr{F}_{m-1}^m$. At any rate, using the coset decomposition proved in Sec. III, we prove this for $m = 2, 3, 4$. It also follows from this decomposition that the orders of the groups $C_r(m)$, at least in these cases, are given by

$$|C_r(m)| = 2^{q_r(m)}, \quad \text{where } q_r(m) = \binom{m+1}{r+1} + \cdots + \binom{m+1}{m}.$$

This is in agreement with a count of the relevant number of faces of the simplex of reference.

Again from I (corollary to Theorem 2.3), we have a chain of subgroups

$$C_0 \supset C_1 \supset \cdots \supset C_m = \{1\}.$$

Taking inverse images under the projection π: $G_0 \to C_0$ gives another subgroup chain

$$G_0 \supset G_1 \supset \cdots \supset G_m = \{\pm I\},$$

with

$$G_r = \pi^{-1}(C_r) = \langle \pm\Gamma(\alpha): \alpha \in S_r(m)\rangle.$$

Here, for example, if $\alpha = \{p,q,r\}$, then $\pm\Gamma(\alpha) = \pm\Gamma_p\Gamma_q\Gamma_r$. Clearly

$$|G_r(m)| = 2|C_r(m)| = 2^{1+q_r(m)}.$$

Since this paper deals mainly with the case $m = 4$ let us recall that four-dimensional projective space PG(4, 2) over the field $\mathbb{F}_2$ consists of 31 points, 155 lines, 155 planes, and 31 solids. Each plane has the structure of PG(2, 2) and each solid the structure of PG(3, 2). From Hirschfeld[2] (or from I) we know that through each point there are 15 lines, 35 planes, and 15 solids. Through each lines there are seven planes and seven solids. Through each plane there are three solids. In general, two planes intersect in a point but they can also intersect in a line. More precisely, let π be any plane in PG(4, 2), of the remaining 154 planes, 112 of these meet π in a point and 42 meet π in a line. Similarly, if λ is a line, then there are 112 lines skew to λ and 42 that meet λ. Also let us record the well-known fact[3] that the group of projectivities of PG(4, 2), which can be identified with GL(5,$\mathbb{F}_2$), has order $31 \cdot 30 \cdot 28 \cdot 24 \cdot 16$.

 0022-2488/90/061315-10$03.00

At one stage we will need the following result of Bose and Burton. Let $\theta(m) = 2^{m+1} - 1$ denote the number of points of PG$(m, 2)$. [In I it was denoted $N(0, m)$.] Let α be a set of points of PG$(m, 2)$ such that, for any r-space π_r, $|\alpha \cap \pi_r| \geqslant \theta(s)$. Then $|\alpha| \geqslant \theta(m - r + s)$; moreover, $|\alpha| = \theta(m - r + s)$ if and only if α is an $(m - r + s)$-space.

A proof can be found in Hirschfeld.[4]

II. THE GROUP G_0 AND ITS SUBGROUPS

We now consider the subgroup chain for $m = 4$:

$$G_0 \supset G_1 \supset G_2 \supset G_3 \supset G_4 = \{\pm I\}. \tag{2.1}$$

Let Z_r denote the centralizer of G_r in G_0. The main result is that $Z_3 = G_1$, $Z_2 = G_2$, and $Z_1 = G_3$. Of course, from I [(4.10)] we already know that $G_{4-r} \subseteq Z_r$, for $r = 1, 2, 3$.

Now let α be a figure: we say that *α has an odd relation to S_r* whenever α bears an odd relation to some element $\beta \in S_r$ (as defined in Sec. IV of I), that is, whenever $\epsilon(\alpha,\beta) = -1$, for some $\beta \in S_r$. This is equivalent to saying that there is a $\beta \in S_r$ such that an odd number of points of α do not belong to β. The significance of this is that if α has an odd relation to S_r then $\Gamma(\alpha) \notin Z_r$ because there is an r-space β such that $\Gamma(\alpha)$ and $\Gamma(\beta)$ do not commute.

Now let p, q be distinct points in $S_0(4)$ and let r be the third pont of the line defined by them. Then

$$\Gamma_p \Gamma_q = \Gamma_r \pmod{G_1}$$

and hence we can write

$$G_0 = \cup \{\Gamma_p G_1 \colon p \in S_0(4) \cup \{0\}\} \tag{2.2}$$

with the understanding that $\Gamma_0 G_1 = G_1$. (Note: We are not claiming at this stage that the sets involved in this decomposition are disjoint.)

Theorem 2.1: $Z_3 = G_1$.

Proof: Let x be an element of G_0 and suppose that $x \notin G_1$. Then $x \in \Gamma_p G_1$ for some $p \in S_0(4)$. But for any point p there is clearly a three-space σ such that $p \notin \sigma$, i.e., p has an odd relation to S_3 so that $\Gamma_p \notin Z_3$. Thus $x \notin Z_3$. This proves that $Z_3 \subset G_1$ and hence the result.

Theorem 2.2: $Z_2 = G_2$.

Before giving the proof we prove a result that must surely have been noticed in other contexts.

Lemma 2.3: If a set of points in PG$(4, 2)$ has at least seven members then at least four of the points are coplanar.

Proof: The result is easily seen to be true if all the points lie in a three-dimensional subspace so we may suppose that five of the points $v_1, \ldots, v_5$ (say) form a basis for $V(5)$. Two more points, say x and y, are to be chosen from the remaining 26, namely, $u = v_1 + \cdots + v_5$, $v_{ij} = v_i + v_j$ (ten such points), $v_{ijk} = v_i + v_j + v_k$ (also ten), $u_i = u + v_i$ (five).

If x is a v_{ij} or a v_{ijk}, then x and three of the v's are coplanar.

If $x = u_i$ and $y = u_j$, then x, y, v_i, v_j are coplanar.

If $x = u_i$ and $y = u$, then x, y, v_i are collinear so there are four coplanar points in this case as well.

Proof of Theorem 2.2: We show first that any subset of PG$(4, 2)$ with six or less points has an odd relation to S_2. Let $\alpha \subset$ PG$(4, 2)$ and suppose $|\alpha| = 2$, 4, or 6. Take any $p \in \alpha$. Then there is a line λ through p that does not contain any other point of α. (There are 15 lines through p and $|\alpha \setminus \{p\}| \leqslant 5$.) Now there is a plane π through λ that does not contain any other point of α. (There are seven planes through λ.) On the other hand, if $|\alpha| = 1, 3, 5$, we can repeat this argument starting with $p \notin \alpha$. In each case there is a plane π with the property that an odd number of points of α are not contained in π. In other words, $\Gamma(\alpha) \notin Z_2$.

For the next case, suppose that $|\alpha| = 7$. If $\Gamma(\alpha) \in Z_2$, then $\Gamma(\alpha)$ commutes with $\Gamma(\pi)$ for every plane π. By I [(4.4)] every plane meets α and hence by the result at the end of Sec. I with $m = 4$, $r = 2$, $s = 0$, α is itself a plane. Thus $\Gamma(\alpha) \in G_2$.

The proof can now be completed by an inductive argument. Let α be a figure in PG$(4, 2)$. We have to show that

$$\text{if } \Gamma(\alpha) \in Z_2, \quad \text{then } \Gamma(\alpha) \in G_2. \tag{2.3}$$

Suppose $|\alpha| = n$; then (2.3) has already been proved for $n \leqslant 7$. Take as an inductive hypothesis (2.3) for all figures with less than n points. Since we may suppose $n > 7$, by Lemma 2.3, there are four points of α that are coplanar. Let π' by the plane defined by them. Now $\Gamma(\pi') \in Z_2$; thus $\Gamma(\alpha\pi') \in Z_2$ and $|\alpha\pi'| < |\alpha|$. By the inductive hypothesis $\Gamma(\alpha\pi') \in G_2$; thus $\Gamma(\alpha) \in G_2$. This completes the proof.

Theorem 2.4: $Z_1 = G_3$.

Proof: We have to show that if $\Gamma(\alpha) \in Z_1$, then $\Gamma(\alpha) \in G_3$. Notice that if α is a figure (a small set), then $|\alpha| \leqslant 15$. We claim that if $|\alpha| < 15$, then $\Gamma(\alpha) \notin Z_1$: whilst if $|\alpha| = 15$ and $\Gamma(\alpha) \in Z_1$, then α is a three-space.

The first part, $|\alpha| < 15$, is clear because if $|\alpha|$ is odd, take $p \notin \alpha$. There are 15 lines through p so there is a line λ that does not meet α. If $|\alpha|$ is even, take $p \in \alpha$ and again there is a line λ that does not meet $\alpha \setminus \{p\}$. In each case, we have found a line λ such that an odd number of points of α do not belong to λ. So α has an odd relation to S_1.

On the other hand, suppose $|\alpha| = 15$ and $\Gamma(\alpha) \in Z_1$. Then α meets every line. Using the result at the end of Sec. I, $|\alpha \cap \pi_1| \geqslant 1 = \theta(0)$. But $|\alpha| = 15 = \theta(4 - 1 + 0)$; so α is a three-space. Thus $\Gamma(\alpha) \in G_3$. The proof is now complete.

Remark: We do not need to generalize these results at this stage, but it is easy to see that Theorems 2.1 and 2.4 generalize to give $Z_{m-1} = G_1$ and $Z_1 = G_{m-1}$ in PG$(m, 2)$. This is part of conjecture B of I.

Remark: An immediate consequence of Theorem 2.2 is that G_2 is a maximum Abelian normal subgroup of G_0. (This was also conjectured in I.) We can express G_2 in the form

$$G_2 = K_2 \times \{I, -I\}, \quad \text{with } K_2 \simeq C_2 \simeq (\mathbb{Z}_2)^{15}. \tag{2.4}$$

A choice of a set of 15 generators for K_2 will provide us with a complete commuting set of 15 operators, each operator having eigenvalues ± 1. The 2^{15} sets of simultaneous eigenvalues will then serve to label the 2^{15} spinor states of our irreducible representation of Cliff$(0, 31)$. Looking ahead to Theorem 3.2, a suitable choice of a complete commuting set of 15 operators is the set

$$\{\Gamma(\alpha);\ \alpha \in \mathscr{F}_2 \cup \mathscr{F}_3\} \tag{2.5}$$

associated with the 10 two-faces and 5 three-faces of a chosen simplex of reference for PG$(4, 2)$.

III. PG (4, 2) AND THE SUBGROUP CHAIN $C_0 \supset C_1 \supset \cdots \supset C_4 = \{1\}$

We wish to find various quotient groups of the chain and to use the coset decompositions to gain insight into the groups C_i. To this end we introduce two "intermediate" groups. Choose a distinguished solid δ which will remain fixed throughout the discussion. Let $C_{2,\delta}$ be the subgroup of C_0 generated by the planes of PG(4, 2) together with the lines that lie in δ. Let $C_{3,\delta}$ be the subgroup generated by the solids of PG(4, 2) together with the planes that lie in δ. Thus we are led to the subgroup chain

$$C_0 \supset C_1 \supset C_{2,\delta} \supset C_2 \supset C_{3,\delta} \supset C_3 \supset C_4 = \{1\}. \tag{3.1}$$

Recall that $S_0 = \{s_1, s_2, \ldots, s_{31}\}$. To avoid an excess of suffixes let us also write p, q etc. for the points of S_0.

Let v be a point that is not in δ. Let δ' be a plane in δ and let v' be a point in δ but not in δ'. We write ν_p and ν'_p for the joins $j(v, p)$ and $j(v', p)$, respectively. When λ is a line, we write α_λ and α'_λ for the joins $j(v, \lambda)$ and $j(v', \lambda)$, respectively. Thus ν_p and ν'_p are lines while α_λ and α'_λ are planes.

Before starting the main theorem we describe the decompositions involved.

Let λ be a line in PG(4, 2). Suppose that λ does not lie entirely in δ. Then λ will meet δ at a point p and α_λ will meet δ in a line λ'. The lines λ, ν_p, λ' are coplanar and concurrent. By Theorem 2.3 of I, we have $\lambda\, \nu_p\, \lambda' = \pi$ for some plane π. Thus we can write $\lambda = \nu_p\, \lambda' \pi$. If $\lambda \subset \delta$, we write $\lambda = \nu_0 \lambda \pi_0$, where ν_0 and π_0 are interpreted as 1. So

$$\lambda \in \nu_p C_{2,\delta}, \quad \text{where } p \in S_0(\delta) \cup \{0\}. \tag{3.2}$$

Suppose that λ lies in δ but not in δ' and that $v' \notin \lambda$. Then λ will meet δ' at a point p and the plane α'_λ will meet δ' in a line λ'. As above, there is a plane π' such that $\lambda \nu'_p\, \lambda' = \pi'$, so that $\lambda = \nu'_p\, \lambda' \pi'$. If λ lies in δ', we write $\lambda = \nu'_0\, \lambda'\, \pi'_0$; while if v' lies on λ, we write $\lambda = \nu'_p\, \lambda_0\, \pi'_0$, where again ν'_0, λ_0, and π'_0 are interpreted as 1. So

$$\lambda \in \nu'_p\, \lambda' C_2,$$
$$\text{where } p \in S_0(\delta') \cup \{0\}, \quad \lambda' \in S_1(\delta') \cup \{0\}. \tag{3.3}$$

Let π be a plane in PG (4, 2) that does not lie in δ or contain v. Then the join $j(v, \pi)$ is a solid that meets δ in a plane π'. Thus π meets δ in a line λ'. The three planes π, π', and α'_λ are contained in $j(v, \pi)$ and each one contains the line λ'. Thus $\pi\pi'\alpha_{\lambda'} = j(v, \pi)$ (Theorem 2.3 of I again).

The plane π also meets δ' in a point p. Suppose that λ' does not contain v'; then $\alpha'_{\lambda'}$ meets δ' in a line λ. The lines ν'_p, λ', and λ are concurrent and coplanar; hence the planes $\alpha_{\nu'_p}$, $\alpha_{\lambda'}$, and α_λ have a comon line and lie in a solid σ. Thus $\alpha_{\nu'_p}\, \alpha_{\lambda'}\, \alpha_\lambda = \sigma$. Combining these two expressions gives

$$\pi = \alpha_{\lambda'} \pi' j(v, \pi) = \alpha_{\nu'_p}\, \alpha_\lambda \pi' \sigma j(v, \pi).$$

If $v' \in \lambda'$, then we can take $\lambda = 0$. If $\lambda' \subset \delta'$, we can take $p = 0$. If $\pi \subset \delta$, we can take $p = 0$ and $\lambda = 0$. If $v \in \pi$, we go straight to the second step of the decomposition. Thus we can write

$$\pi \in \alpha_{\nu'_p}\, \alpha_\lambda C_{3,\delta},$$
$$\text{where } p \in S_0(\delta') \cup \{0\}, \quad \lambda \in S_1(\delta') \cup \{0\}. \tag{3.4}$$

We are now in a position to prove the main theorem of this section.

Theorem 3.1: (i) The disjoint cosets of C_1 in C_0 are given by

$\{p\, C_1:\ p \in S_0(4) \cup \{0\}\}$;

(ii) $C_0/C_1 \simeq V(5)$;

(iii) the disjoint cosets of $C_{2,\delta}$ in C_1 are

$\{\nu_p C_{2,\delta}:\ p \in S_0(\delta) \cup \{0\}\}$;

(iv) $C_1/C_{2,\delta} \simeq V(4)$;

(v) the disjoint cosets of C_2 in $C_{2,\delta}$ are

$\{\nu'_p\, \lambda' C_2:\ p \in S_0(\delta') \cup \{0\},\ \lambda' \in S_1(\delta') \cup \{0\}\}$.

(vi) $C_{2,\delta}/C_2 \simeq V(3) \times \widehat{V}(3)$;

(vii) the disjoint cosets of $C_{3,\delta}$ in C_2 are

$\{\alpha_{\nu'_p}\, \alpha_\lambda C_{3,\delta}:\ p \in S_0(\delta') \cup \{0\},\ \lambda \in S_1(\delta') \cup \{0\}\}$;

(viii) $C_2/C_{3,\delta} \simeq V(3) \times \widehat{V}(3)$;

(ix) the disjoint cosets of C_3 in $C_{3,\delta}$ are

$\{\pi\, C_3:\ \pi \in S_2(\delta) \cup \{0\}\}$;

(x) $C_{3,\delta}/C_3 \simeq \widehat{V}(4)$;

(xi) $C_3 \simeq \widehat{V}(5)$.

Proof: Take distinct points $p, q \in S_0(4)$ and let r be the third point of the line defined by them. Then

$$pq = r \pmod{C_1}. \tag{3.5}$$

It follows that C_0 is the union of the cosets listed in (i). Thus $|C_0| \leqslant |C_1| 2^5$ with equality if and only if the cosets are disjoint.

Part (ii) follows from (3.5) once it has been shown that the cosets are disjoint.

If p and q are distinct points of $S_0(\delta)$ and r is the third point of the line defined by them, then

$$\nu_p \nu_q = \nu_r \pmod{C_{2,\delta}}. \tag{3.6}$$

It follows from (3.2) that C_1 is the union of the cosets listed in (iii). Thus $|C_1| \leqslant |C_{2,\delta}| 2^4$ with equality if and only if the cosets are disjoint.

Part (iv) follows from (3.6) once it has been shown that the cosets are disjoint.

Parts (v)–(x) follow in the same way from (3.4) giving $|C_{2,\delta}| \leqslant |C_2| 2^6$, $|C_2| \leqslant |C_{3,\delta}| 2^6$, and $|C_{3,\delta}| \leqslant |C_3| 2^4$ with equality in each case if and only if the corresponding cosets are disjoint.

Part (xi) is dealt with in the Introduction. Combining the inequalities for the order for the various groups we have that $|C_0| \leqslant 2^5 \cdot 2^4 \cdot 2^6 \cdot 2^6 \cdot 2^4 \cdot 2^5$ with equality if and only if the cosets are disjoint in every decomposition. But we know that $|C_0| = 2^{30}$; so the proof of the theorem is complete.

The groups involved in these decompositions can all be regarded as $\mathbb{F}_2$ vector spaces so that any extensions involved are split extensions.

Thus, using the second isomorphism theorem for groups, we have that $C_{3,\delta}/C_3$ is a subgroup of C_2/C_3 with quotient $C_2/C_{3,\delta}$ and hence

$$C_2/C_3 \simeq C_2/C_{3,\delta} \times C_{3,\delta}/C_3 \simeq V(3) \times \widehat{V}(3) \times \widehat{V}(4).$$

Similarly

$$C_1/C_2 \simeq C_1/C_{2,\delta} \times C_{2,\delta}/C_2 \simeq V(4) \times V(3) \times \hat{V}(3).$$

From these isomorphisms arise two further coset decompositions: the disjoint cosets of C_2 in C_1 are

$$\{\nu_p\, \nu'_q\, \lambda' C_2:\ p \in S_0(\delta) \cup \{0\},\ q \in S_0(\delta') \cup \{0\},\ \lambda' \in S_1(\delta') \cup \{0\}\}; \quad (3.7)$$

the disjoint cosets of C_3 in C_2 are

$$\{\alpha_{\nu'_p}\, \alpha_\lambda\, \pi\, C_3:\ p \in S_0(\delta') \cup \{0\},\ \lambda \in S_1(\delta') \cup \{0\},\ \pi \in S_2(\delta) \cup \{0\}\}. \quad (3.8)$$

The isomorphisms of the theorem can be summarized by the following diagram:

$$\begin{array}{ccccccccc} C_0 & \supset & C_1 & \supset & C_2 & \supset & C_3 & \supset & C_4 \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \\ V(5) & & V(4)\times V(3)\times \hat{V}(3) & & V(3)\times \hat{V}(3)\times \hat{V}(4) & & \hat{V}(5) & & \end{array}$$

The isomorphisms of the theorem could be established in other ways. They could be cast in a more algebraic framework by assigning to each line, plane, etc., its Plücker (Grassmann) coordinates. For example, a line in PG(4, 2) has ten coordinates and this would give rise to a vector in $V(4) \times V(3) \times \hat{V}(3)$. Different partitions of the coordinates as $4 + 3 + 3$ would correspond to different choices of δ and v.

Alternatively a third possible line of proof would involve the following ideas. Assigning to each line in PG(4, 2) the point where it meets δ (0 if the line is in δ), gives a homomorphism not from C_1 but from the *free* group generated by lines to $V(4)$. To get a homomorphism from C_1 it would be necessary to show that relations between lines are taken to zero. What are these relations? We conjecture that in PG(3, 2) the only relations between lines would arise from sets of five skew lines and reguli (see Hirschfeld[5]). We do not know what they would be in four dimensions.

Similarly we would have to assign to each plane the point where it meets δ' and the projection of the line where it meets δ. This time we would have to ask for relations between planes.

The group C_0 and its subgroups are generated by points, lines, etc. As explained in the Introduction the coset decomposition can be utilized to give a more succinct system of generators.

Theorem 3.2: For $m = 2, 3, 4$, the group $C_r(m)$ is generated by the union $\mathscr{F}^m_r \cup \mathscr{F}^m_{r+1} \cup \cdots \cup \mathscr{F}^m_{m-1}$, where $\mathscr{F}^m_r$ denotes the set of r-dimensional faces of the chosen simplex of reference $\mathscr{S}$ in PG$(m, 2)$.

Proof: In the Introduction it was shown that $C_{m-1}(m)$ is generated by $\mathscr{F}^m_{m-1}$.

On the other hand, we know, at least when $m = 2, 3, 4$, that $C_0(m)$ is the disjoint union of the cosets $\{p\, C_1(m):\ p \in S_0(m) \cup \{0\}\}$. So it follows that $C_0(m)/C_1(m)$ is generated by $\mathscr{F}^m_0$. These two steps deal with the $m = 2$ case.

Consider PG(3, 2). Choose the distinguished plane δ to be a face of $\mathscr{S}$ and the distinguished point v to be the remaining vertex. From I (Lemma 3.1) we have the coset decomposition of C_1 as the union of $\{\nu_p \lambda C_2:\ p \in S_0(\delta) \cup \{0\},\ \lambda \in S_1(\delta) \cup \{0\}\}$. So it is only necessary to show that each ν_p and λ can be written in terms of the required generators. This is already clear for λ by considering δ as PG(2, 2), and it also follows for ν_p because we know that $p \in S_0(\delta) \cup \{0\}$ can be written in terms of $\mathscr{F}^2_0 \cup \mathscr{F}^2_1$ [again in PG(2, 2)], and ν_p, being the join $j(v, p)$, can be written in terms of $\mathscr{F}^3_1 \cup \mathscr{F}^3_2$. This deals with PG(3, 2).

The $m = 4$ case can be dealt with in the same way. Choose δ to be three-dimensional face of $\mathscr{S}$ and v the remaining vertex. Then choose δ' to be a two-dimensional face of $\mathscr{S}$ in δ and v' the remaining vertex in δ. Now consider the decompositions (3.7) and (3.8). Each term is either the join with a vertex or already lies in a lower-dimensional face.

This completes the proof.

IV. THE GROUP C_2 IN CLOSE-UP

In the case $m = 3$ we know that the group $C_1(3)$, generated by the lines of PG(3, 2), is partitioned into seven orbits under the action of the subgroup $\Omega(3) \simeq GL(4;\mathbb{F}_2)$ of Aut $C_0(3)$. For details, see Lemma 3.3 of I. The seven orbits, say $\theta_0, \theta_1, \ldots, \theta_6$, are as follows:

$$\theta_0 = \{\varnothing\},\quad \theta_1 = \{35 \text{ lines}\},\quad \theta_2 = \{105 \text{ two-frames}\},$$
$$\theta_3 = \{280 \text{ skew pairs}\},\quad \theta_4 = \{15 \text{ planes}\}, \quad (4.1)$$
$$\theta_5 = \{168 \text{ three-frames}\},\quad \theta_6 = \{420 \text{ tripods}\}.$$

The figures belonging to the orbits $\theta_0, \theta_1, \ldots, \theta_6$ have respective sizes 0, 3, 4, 6, 7, 5, 7. Altogether we have, of course, $1024 = |C_1(3)|$ figures.

In this section we will determine the orbits of $C_2 = C_2(4)$ under the action of the subgroup $\Omega \simeq GL(5;\mathbb{F}_2)$ of Aut C_0. To this end note that a figure ψ formed from the product of lines $\lambda, \lambda', \ldots$ lying in a solid δ gives rise, upon choosing a vertex $v \notin \delta$, to a figure $\Psi \in C_2$ formed from the product of planes $\alpha, \alpha', \ldots$, where $\alpha = j(v, \lambda)$, $\alpha' = j(v, \lambda'), \ldots$. Each of the orbits $\theta_1, \ldots, \theta_6$ gives rise in this way to corresponding orbits $\Theta_1, \ldots, \Theta_6$ for the group $C_2 = C_2(4)$. So already we have seven orbits for C_2, namely:

$$\Theta_0 = \{\varnothing\},\quad \Theta_1 = \{155 \text{ planes}\},\quad \Theta_2 = \{465 \text{ cubes}\},$$
$$\Theta_3 = \{8680 \text{ slew pairs}\},\quad \Theta_4 = \{31 \text{ solids}\}, \quad (4.2)$$
$$\Theta_5 = \{5208 \text{ pentapods}\},\quad \Theta_6 = \{4340 \text{ triblades}\}.$$

In the next few paragraphs we will explain the terminology employed in the list (4.2), and will also justify the assertions there concerning $|\Theta_i|$ (in the less obvious cases, namely, for $i = 2, 3, 5, 6$). Granted the values of $|\Theta_i|$, $i = 0, 1, \ldots, 6$, note that we have so far accounted for 18 880 figures of C_2. Since $|C_2| = 2^{15} = 32\,768$, it follows that we have still to consider 13 888 figures of C_2. We shall see in a moment that these remaining figures form a single Ω-orbit Θ_7 in C_2.

Concerning Θ_2, recall that a two-frame ψ arises as the product $\lambda\mu$ of two lines lying in a plane α. Equivalently, expressing α as a star $\lambda\mu\nu$ of lines, we can view ψ as the complement $\alpha\nu = \alpha\backslash\nu$ within α of the line ν. This second point of view has the advantage that ψ has the unique expression $\psi = \alpha\nu$, $\nu \subset \alpha$, in contrast to three different expressions $\psi = \lambda\mu = \lambda'\mu' = \lambda''\mu''$ arising from the three choices of the point $p \in \nu$ to form $\text{star}_1(p, \alpha)$. Viewing ν as line at infinity in the plane α the four points of $\psi = \alpha\nu$ can be considered as the four points, which could perhaps be referred to as a "square," of a two-dimensional affine geometry AG(2, 2). Going up one dimension, consider the product $\psi = \alpha\beta$ of two planes lying inside a solid σ. Expressing σ as a star $\alpha\beta\gamma$ of planes we can view ψ as the complement $\sigma\gamma = \sigma\backslash\gamma$ within σ of the plane γ. Viewing γ as a plane at infinity in the solid σ, the eight points of $\psi = \sigma\gamma$ can be considered as the eight points, referred to in (4.2) as a "cube," of a three-dimensional affine geometry AG(3, 2). Now in the geometry PG (4, 2) there are 31 choices for the solid σ, and within the chosen σ there are 15 choices for the plane γ. Consequently

$$|\Theta_2| = 31\times 15 = 465. \tag{4.3}$$

Caution: We can compute $|\Theta_2|$ also by calculating first the number P ($=\frac{1}{2}\times 155\times 42$) of pairs α, β of planes such that α intersects β in some line λ. But take note that $|\Theta_2| = P/7$, because each cube $\psi = \sigma\gamma$ can be expressed in seven different ways as the product $\alpha\beta$ of two planes, corresponding to the seven choices of line $\lambda \subset \gamma$ to form $\text{star}_2(\lambda, \sigma)$. In AG(3, 2) terms we have seven different $4 + 4$ decompositions of the eight points of the "cube" ψ into a pair of parallel affine planes (or "squares")—parallel because α, β meet in a line λ of the plane at infinity γ. It is perhaps worth remarking that these seven different decompositions of an affine cube (over $\mathbb{F}_2$) into pairs of parallel planes entered into some earlier work[6] on eight-dimensional ternary composition algebras.

Concerning Θ_3, by a "slew pair" we mean a figure $\psi = \alpha\beta$ formed from a pair α, β of planes that are "as skew as possible" (granted that they lie in a projective space of dimension 4); that is, α, β intersect only at a single point p. (We imagine β as obtained from α by "pivoting," or "slewing," around this point p; hence the terminology.) Note that $|\psi| = 12$. Also, in contrast to the case of $\psi \in \Theta_2$, a slew pair ψ has unique expression in the form $\psi = \alpha\beta$, and defines also a unique "pivot" $p = \alpha \cap \beta$. Now there are 31 choices of pivot p, and any solid σ not passing through $p = \alpha \cap \beta$ will intersect $\psi = \alpha\beta$ in a pair of skew lines. As in (4.1) there are 280 choices of skew pairs in σ. Consequently

$$|\Theta_3| = 31\times 280 = 8680. \tag{4.4}$$

Alternatively, we could say that as a given plane is met in a point by 112 planes then $|\Theta_3| = \frac{1}{2}\times 155\times 112$.

Concerning Θ_5, by a "pentapod" we mean a product of five lines that join the five points of a three-frame to a common "apex" v lying outside the three-space σ of the three-frame. Clearly each $\psi \in \Theta_5$ has size $|\psi| = 11$. Just as a three-frame can be expressed in the form $\lambda\mu\nu$, where $\lambda\mu$ is a skew pair and where the line ν is a transversal of λ and μ, so a pentapod ψ can be expressed in the form $\psi = \alpha\beta\gamma$, where $\alpha\beta$ is a slew pair and where γ is a plane which is a "transversal" of α and β in the sense that γ intersects both α and β in lines (which necessarily pass through the pivot $\alpha \cap \beta$). Now there are 31 choices for the apex v of a pentapod, and any solid σ not passing through v will intersect the pentapod in a three-frame. As in (4.1) there are 168 choices of three-frames in σ. Consequently

$$|\Theta_5| = 31\times 168 = 5208. \tag{4.5}$$

Concerning Θ_6, by a "triblade" we mean a figure $\psi = \alpha\beta\gamma$ formed from the product of three planes α, β, γ that intersect in a common line λ (the "axis" of ψ) but that do not lie within any common solid. Clearly $|\psi| = 3 + 4 + 4 + 4 = 15$. If σ is any solid not containing the axis λ of the triblade $\psi = \alpha\beta\gamma$, then $\psi \cap \sigma$ is a tripod with apex $p = \lambda \cap \sigma$ and with legs $\alpha \cap \sigma$, $\beta \cap \sigma$, and $\gamma \cap \sigma$. Now there are 155 choices for the axis λ, and within σ there are $7\cdot 6\cdot 4/3! = 28$ choices for the three legs of a tripod having p as apex. Consequently

$$|\Theta_6| = 155\times 28 = 4340. \tag{4.6}$$

Incidentally, just as a tripod can be thought of (see I) as (the complement within three-space of) a "pierced plane" $\alpha\lambda$, where the line λ does not lie in the plane α, so can a triblade be thought of as (the complement of) a "sliced solid" $\sigma\alpha$, where the plane α does not lie in the solid σ. Now there are 31 choices of solid σ, and $155 - 15 = 140$ choices of plane α not lying in σ. Consequently we have the following check of (4.6):

$$|\Theta_6| = 31\times 140 = 4340.$$

As noted after (4.2) there are 13 888 figures of $C_2(4)$ that do not arise, in the above manner, out of figures of $C_1(3)$. For a start, consider a figure of the kind

$$\psi = \alpha\beta\gamma, \tag{4.7}$$

where α, β, γ are pairwise slew planes. The plane α contributes five points to ψ, the remaining two points of α being the pivots $b' = \alpha \cap \gamma$ and $c' = \alpha \cap \beta$. Similarly β does not contribute the pivots c' and $a' = \beta \cap \gamma$, and γ does not contribute the pivots a' and b', and so $|\psi| = 5 + 5 + 5 = 15$. Such a figure ψ, formed from three mutually slew planes, will be referred to as a 15_3. Now the points $a = b' + c'$, $b = c' + a'$, $c = a' + b'$ constitute a line λ that is a transversal for the three planes α, β, γ. Let μ_a, ν_a be the lines of α through a other than $j(b', c')$, and let μ_b, ν_b and μ_c, ν_c be similarly defined. Then the 15_3 figure $\psi = \alpha\beta\gamma$ can be expressed in the form

$$\psi = \lambda\mu_a\nu_a\mu_b\nu_b\mu_c\nu_c \tag{4.8}$$

and could perhaps also be referred to as a "sixwig" consisting of a "body" λ together with the three pairs of "legs," each point p of the body contributing a pair of legs μ_p, ν_p.

The number N of mutually slew triads of planes α, β, γ can be computed as follows. There are $31\times 30\times 28$ choices for the three pivot points a', b', c'. There are seven planes through $j(b', c')$; discounting the plane of the pivots, this leaves six choices for the plane α. Of the corresponding choices for the plane β a further two are discounted because they intersect α in a line, so that there are four choices for β. Similarly there are $6-2-2=2$ choices for γ. Consequently

$$N = 31\times30\times28\times6\times4\times2/3! = 15\times13\,888. \quad (4.9)$$

Clearly N is *not* the number of 15_3 figures in C_2 since it is much larger than the order of C_2! In fact, as described in the Appendix, a 15_3 figure ψ is a surprisingly symmetric object that can be expressed as the product $\alpha\beta\gamma$ of mutually slew planes in precisely 15 ways. Consequently, if Θ_7 denotes the set of all the 15_3 figures, we have

$$|\Theta_7| = N/15 = 13\,888. \quad (4.10)$$

One easily sees that Θ_7 forms a single Ω-orbit, and so, since

$$|\Theta_0| + |\Theta_1| + \cdots + |\Theta_7| = 32\,768 = 2^{15} = |C_2|,$$

we have arrived at the complete list

$$\Theta_0, \Theta_1, \Theta_2, \Theta_3, \Theta_4, \Theta_5, \Theta_6, \Theta_7 \quad (4.11)$$

of the Ω-orbits constituting the subgroup C_2 of C_0. The figures of these eight orbits are of respective sizes

$$0, 7, 8, 12, 15, 11, 15, 15. \quad (4.12)$$

Remark: We can make the observation that, as is to be expected, the number of elements in each orbit divides the order of the group $GL(5, \mathbb{F}_2)$.

Remark: The 15 different $5+5+5$ decompositions of the 15 points of a 15_3 figure ψ alluded to above, each decomposition being associated with representation of ψ of the kind (4.7), can be given a more colorful and detailed description in terms of sixwig decompositions of the kind (4.8). One finds that a sixwig has a remarkable propensity for reconstituting itself, in that, for a start, each of its six legs can be thought of as the body of another sixwig formed out of the same fifteen points. Moreover, for the sixwig ψ in (4.8), through each of the four points of the legs μ_a, ν_a (leaving aside their point of attachment a to the body λ) there pass two transversals of the original $5+5+5$ decomposition. These further eight transversals of the original structure can equally be thought of as the body of a sixwig formed from the same 15 points of ψ. Thus ψ can be thought of as a sixwig in $1+6+8=15$ different ways, the 15 different choices of body comprising in fact the totality of lines that lie entirely inside the figure ψ. As described more fully in the Appendix, all 15 points of ψ enter on an equal footing, as do all 15 of the lines, each point being the apex of a tripod of lines and each line containing three points of ψ (indeed, since we are working over the field $\mathbb{F}_2$, each line consists solely of three points of ψ).

Since our arithmetic has assured us that the list of (4.10) is complete, any other apparently different figure of C_2 must, in fact, be one of the above in disguise. We have already encountered the example of a sliced solid being the (complement of) a triblade. A more interesting example is that of a figure $\psi = \sigma\alpha\alpha'$ formed from a solid σ and a slew pair α, α' of planes whose pivot v does not lie in σ. Clearly $|\psi| = 15$, since ψ consists of nine points of the solid σ together with two triangles, one in the plane α and the other in the plane α'. Now one can choose the pivot v in 31 ways and a solid σ not passing through v in $31-15=16$ ways. Within σ there are 280 skew pairs λ, μ of lines, whose joins to v will yield 280 slew pairs of planes α, α' having v as pivot. Consequently the number N' of such $(\sigma, \alpha, \alpha')$-configurations is

$$N' = 31\times16\times280 = 138\,880, \quad (4.13)$$

again much larger than $|C_2|$! Now, in fact (again see the Appendix), the figure ψ in question can be seen to be a 15_3 figure, and the symmetry of a 15_3 figure ψ is such that it can be expressed in the form

$$\psi = \sigma\,\alpha\,\alpha' \quad (4.14)$$

(where α, α' are a slew pair of planes whose pivot v lies outside the solid σ) in precisely ten distinct ways. Consequently the number of such figures is

$$N'/10 = 13\,888, \quad (4.15)$$

in agreement with (4.10).

V. SOME PROJECTIVE GEOMETRY SPIN-OFFS

The result $Z_2 = G_1$ of Theorem 5.2 of I yields the following theorem in solid projective geometry over $\mathbb{F}_2$.

Theorem 5.1: A subset φ of PG(3,2) bears an even relation to every plane if and only if φ (or equally well φ^c) is a product of lines.

In more detail, after making use of (4.1), we have the following results.

Corollary 5.2: (a) Let $H_k = H_k(\varphi)$ denote the proposition H_k: φ is a k-set that has odd intersection with every plane of PG(3, 2). Then the following implications hold:

$H_3 \Leftrightarrow \varphi$ is a line,

$H_5 \Leftrightarrow \varphi$ is a three-frame,

$H_7 \Leftrightarrow \varphi$ is either a plane or a tripod,

$H_9 \Leftrightarrow \varphi^c$ is a skew pair of lines

(equivalently, φ is a regulus),

$H_{11} \Leftrightarrow \varphi^c$ is a two-frame.

(Also no φ satisfies H_1 or H_{13}.)

(b) Let J_k be obtained from H_k by replacing "odd intersection" by "even intersection." Then the following implications hold:

$J_{12} \Leftrightarrow \varphi^c$ is a line,

$J_{10} \Leftrightarrow \varphi^c$ is a three-frame,

$J_8 \Leftrightarrow$ either φ^c is a plane, or φ is a pierced plane,

$J_6 \Leftrightarrow \varphi$ is a skew pair of lines,

$J_4 \Leftrightarrow \varphi$ is a two-frame.

(Also no φ satisfies J_{14} or J_2.)

(c) No φ satisfies H_k for k even, or J_k for k odd.

Remark: Part (b) is merely a rewrite of part (a). The four-dimensional case of part (c) is proved below.

Similarly the result $Z_2 = G_2$ of Theorem 2.2 of the present paper yields a corresponding theorem in four-dimensional projective geometry.

Theorem 5.3: A subset φ of PG(4, 2) bears an even relation to every plane if and only if φ (or equally well φ^c) is a product of planes.

In more detail, after making use of (4.2) and (4.12), we have the following results.

Corollary 5.4: Let H_k and J_k denote the same propositions as in Corollary 5.2, except that PG(4, 2) now takes the place of PG(3, 2). Then we have the following.

(a) No φ satisfies H_k for odd $k<31$ except when $k=7, 11, 15, 19, 23$, in which cases the following implications hold:

$H_7 \Leftrightarrow \varphi$ is a plane,

$H_{11} \Leftrightarrow \varphi$ is a pentapod,

$H_{15} \Leftrightarrow$ either φ is a solid, or φ is a triblade,

or φ is a 15_3,

$H_{19} \Leftrightarrow \varphi^c$ is a slew pair of planes,

$H_{23} \Leftrightarrow \varphi^c$ is a cube.

(b) No φ satisfies J_k for even $k>0$ except when $k=24, 20, 16, 12, 8$, in which cases the following implications hold:

$J_{24} \Leftrightarrow \varphi^c$ is a plane,

$J_{20} \Leftrightarrow \varphi^c$ is a pentapod,

$J_{16} \Leftrightarrow$ either φ^c is a solid, or φ is a sliced solid,

or φ^c is a 15_3,

$J_{12} \Leftrightarrow \varphi$ is a slew pair of planes,

$J_8 \Leftrightarrow \varphi$ is a cube.

(c) No φ satisfies H_k for k even, or J_k for k odd.

Remark: Part (c) says that an even set cannot have odd intersection with every plane and that an odd set cannot have even intersection with every plane.

Proof of (c): Let φ be any subset of PG(4, 2). Suppose that for every plane π the elements $\Gamma(\varphi)$ and $\Gamma(\pi)$ of G_0 anticommute. Let $\pi_1, \ldots, \pi_7$ be the seven planes through a line. Then, using Theorem 2.3 of I, we have that $\Gamma(\pi_1)\cdots\Gamma(\pi_7) = \pm I$, so that $\Gamma(\varphi)$ anticommutes with $\pm I$, which is absurd. Hence there is a plane π such that $\Gamma(\varphi)$ and $\Gamma(\pi)$ commute, in other words, an even number of points of φ are not in π.

In the case of a 15_3 figure ψ of PG (4, 2), we can appeal to our knowledge of ψ as given in the Appendix to produce details concerning its odd intersection with all the 155 planes of PG(4, 2). Let us call a plane α an s-plane whenever $|\alpha \cap \psi| = s$. Then one obtains the following results:

(i) there are 45 five-planes, each five-plane intersecting ψ in one of the 45 bipods contained in ψ;

(ii) for a three-plane α there are three possibilities for $|\alpha \cap \psi|$, namely,

(a) one of the 15 lines λ of L [with $\alpha = j(c, \lambda)$],

(b) one of the 20 triangles contained in ψ,

(c) one of the 60 triads contained in ψ, (5.1)

altogether there are therefore 95 three-planes, those of kind (a) or (b) accounting for all of the 35 planes through the center c of ψ;

(iii) there are 15 one-planes α_{ij}, say, one for each of the 15 points p_{ij} of ψ.

In this way we account for all the 155 ($=45+95+15$) planes in PG(4, 2).

Similarly we can use the results of the Appendix to produce the following details concerning the odd intersection of a 15_3 figure ψ with all the 31 solids of PG(4, 2):

(i) there are ten nine-solids, namely, the ten solids $\sigma_{ijk} = \sigma_{lmn}$ of (A15), which intersect ψ in one of its ten nonads;

(ii) there are 15 seven-solids, namely, the 15 solids σ_{ij} introduced after Eq. (A14), which intersect ψ in one of its 15 tripods (these 15 seven-solids accounting for all the solids through the center c of ψ);

(iii) there are six five-solids, say σ_i, each five-solid intersecting ψ in one its six pentads π_i. (5.2)

In this way we account for all the 31 solids in PG(4, 2).

APPENDIX: THE SYMMETRICAL CONFIGURATION 15_3 IN PG (4, 2)

In a five-dimensional vector space, over a field F, we may choose six vectors $v_0, v_1, v_2, v_3, v_4, v_5$ such that

$$v_0+v_1+v_2+v_3+v_4+v_5=0 \tag{A1}$$

and such that no five of the vectors are linearly dependent. Define 15 further vectors $p_{ij}=p_{ji}$ by

$$p_{ij}=v_i+v_j, \quad 0 \leqslant i \neq j \leqslant 5. \tag{A2}$$

Consider now the corresponding 15 points in the projective geometry PG(4, F), and let us denote these points by their representative nonzero vectors p_{ij}. *Suppose for the moment that char* $\mathrm{F} \neq 2$. Then the only collinearities (of three or more points) amongst the 15 points derive, by way of (A1), solely from the 15 different partitions of the six symbols 012345 into three pairs. That is, the three distinct points

$$p_{ij},\ p_{kl},\ p_{mn} \tag{A3}$$

lie on a line if and only if $ijklmn$ is a permutation of 012345. In what follows let us agree, except when anything is said to the contrary, that indices $i, j, k, \ldots$ range from 0 to 5 and that $ijklmn$ denotes an arbitrary permutation of 012345. The line containing the points (A3) will be denoted $\lambda(ij, kl, mn)$, or equally $\lambda(lk, ji, mn)$, etc.

Let ψ denote the set of 15 projective points defined via (A2) and let L denote the set of 15 projective lines defined via (A3). Now for fixed ij in (A3) there are precisely three partitions of the remaining four indices into three pairs, namely, (kl, mn), (km, nl), and (kn, lm). Consequently each point p of ψ lies on precisely three lines of L, just as each line λ of L contains precisely three points of ψ. The configuration (ψ, L) consisting of 15 points and 15 lines is a well-known configuration of four-dimensional projective geometry,[7] and is commonly denoted by the symbol 15_3. It is a highly symmetrical configuration in that all 15 points enter into it on an equal footing, as do all 15 lines; moreover, there is an obvious *duality* between points of ψ and lines of L, for example, dual to the notion of a *tripod with apex* $p \in \psi$, which by definition consists of the three lines $\lambda, \lambda', \lambda''$ of L that intersect in p (and which therefore determine six other points of ψ, a pair on each of the lines $\lambda, \lambda', \lambda''$), there is the notion of a *tripoint with body* $\lambda \in L$, which by defintion consists of the three points p, p', p'' of ψ that lie on λ (and which therefore determine six other lines of L, a pair through each of the points p, p', p'').

Actually we will often think of a tripod in terms of the seven points of intersection with ψ of its three legs, $\lambda, \lambda', \lambda''$. The seven-set $\subset \psi$ which is determined in this way by a point p_{ij} of ψ will be denoted by $\tau(p_{ij})$, and will also be referred to as the *tripod with apex* p_{ij}. Similarly we may think of a tripoint in terms of its associated *sixwig with body* λ, a seven-set $\subset L$ consisting of the body λ together with six *legs* attached in pairs at each of the points p, p', p''.

Given two distinct points $p, q \in \psi$ there are two possibilities: either p, q lie on a line of L, in which case we will say that p and q are *connected*, or they do not, in which case we will say they are *unconnected*.

In the former case we will refer to the two-set $\{p, q\} \subset \psi$ as a *bipoint*, and in the latter case as a *dyad*. Dually, given two distinct lines $\lambda, \mu \in L$, either they meet in a point of ψ, in which case we will say that λ, μ form a *bipod*, or they do not, in which case they form a *skew pair*. (Actually we will usually think of a bipod in terms of the five points of intersection with ψ of its two legs λ, μ, and this five-set will also be referred to as a bipod.) Observe that each point $p \in \psi$ is connected to six other points of ψ, namely, those forming the legs of the tripod whose apex is p, and is therefore unconnected to the remaining eight points of ψ. Dually each line $\lambda \in L$ meets six other lines of L, namely, the six legs of the sixwig whose body is λ, and is skew to the remaining eight lines of L.

By a *pentad* we will mean a totally unconnected set of five points of ψ; the five tripods having the five points as apexes then account for all the 15 lines of L. Dually we also have the notion of a *skew pentad*, or *spread*, of lines of L, whose associated tripoints account for all the 15 points of ψ. Now there are six obvious pentads (of points), say π_i, $i = 0, 1, \ldots, 5$, where

$$\pi_i = \{p_{ij};\ j \neq i\}. \tag{A4}$$

If we recall from (A3) that

$$p_{ab} \text{ is connected to } p_{cd} \Leftrightarrow \{a, b\} \cap \{c, d\} = \varnothing, \tag{A5}$$

we easily see that there are no other pentads other than the six of the kind (A4), and that there exist no totally unconnected sets of size > 5. By the terms tetrad, triad, dyad we will mean subsets of a pentad of sizes 4, 3, 2, respectively. Note therefore that there exist 30 tetrads, 60 triads, and 60 dyads.

Now a totally unconnected three-set is not necessarily a triad, since it is possible that it cannot be extended so as to form a pentad. For example, all three-sets of the form

$$\tau_{ijk} = \{p_{ij}, p_{jk}, p_{ki}\} \tag{A6}$$

are easily seen to be already maximal unconnected sets. It follows easily from (A5) that all maximal unconnected three-sets are, in fact, of the form (A6). Such three-sets of ψ will be referred to as *triangles*. There are thus $\binom{6}{3} = 20$ triangles, which occur in *complementary pairs* τ_{ijk}, τ'_{ijk}, where

$$\tau'_{ijk} = \tau_{lmn}. \tag{A7}$$

Given the triangle τ_{ijk}, the points of the complementary triangle are distinguished by the property that they are connected to all three points of τ_{ijk}. The remaining nine points of ψ, namely,

$$\nu_{ijk} = \nu_{lmn} = \{p_{ad};\ a \in \{i,j,k\},\ d \in \{l,m,n\}\}, \tag{A8}$$

can be usefully displayed in the 3×3 array

$$\nu_{ijk} = \nu_{lmn} = \begin{bmatrix} il & jm & kn \\ km & in & jl \\ jn & kl & im \end{bmatrix} \tag{A9}$$

formed by "addition" of the two arrays

$$\begin{bmatrix} i & j & k \\ k & i & j \\ j & k & i \end{bmatrix} \text{ and } \begin{bmatrix} l & m & n \\ m & n & l \\ n & l & m \end{bmatrix}. \tag{A10}$$

Here we have let jm denote the point p_{jm}, and k denote the vector v_k. Observe that the rows of (A9) form a *triplet* of mutually skew lines of L, and that the columns of (A9) form a *complementary triplet* of mutually skew lines, each column being a transversal of the three rows and each row being a transversal of the three columns. It is easily seen that each of these triplets is "maximally skew," unlike *skew triads* of lines consisting of three lines selected from a skew pentad of lines. Such nine-sets of points as (A9), obtained from ψ by deleting a complementary pair of triangles, will be termed *nonads*. Since ψ contains 20 triangles, it will contain ten nonads.

The points of a nonad have the property that each one is connected to just one point of each of the complementary triangles. Thus if *abc* is any permutation of *ijk*, and *def* is any permutation of *lmn*, then the point p_{ad} in (A8) lies on the line $\lambda(ad, bc, ef)$. [The other two lines through p_{ad} are, of course, that row and that column of the array (A9) which meet at *ad*.] The nine points of the nonad (A9) in this way give rise to a *nonad of lines* that are *transversals* of the (ν, τ, τ') decomposition of the $(9 + 3 + 3 =)15$ points of ψ, and that account for the $3 \times 3 = 9$ lines joining points of a triangle to points of the complementary triangle.

Observe that we have shown that ψ possesses precisely *ten* different $(9, 3, 3)$-decompositions,

$$\psi = \nu_{ijk} \cup \tau_{ijk} \cup \tau'_{ijk}, \tag{A11}$$

into a nonad $\nu_{ijk} = \nu_{lmn}$ and a complementary pair of triangles τ_{ijk} and $\tau'_{ijk} = \tau_{lmn}$. Dually L possesses precisely ten different $(9, 3, 3)$-decompositions,

$$L = T_{ijk} \cup R_{ijk} \cup R'_{ijk}, \tag{A12}$$

into a complementary pair of skew triplets of lines R_{ijk}, R'_{ijk} [the rows and columns of the array (A9)] and a nonad of transversals, T_{ijk} say (as in the preceding paragraph). Just as the joins of τ to τ' yield the nine lines of T, the intersections of R and R' yield the nine points of ν. Also just as each line of T_{ijk} is transversal of the (ν, τ, τ')-decomposition (A11), each point of ν_{ijk} is a "meet" of the (T, R, R')-decomposition (A12). Incidentally a $(9, 3, 3)$-decomposition of a given 15_3 configuration is determined by a choice of dyad, for a dyad is necessarily of the form $\{p_{ij}, p_{ik}\}$ and hence determines a triangle τ_{ijk} (and hence also the complementary triangle $\tau'_{ijk} = \tau_{lmn}$).

While on the topic of decompositions of ψ, note that each line $\lambda \in L$ gives rise to a (5, 5, 5) decomposition of ψ as follows. If $\lambda = \lambda(ij, kl, mn)$, let $\beta_{ij}, \beta_{kl}, \beta_{mn}$ denote the bipods, whose apexes are, respectively, p_{ij}, p_{kl}, p_{mn}, formed from the three pairs of legs of the sixwig with body λ. Viewing these bipods in terms of the five-sets of points in which they intersect ψ, we have arrived at a (5, 5, 5) decomposition of ψ:

$$\psi = \beta_{ij} \cup \beta_{kl} \cup \beta_{mn}. \tag{A13}$$

There are precisely 15 such decompositions, one for each choice of line $\lambda \in L$.

Finally (before dealing with the case that interests us, namely, $\mathbb{F} = \mathbb{F}_2$), let us briefly mention some further aspects of the high symmetry of our 15_3 configuration. One easily checks that each point of ψ lies in precisely six nonads, each line of L lies in precisely four nonads, and each bipod lies in precisely two nonads. In fact, two distinct nonads always intersect in the five points of a bipod. In detail, ν_{ijk} and ν_{ijl} are seen to intersect in the bipod obtained from the tripod $\tau(p_{kl})$ by deletion of the points p_{ij}, p_{mn}, and ν_{ijk} and ν_{ilm} are seen to intersect in the bipod obtained from the tripod $\tau(p_{in})$ by deletion of the points p_{jk}, p_{lm}.

Suppose now, at last, that $\mathbb{F} = \mathbb{F}_2$. Then the above construction of a 15_3 configuration, via (A1) and (A2), fails because, in addition to the 15 collinearities (A3), there are now 20 others consisting of "triangles" p_{ij}, p_{jk}, p_{ki}. This applies for any field $\mathbb{F}$ of characteristic 2, and, in fact, we see that the configuration arising from (A1) and (A2) is merely that of the 15 points and 35 lines of a PG(3, $\mathbb{F}_2$) subgeometry of PG(4, $\mathbb{F}$).

Let us therefore start afresh in the following manner. Choose a simplex of reference for PG(4, 2), and let its vertices be denoted v_1, v_2, v_3, v_4, v_5, and let its associated "center," or unit point, be

$$c = v_1 + v_2 + v_3 + v_4 + v_5. \tag{A14}$$

Setting $v_0 = 0$, we define a set $\bar{\psi}$ of 15 points $\bar{p}_{ij} = \bar{p}_{ji}$, $0 \leqslant i \neq j \leqslant 5$, and a second set ψ of 15 points $p_{ij} = p_{ji}$, $0 \leqslant i \neq j \leqslant 5$, by

$$\bar{p}_{ij} = v_i + v_j, \quad p_{ij} = c + \bar{p}_{ij}. \tag{A15}$$

An easy check now shows that three distinct points p_{ij}, p_{kl}, p_{mn}, chosen from the second set ψ of 15 points, lie on a line if and only if $ijklmn$ is a permutation of 012345—exactly as previously in (A3). *Thus we have constructed* a 15_3 configuration (ψ, L) over $\mathbb{F}_2$ (or indeed over any field of characteristic 2). Moreover, with one slight exception, *everything we said from (A 3) onwards applies equally to the present* 15_3 *configuration over* $\mathbb{F}_2$. [The exception is (A10), though this can serve still as a mnemonic for the construction of the 20 different nonads of the configuration.]

There are some features peculiar to characteristic 2. For a start the points of each pentad π_i in (A4) sum to zero, and so each pentad π_i is a three-frame, which defines some solid, say σ_i, $0 \leqslant i \leqslant 5$. Next note that if α_{ijk} denotes the plane of the triangle τ_{ijk} in (A6), then each of the 20 planes α_{ijk} passes through the center c, since from (A14) and (A15) we have

$$p_{ij} + p_{jk} + p_{ki} = c. \tag{A16}$$

The center c is thus a privileged 16th point uniquely determined by the 15 points of ψ. Each of the 15 solid σ_{ij}, say, determined by the 15 tripods $\tau(p_{ij})$ also passes through c, since each tripod contains triangles. Incidentally, over $\mathbb{F}_2$, the priveleged center c "of" ψ is picked out by the property that it alone [amongst the 31 points of PG (4, 2)] does not lie on any of the joins $j(p, q), p, q \in \psi$. Of course, one feature that is obviously peculiar to $\mathbb{F}_2$ is the fact that each line of L consists entirely of (three) points of ψ.

Concerning the ten (9, 3, 3)-decompositions of the kind (A11), observe that for each of them the associated complementary pair of slew planes, α_{ijk} and $\alpha'_{ijk} = \alpha_{lmn}$, have the same pivot, namely, the center c of ψ. Concerning the (9, 3, 3)-decomposition (A12), we may now (over $\mathbb{F}_2$) refer to R_{ijk} and $R'_{ijk} = R_{lmn}$ as *complementary reguli* (see, e.g., Hirschfeld[5]). Of course, over $\mathbb{F}_2$, the (9, 3, 3)-decomposition (A11) can be expressed in the form

$$\psi = \sigma_{ijk}\alpha_{ijk}\,\alpha'_{ijk} \tag{A17}$$

encountered in (4.14), where σ_{ijk} denotes the solid defined by the nonad ν_{ijk} (or equally by the regulus R_{ijk}). We have thus justified our computation after (4.14) of $|\Theta_7|$. Similarly the (5, 5, 5)-decomposition (A13) can be expressed in the form

$$\psi = \alpha_{ij}\alpha_{kl}\alpha_{mn} \tag{A18}$$

encountered in (4.7), where α_{ij} denotes the plane of the bipod β_{ij}, etc. We have thus justified our computation of $|\Theta_7|$ in (4.10).

Remark: Over $\mathbb{F}_2$ the figure $\bar{\psi}$, consisting of the first set $\{\bar{p}_{ij}\}$ of 15 points in (A15), possesses properties that are in many ways "opposite" to those of ψ. For example, note that the analog of the 20 pairs $\tau_{ijk}, \tau'_{ijk} = \tau_{lmn}$ of triangles of ψ are the 20 pairs $\lambda_{ijk}, \lambda'_{ijk} = \lambda_{lmn}$ of lines of $\bar{\psi}$, where

$$\lambda_{ijk} = \bar{p}_{ij}\bar{p}_{jk}\bar{p}_{ki}. \tag{A19}$$

These 20 lines $\bar{L}$, say, are the only lines formed from points of $\bar{\psi}$, and so we have a configuration $(\bar{\psi}, \bar{L})$ consisting of 15 points and 20 lines such that each line of $\bar{L}$ contains three points of $\bar{\psi}$ and through each point of $\bar{\psi}$ pass four lines of $\bar{L}$. Just as each point of τ_{ijk} connects, within the configuration (ψ, L), to each point of τ'_{ijk}, no point of λ_{ijk} connects, within the configuration $(\bar{\psi}, \bar{L})$, to any point of λ'_{ijk}. Given a line $\lambda_{ijk} \in L$, and hence given also the line λ'_{ijk}, the remaining $(15-3-3=)9$ points of $\bar{\psi}$ can be obtained as follows. Join λ_{ijk} in turn to each of the three points of the triangle τ'_{ijk} to obtain three one-planes of the (ψ, L) configuration [see (5.1)], namely, $\alpha_{lm}, \alpha_{mn}, \alpha_{nl}$. Of course these planes are six-planes of the $(\bar{\psi}, \bar{L})$ configuration and, discounting the three points of λ_{ijk}, they contribute the required $3+3+3=9$ points of $\bar{\psi}$ lying outside λ_{ijk} and λ'_{ijk}.

The analogs of the six pentads π_i in (A4), the totally unconnected five-sets in ψ, are the six *simplices*

$$\bar{\pi}_i = \{\bar{p}_{ij};\ j \neq i\}, \quad 0 \leqslant i \leqslant 5, \tag{A20}$$

which are picked out by their being the completely connected five-sets in $\bar{\psi}$. Each simplex generates the remaining ten points of $\bar{\psi}$ by way of the ten further points lying on the ten edges (joins of pairs of vertices) of the simplex. Note also that all six simplices (A20) have the same center

$c = \Sigma_{j \neq i} \bar{p}_{ij}$. The center c of $\bar{\psi}$ is also picked out by the property that it alone [amongst the 31 points of PG(4, 2)] does not lie on any of the joins $j(p, q)$, $p, q \in \bar{\psi}$.

It is worth noting that we can compute $|\Theta_7|$ equivalently in terms of the number of $(\bar{\psi}, \bar{L})$ configurations. Consequently we have $|\Theta_7| = N''/6$, where N'' denotes the number of (nonordered) simplicies of PG(4, 2) and where we need to divide by six because each $\bar{\psi}$ figure contains precisely six simplices of the kind (A20) (which generate $\bar{\psi}$ in the manner described above). Consequently

$$|\Theta_7| = 31\times 30\times 28\times 24\times 16/(5!\times 6) = 13\,888, \tag{A21}$$

in agreement with (4.10) and (4.15).

Remark: Our construction of a 15_3 configuration over the field $\mathbb{F}_2$ did not treat the six vectors $v_0, \ldots, v_5$, which span $V(5)$, in a completely democratic fashion since we imposed the condition $v_0 = 0$. In fact, any condition

$$\sum_{i=0}^{5} a_i v_i = 0, \quad a_i \in \mathbb{F}_2, \tag{A22}$$

could be used instead of $v_0 = 0$, *provided that an odd number of the a_i are nonzero,*

$$\sum_{i=0}^{5} a_i = 1, \tag{A23}$$

and provided that we define c by

$$c = \sum_{i=0}^{5} v_i. \tag{A24}$$

A more democratic treatment is possible if we view $V(5)$ as a subspace of $V(6)$ (cf. Edge[8]). Choose any basis $\{v_0, \ldots, v_5\}$ for $V(6;\mathbb{F}_2)$ and let $V(5)$ denote the associated "even" subspace, consisting of vectors having an even number of nonzero coordinates. The 31 points of $\mathrm{PG}(4, 2) = V(5)\setminus\{0\}$ in this view consist of the point c, *defined now by (A24)*, together with the two sets $\bar{\psi}$, ψ of 15 points defined by (A15).

[1]R. Shaw, J. Math. Phys. **30**, 1971 (1989).

[2]J. W. P. Hirschfeld, *Projective Geometries over Finite Fields* (Clarendon, Oxford, 1979).

[3]See, for example, L. E. Dickson, *Linear Groups with an Exposition of the Galois Field Theory* (Dover, New York, 1958).

[4]This is proved in Ref. 2, following the original proof given by R. C. Bose and R. C. Burton, J. Comb. Theory **1**, 96 (1966).

[5]J. W. P. Hirschfeld, *Finite Projective Spaces of Three Dimensions* (Clarendon, Oxford, 1985).

[6]R. Shaw, J. Phys. A: Math. Gen. **21**, 7 (1988).

[7]H. F. Baker, *Principles of Geometry*, Vol. 4 (Cambridge U. P., Cambridge, 1925); H. W. Richmond, Math. Ann. **53**, 161 (1900); H. S. M. Coxeter, Bull. Am. Math. Soc. **56**, 413 (1950).

[8]W. L. Edge, Canadian J. Math. **11**, 625 (1959).

Explicit orthonormal Clebsch–Gordan coefficients of SU(3)

Sigitas Ališauskas[a)]
Institute of Physics, Academy of Sciences of the Lithuanian SSR, Vilnius 232600, USSR

(Received 20 June 1989; accepted for publication 10 January 1990)

The construction of the explicit algebraic-polynomial expressions for the nonmultiplicity-free orthonormal Clebsch–Gordan (Wigner) coefficients of $SU(3)\supset U(2)$ is completed in the case of the paracanonical coupling scheme related with the explicit minimal biorthogonal systems by means of the Hecht or Gram–Schmidt process. The direct and inverse orthogonalization coefficients (the first of them being equivalent to the boundary orthonormal isofactors) are expressed, up to explicitly given multiplicative factors, in terms of the numerator and denominator polynomials related with the auxiliary A_λ function of Louck, Biedenharn, and Lohe that appears as a fragment of the denominator G-functions of canonical SU(3) tensor operators.

I. INTRODUCTION

In a previous paper[1] (referred to hereafter as Paper I), the alternative approaches to the nonmultiplicity-free Clebsch–Gordan (Wigner) coefficients of $SU(3)\supset U(2)$ have been considered. As it was demonstrated, the Gram–Schmidt processes applied to the explicit biorthogonal systems of the SU(3) isofactors (reduced Clebsch–Gordan–Wigner coefficients) lead to the different algebraic systems of the orthonormal SU(3) isofactors that are determined by the additional selection rules, or by the null space structure and symmetries of isofactors. They include two versions of the paracanonical coupling[1] (suggested in fact by Hecht[2]), three versions of the canonical coupling (introduced and developed by Biedenharn and collaborators[3,4]), along with six versions of the pseudocanonical coupling.[1] In Paper I, the mutual expansion of the biorthogonal systems associated with the canonical and paracanonical systems has been established, as well as the symmetry and the main features of the polynomial structure of the boundary paracanonical SU(3) isofactors (orthogonalization coefficients). Following the ideas of Ref. 3, the boundary paracanonical or pseudocanonical isofactors have been expressed in Paper I in terms of the specific numerator and denominator functions (the last functions being special cases of the first ones). For fixed value of a definite parameter and the multiplicity labels, the numerator–denominator functions can be expressed as polynomials of definite degree in the remaining (five) independent parameters. The corresponding numerator–denominator polynomials were expressed originally in terms of the overlap determinants divided by the product of the linear functions. Unfortunately, the author of Paper I was not acquainted with the conjecture,[5] proved in Ref. 4, about the explicit polynomial form of the denominator function of the canonical SU(3) tensor operators.

In the present paper the remarkable ideas of Refs. 4 and 5 are generalized for explicit expression of the boundary paracanonical orthonormal isofactors (the direct orthogonalization coefficients, depending on eight parameters) as well as for the inverse orthogonalization coefficients of the minimal biorthogonal system of the SU(3) isofactors. The solution of the boundary value problem (see Paper I and references therein) allows one to write the explicit expressions for the general paracanonical orthonormal SU(3) isofactors (depending on 12 parameters).

In the present section the main definitions and the properties of the paracanonical isofactors are given. In Sec. II the polynomial structure of the boundary paracanonical isofactors is discussed, as well as the reduction formulas, symmetries, and zeros of the numerator–denominator functions, which determine the uniqueness of these functions. In Sec. III an expression (in two different forms) is presented for the numerator–denominator polynomials, that is proved on the grounds of the results of Ref. 4. In Sec. IV the explicit construction that leads to the inverse (dual) orthogonalization coefficients is presented.

The irreducible representations (irreps) of SU(3) will be denoted as (ab), where $a=m_{13}-m_{23}$, $b=m_{23}-m_{33}$, and $[m_{13}m_{23}m_{33}]$ is the Young frame. The basis states are labeled by the hypercharge $y=m_{12}+m_{22}-\frac{2}{3}(m_{13}+m_{23}+m_{33})$, the isospin $i=\frac{1}{2}(m_{12}-m_{22})$, and its projection $i_z=m_{11}-\frac{1}{2}(m_{12}+m_{22})$, where the integers m_{ij} form the Gel'fand–Tzetlin pattern. Sometimes the parameter

$$z=\tfrac{1}{3}(b-a)-\tfrac{1}{2}y=m_{23}-\tfrac{1}{2}(m_{12}+m_{22}) \tag{1.1}$$

is more convenient than y because the linear combinations

$$i\pm z,\quad a+z-i,\quad b-z-i \tag{1.2}$$

are non-negative integers. In the case of the coupling $(a'b')\times(a''b'')$ to (ab),

$$z=z'+z''+v, \tag{1.3}$$

where

$$v=\tfrac{1}{3}(a'+a''-a-b'-b''+b) \tag{1.4}$$

is an integer. The parameters of the highest weight state take the values

$$y_0=\tfrac{1}{3}(a+2b),\quad i_0=-z_0=\tfrac{1}{2}a, \tag{1.5a}$$

while for the lowest weight state

$$\bar{y}_0=-\tfrac{1}{3}(2a+b),\quad \bar{i}_0=\bar{z}_0=\tfrac{1}{2}b. \tag{1.5b}$$

[a)] Present address: Institute of Theoretical Physics and Astronomy, Academy of Science of Lithuania, Vilnius 232600, Lithuania.

 0022-2488/90/061325-08$03.00

For the first version of the paracanonical coupling the multiplicity label $\widetilde{I}$ (intrinsic isospin of the Gel'fand–Weyl–Biedenharn pattern) satisfies the conditions

$$\widetilde{I}\pm\tilde{z}\geqslant 0,\quad a+\tilde{z}-\widetilde{I}\geqslant 0,\quad b-\tilde{z}-\widetilde{I}\geqslant 0, \tag{1.6a}$$

$$\widetilde{I}\pm\tilde{i}_z\geqslant 0,\quad i_0'+\bar{i}_0''-\widetilde{I}\geqslant 0, \tag{1.6b}$$

$$\widetilde{I}\geqslant B, \tag{1.6c}$$

where

$$B=\tfrac{1}{2}(a+b-b'-a''+|v|), \tag{1.7a}$$

$$\tilde{i}_z=i_0'-\bar{i}_0''=\tfrac{1}{2}(a'-b''),\quad \tilde{z}=\tfrac{1}{2}(b''-a')+v. \tag{1.7b}$$

The paracanonical isofactors (to here their first version if it is not asserted otherwise) satisfy the following symmetries (cf. Ref. 6):

$$(a'b'y'i'a''b''y''i''\|abyi;\widetilde{I})=(b''a''-y''i''b'a'-y'i'\|ba-yi;\widetilde{I}) \tag{1.8a}$$

$$=(-1)^{v+z'+i''-i}\left[\frac{\dim(ab)(2i''+1)}{\dim(a''b'')(2i+1)}\right]^{1/2}(ba-yi\,a'b'y'i'\|b''a''-y''i'';\widetilde{I}'') \tag{1.8b}$$

$$=(-1)^{v+i'-z''-i}\left[\frac{\dim(ab)(2i'+1)}{\dim(a'b')(2i+1)}\right]^{1/2}(a''b''y''i''ba-yi\|b'a'-y'i';\widetilde{I}'), \tag{1.8c}$$

where [see correspondence (2.17) of Paper I]

$$\widetilde{I}=\tfrac{1}{2}(b-b'-v)+\widetilde{I}'=\tfrac{1}{2}(a-a''+v)+\widetilde{I}'', \tag{1.9a}$$

$$\dim(ab)=\tfrac{1}{2}(a+1)(b+1)(a+b+2). \tag{1.9b}$$

[Contrary to the case of the canonical tensor operators, the contragrediency transformation ($a'\leftrightarrow b'$, $a''\leftrightarrow b''$, $a\leftrightarrow b$) interchanges the two versions of the paracanonical coupling as well as the transposition of the states to be coupled.]

The paracanonical isofactors vanish unless

$$i-\widetilde{I}\leqslant i_0'+z'+\bar{i}_0''-z'',\quad i'+\tfrac{1}{2}(b-b'-v)-\widetilde{I}\leqslant i_0''+z''+i_0+z,\quad i''+\tfrac{1}{2}(a-a''+v)-\widetilde{I}\leqslant\bar{i}_0-z+\bar{i}_0'-z'. \tag{1.10}$$

After all, the paracanonical splitting is completely determined by the condition of vanishing[2] of the boundary isofactors

$$(a'b'y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\bar{y}i;\widetilde{I})=0$$

for $\widetilde{I}\leqslant i$ [which belong to conditions (1.10)].

Now, the general paracanonical isofactors of SU(3) may be expanded as follows:[1]

$$(a'b'y'i'a''b''y''i''\|abyi;\widetilde{I})$$

$$=\sum_{\tilde{i}}((a'b'y'i'a''b''y''i''\|abyi))^{-,+,\tilde{i}}(a'b'y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\bar{y}\tilde{i};\widetilde{I}) \tag{1.11a}$$

$$=\sum_{\rho j}(a'b'y'i'a''b''y''i''\|abyi;\rho)(a'b'y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\bar{y}j;\rho)A_{jI}^{(a'b'a''b'';ab)}$$

$$\equiv\sum_{j}(a'b'y'i'a''b''y''i''\|abyi)_{-,+,j}A_{jI}^{(a'b'a''b'';ab)}, \tag{1.11b}$$

where the first factors in the rhs belong to the minimal biorthogonal system of the SU(3) isofactors, expressed according to Eqs. (2.22), (3.2a), and (3.2b) of Paper I. The boundary isofactors and the inverse orthogonalization coefficients satisfy the biorthogonality relations:

$$\sum_{\tilde{i}}(a'b'y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\bar{y}\tilde{i};\widetilde{I})A_{\tilde{i}J}^{(a'b'a''b'';ab)}=\delta_{\widetilde{I}J}, \tag{1.12a}$$

$$\sum_{\widetilde{I}}A_{jI}^{(a'b'a''b'';ab)}(a'b'y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\bar{y}\tilde{i};\widetilde{I})=\delta_{\tilde{i}j}. \tag{1.12b}$$

In order to represent the symmetry of the boundary isofactors and the numerator–denominator polynomials in Paper I, the following 3×6 array $|q_{\alpha\beta}|$ was introduced:

$$|q_{\alpha\beta}|=\begin{vmatrix} b'-a''+a+v & a'-b''+b-v & b-v & b & b' & b'+v\\ a & a+v & a-a'+b''+v & a''-b'+b-v & a''-v & a''\\ a'-v & a' & b'' & b''+v & b'+b''-b+v & a'+a''-a-v\end{vmatrix}, \tag{1.13}$$

where min $q_{\alpha\beta} + 1$ ($\alpha = 1,2,3; \beta = 1,2,...,6$) gives the external multiplicity of the irrep (ab) in $(a'b') \times (a''b'')$ and

$$q_{\alpha 2} - q_{\alpha 1} = q_{\alpha' 4} - q_{\alpha' 3} = q_{\alpha'' 6} - q_{\alpha'' 5} = v, \quad q_{\alpha\beta} - q_{\alpha\beta'} = q_{\alpha'\beta} - q_{\alpha'\beta'} . \tag{1.14}$$

The function [see (4.8) of Paper I]

$$\frac{(a'b'y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\tilde{y}i;\tilde{I})}{[\dim(ab)(a'+1)(b''+1)M(a'+b',b')M(a''+b'',a'')]^{1/2}} \tag{1.15}$$

[where $M(h_1h_2) = (h_1+1)!h_2!/(h_1-h_2+1)$] is invariant under the 24 transformations of array (1.13), generated by the transpositions of rows, by the transposition of the left and middle couples of columns, and by the transposition of the even and odd columns in array (1.13).

II. POLYNOMIAL STRUCTURE OF THE BOUNDARY PARACANONICAL ISOFACTORS

The boundary paracanonical isofactors for $q_{11} \equiv b' - a'' + a + v \equiv b - 2\tilde{z}$, $i - \tilde{z}$, and $b - \tilde{z} - \tilde{I}$ fixed may be expressed [cf. Eq. (4.2) of Paper I] as follows:

$$(a'b''y_0'i_0'a''b''\bar{y}_0''\bar{i}_0''\|ab\tilde{y}i;\tilde{I}) = \left[\frac{2\dim(ab)b'^{(b-\tilde{z}-\tilde{I})}(a'+b'+1)^{(b-\tilde{z}-\tilde{I})}(b-\tilde{z}+i+1)^{(b-\tilde{z}-\tilde{I})}}{a''^{(-1)(\tilde{I}-\tilde{z}+1)}(a''+b''+1)^{(-1)(\tilde{I}-\tilde{z}+1)}(i+\tilde{z})^{(-1)(\tilde{I}-\tilde{z}+1)}}\right]^{1/2} (-1)^{\tilde{I}-i}\mathbb{K}_{\tilde{I},i}\, g_{\tilde{I},i}[g_{\tilde{I},\tilde{I}}\, g_{\tilde{I}+1,\tilde{I}+1}]^{-1/2} \tag{2.1}$$

for $\min(b-\tilde{z}, a+\tilde{z}, i_0' + \bar{i}_0'') \geqslant \tilde{I} \geqslant i \geqslant \max(|\tilde{i}_z|, |\tilde{z}|)$ and vanish otherwise. Here

$$X^{(y)} = (X-y)^{(-1)y} = X(X-1)\cdots(X-y+1) = X!/(X-y)!, \tag{2.2}$$

$$\mathbb{K}_{j,i} = [(j+\tilde{i}_z)^{(j-i)}(j-\tilde{i}_z)^{(j-i)} \times (i_0' + \bar{i}_0'' - i)^{(j-i)}(i_0' + \bar{i}_0'' + j + 1)^{(j-i)} \times (a+\tilde{z}-i)^{(j-i)}(a+\tilde{z}+j+1)^{(j-i)} \times (j-\tilde{z})^{(j-i)}(b-\tilde{z}-i)^{(j-i)}]^{1/2}/(j-i)!. \tag{2.3}$$

The functions $g_{\tilde{I},i}(q_{\alpha\beta})$ (where $b - \tilde{z} \geqslant \tilde{I} \geqslant i \geqslant \tilde{z}$ and parameters q_{11}, $b - \tilde{z} - I$, and $\tilde{I} - i$ are fixed) are the polynomials with some rational (for $\tilde{I} = i$ integer) coefficients in five free [see Eqs. (1.14)] parameters of the total degree

$$3(b-\tilde{z}-\tilde{I}+1)(\tilde{I}-\tilde{z}) - 3(\tilde{I}-i) . \tag{2.4}$$

Our definition of $g_{\tilde{I},i}$ (together with the phase factor) is slightly changed to compare with Eqs. (4.6a) and (4.6b) of Paper I by substituting $\mathbb{K}_{\tilde{I},i}$ instead of $K_{\tilde{I},i}$. The multiplication of our polynomial $g_{\tilde{I}i}$ ($\tilde{I} \neq i$) by the maximal rational factor of

$$\left[\binom{\tilde{I}-\tilde{z}}{\tilde{I}-i}\binom{b-\tilde{z}-i}{\tilde{I}-i}\right]^{1/2} \tag{2.5}$$

[where $\binom{m}{n}$ are binomial coefficients] makes all its coefficients integer. Since the determination of this factor presents a number–theoretical problem, our definition of $g_{\tilde{I},i}$ is more convenient for formulating of the reduction formulas.

The polynomials $g_{\tilde{I},i}$ are invariant with respect to the S_4 group generated by the substitutions

$$a' \to -a' - 2, \quad b' \to a' + b' + 1 \quad (v \to v - a' - 1), \tag{2.6a}$$

$$a'' \to a'' + b'' + 1, \quad b'' \to -b'' - 2 \quad (v \to v + b'' + 1), \tag{2.6b}$$

and the transposition of the second and third rows of array (1.13),

$$a \to a' - v, \quad a' \to a + v, \quad b'' \to a - a' + b'' + v, \quad a'' \to a' + a'' - a - v . \tag{2.6c}$$

The invariance of $g_{\tilde{I},i}$ with respect to the substitution (2.6a) follows from Eqs. (2.20), (3.3), (3.5), or (3.6) of Paper I; the invariance with respect to the composition of substitutions (2.6a) and (2.6b) is discussed in Paper I. Besides, the polynomials $g_{\tilde{I},\tilde{I}}$ remain unchanged after the permutation of the middle and the right couples of columns of array (1.13).

The remaining symmetry properties of function (1.15) together with Eq. (2.1) allow one to deduce the following reduction formulas of $g_{\tilde{I},i}$, respectively, for $q_{11} - b + \tilde{z} + \tilde{I} \leqslant q_{21} \leqslant q_{11}$, for $-i + \tilde{z} \leqslant v \leqslant 0$, or for $q_{11} - i + \tilde{z} \leqslant q_{13} \leqslant q_{11}$:

$$g_{\tilde{I},i}(q_{\alpha\beta}) = \prod_{s=1}^{\tilde{I}-\tilde{z}} (b'+v+s)^{(q_{11}-q_{21})}(b'+b''+v+1+s)^{(q_{11}-q_{21})} \prod_{s=1}^{i-\tilde{z}} (b+s)^{(q_{11}-q_{21})} \times \prod_{s=i-\tilde{z}+1}^{\tilde{I}-\tilde{z}} (b+1+s)^{(q_{11}-q_{21})} g_{\tilde{I},i}(q_{1\beta} \leftrightarrow q_{2\beta}) , \tag{2.7a}$$

$$= (i-\tilde{i}_z)^{(-v)} \prod_{s=1}^{b-\tilde{z}-\tilde{I}+1} (b'+1-s)^{(-v)}(b'+b''+a+v+3-s)^{(-v)}$$

$$\times \prod_{s=1}^{b-\tilde{z}-\tilde{I}} (b-v+1-s)^{(-v)} g_{\tilde{I},i}(q_{\alpha 1}\leftrightarrow q_{\alpha 2}, q_{\alpha 3}\leftrightarrow q_{\alpha 4}, q_{\alpha 5}\leftrightarrow q_{\alpha 6}), \quad (2.7b)$$

$$= \prod_{s=1}^{b-\tilde{z}-\tilde{I}+1} (a'+b'+2-s)^{(q_{11}-q_{13})}(a+b'+v+2-s)^{(q_{11}-q_{13})}$$

$$\times (i+\tilde{i}_z)^{(q_{11}-q_{13})} \prod_{s=1}^{b-\tilde{z}-\tilde{I}} (q_{12}+1-s)^{(q_{11}-q_{13})} g_{\tilde{I},i}(q_{\alpha 1}\leftrightarrow q_{\alpha 3}, q_{\alpha 2}\leftrightarrow q_{\alpha 4}). \quad (2.7c)$$

The zeros of the polynomial $g_{\tilde{I},\tilde{I}}$ form at least the weight space $\mathbb{W}_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v,-b'-v)$, where the lattice (integer) points (b',v) are restricted by the conditions

$$0 \leqslant b' \leqslant q_{11}-1, \quad -q_{11} \leqslant v \leqslant -1,$$
$$-\tilde{I}+\tilde{z} \leqslant b'+v \leqslant b-\tilde{z}-\tilde{I}-1. \quad (2.8)$$

The points of $\mathbb{W}_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v-b'-v)$ in the Möbius plane $\mathbb{M}$ are in one-to-one correspondence with those of the weight space of the irrep $(b-\tilde{z}-\tilde{I}, \tilde{I}-\tilde{z}-1)$ of SU(3). The corresponding zeros have the multiplicity of the weight

$$M_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v) = \min(b-\tilde{z}-\tilde{I}+1, \tilde{I}-\tilde{z}, 1+d(b',v)), \quad (2.9)$$

where $d(b',v)$ is the distance from the lattice point (b',v) to the nearest boundary point of $\mathbb{W}_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v,-b'-v)$ in $\mathbb{M}$.

The zeros of the polynomials $g_{\tilde{I},i}$ form at least the truncated weight space $\mathbb{W}_{q_{11}[\tilde{I}-i]}^{b-\tilde{z}-\tilde{I}+1}(b',v,-b'-v)$, which is obtained from $\mathbb{W}_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v,-b'-v)$ with the decreased by unity multiplicities of the lattice points corresponding to the solutions of the equation

$$(v-\tilde{z}+\tilde{I})^{(\tilde{I}-i)} = 0. \quad (2.10)$$

The proof follows from the analysis of the null spaces[7] of the paracanonical tensor operators. According to Eq. (2.22) of Paper I, the first factors in the rhs of Eq. (1.11a) vanish (have the simple zeros under the square root) unless

$$\tilde{i} \pm \tilde{z} \geqslant 0, \quad \tilde{i} \pm \tilde{i}_z \geqslant 0, \quad (2.11a)$$
$$b-\tilde{z}-\tilde{i} \geqslant 0, \quad a+\tilde{z}-\tilde{i} \geqslant 0, \quad i_0'+\tilde{i}_0''-\tilde{i} \geqslant 0. \quad (2.11b)$$

Conditions (2.11a) ensure the vanishing of the superfluous paracanonical isofactors with the lower values of $\tilde{I}$ for $q_{21} < q_{11}$, or $q_{31} < q_{11}$, as well as for $q_{41} < q_{11}$. The factor $b'^{(b-\tilde{z}-\tilde{I})}$ in Eq. (2.1) ensures the vanishing of the superfluous isofactors for $q_{51} \equiv b' < q_{11}$ when $v \geqslant 0$. [The first factor in the rhs of Eq. (1.11) vanishes for $b' < 0$.] Finally, the vanishing of the superfluous isofactors for $q_{61} \equiv b'+v < q_{11}$ is ensured by the correlation of the zeros in the numerator and the denominator functions of Eq. (2.1) together with the zeros of the factors $b'^{(b-\tilde{z}-\tilde{I})}$, $\mathbb{K}_{\tilde{I},i}$, and the first factor in the rhs of Eq. (1.11).

Similarly to Lemma 2.2 of Ref. 4, it may be demonstrated that the polynomial in the five free parameters with the weight space of zeros $\mathbb{W}_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v,-b'-v)$ and the S_4 symmetry generated by substitutions (2.6) is of total degree at least $3(b-\tilde{z}-\tilde{I}+1)(\tilde{I}-\tilde{z})$.

Up to multiplicative factor the polynomial $g_{\tilde{I},\tilde{I}}$ is the unique polynomial of the total degree $3(b-\tilde{z}-\tilde{I}+1)(\tilde{I}-\tilde{z})$ with the S_4 symmetry (2.6) and the weight space of zeros $\mathbb{W}_{q_{11}}^{b-\tilde{z}-\tilde{I}+1}(b',v,-b'-v)$ which satisfies the reduction formulas (2.7).

Proof: Suppose this theorem is valid for $q_{11}=1,2,\ldots,q-1$. The reduction formulas (2.7) allow us to express the function $g_{\tilde{I},\tilde{I}}$ for the all values of $q_{21},q_{31},q_{12},q_{13}$ that satisfy the conditions

$$q-b+\tilde{z}+\tilde{I} \leqslant a \leqslant q-1, \quad q-b+\tilde{z}+\tilde{I} \leqslant a'-v \leqslant q-1,$$
$$q-\tilde{I}+\tilde{z} \leqslant q+v \leqslant q-1, \quad q-\tilde{I}+\tilde{z} \leqslant b-v \leqslant q-1. \quad (2.12)$$

The possible variation polynomial should have zeros in the corresponding hyperplanes and therefore, it should include the factor

$$H_q = (v+q)^{(q)}(b-v)^{(q)}(a'-v)^{(q)}a^{(q)}$$
$$\times (a+b+1)^{(q)}(b''+v+q+1)^{(q)}. \quad (2.13)$$

Besides for $b-\tilde{z}-\tilde{I}>0$ and $\tilde{I}-\tilde{z}>1$ the variation polynomial should include as a factor a polynomial with the S_4 symmetry (2.6) and a weight space of zeros $\mathbb{W}_{q_{11}-2}^{b-\tilde{z}-\tilde{I}}(b'-1,v+1,-b'-v)$. Thus the total degree of the variation polynomial $6q$ (for $b-\tilde{z}-\tilde{I}=0$ or $\tilde{I}-\tilde{z}=1$) or $3(\tilde{I}-\tilde{z}-1)(b-\tilde{z}-\tilde{I})+6(b-2\tilde{z})$ exceeds $3(\tilde{I}-\tilde{z})(b-\tilde{z}-\tilde{I}+1)$.

Since polynomials $\mathbb{K}_{\tilde{I}}^2\, g_{\tilde{I},i}$ and $g_{\tilde{I},\tilde{I}}$ have the same weight space of zeros [with additional zeros corresponding to the solutions of Eq. (2.10) outside of region (2.8) in the first case] their total degree should be at least $3(b-\tilde{z}-\tilde{I}+1)(\tilde{I}-\tilde{z})+3(\tilde{I}-i)$. Thus the uniqueness of the polynomials $g_{\tilde{I},i}$ of the total degree (2.4) with S_4 symmetry (2.6) is also ensured.

III. EXPLICIT NUMERATOR–DENOMINATOR POLYNOMIALS

According to Sec. IV of Paper I,

$$g_{b-\tilde{z}+1,b-\tilde{z}+1} = g_{\tilde{z},\tilde{z}} = g_{b-\tilde{z},\tilde{z}} = 1,$$

when the final polynomials

$$g_{b-\tilde{z},i} = (1)^{b-\tilde{z}-i} E_{b-\tilde{z},i}/\mathbf{K}_{b-\tilde{z},i},$$
$$g_{\tilde{z}+1,\tilde{z}+1} = F^{\tilde{z},\tilde{z}}, \quad g_{\tilde{z}+1,\tilde{z}} = F^{\tilde{z}+1,\tilde{z}}/\mathbf{K}_{\tilde{z}+1,\tilde{z}}$$

may be expressed in terms of the Saalschutzian ${}_4F_3(1)$ series that appear[8] also in $6j$ (Racah) coefficients of SU(2). The relations between the terminating Saalschutzian ${}_4F_3(1)$ series[9] or the different expressions[10] for $6j$ coefficients allow one to present these polynomials in different forms.

In analogy with Refs. 4 and 5 some of these expressions may be extrapolated to the following expression:

$$g_{\tilde{I},i}(q_{\alpha\beta}) = \sum_{\mu\nu} \frac{M[(\tilde{I}-\tilde{z})^{b-\tilde{z}-\tilde{I}}, i-\tilde{z}]}{M(\mu)M(\nu)} (i+\tilde{i}_z)^{(i-\tilde{z}-\mu_t)}(i-i_z)^{(i-\tilde{z}-\nu_t)}$$
$$\times \prod_{s=1}^{t-1} (\tilde{I}+i_z+t-s)^{(\tilde{I}-\tilde{z}-\mu_s)}(\tilde{I}-\tilde{i}_z+t-s)^{(\tilde{I}-\tilde{z}-\nu_s)}$$
$$\times \prod_{s=1}^{t} (b'+v+1-s)^{(-1)\mu_s}(a'+a''+b-v+3-s)^{(-1)\mu_s}$$
$$\times \prod_{s=1}^{t} (a'+b'+1-t+s)^{\nu_s)}(a+b'+v+1-t+s)^{(\nu_s)} \tag{3.1a}$$
$$= \sum_{\mu^*\nu^*} \frac{M[(b-\tilde{z}-\tilde{I}+1)^{i-\tilde{z}}(b-\tilde{z}-\tilde{I})^{\tilde{I}-i}]}{M(\mu^*)M(\nu^*)}$$
$$\times \prod_{s=1}^{\tilde{I}-\tilde{z}} (v+s-\Delta_s)^{(-1)(b-\tilde{z}-\tilde{I}+\Delta_s-\mu_s^*)}(b''-a'+v+s-\Delta_s)^{(-1)(b-\tilde{z}-\tilde{I}+\Delta_s-\nu_s^*)}$$
$$\times \prod_{s=1}^{\tilde{I}-\tilde{z}} (b'+v+s)^{(\mu_s^*)}(a'+a''+b-v+2+s)^{(\mu_s^*)}$$
$$\times \prod_{s=1}^{\tilde{I}-\tilde{z}} (a'+b'-b+\tilde{z}+\tilde{I}-s+1)^{(-1)\nu_s^*}(a''+\tilde{I}-\tilde{z}-s+1)^{(-1)\nu_s^*}. \tag{3.1b}$$

Here $t = b-\tilde{z}-\tilde{I}+1$, $\Delta_s = 1$ for $1 \leqslant s \leqslant i-\tilde{z}$ and 0 otherwise. The partitions $\mu = [\mu_1\mu_2\cdots\mu_t]$, $\nu = [\nu_1\nu_2\cdots\nu_t]$ denote the irreps of $U(t)$. [Here, $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_t \geqslant 0$, etc.] The Young frames μ and ν accept such values that the irrep $\lambda' = [(\tilde{I}-\tilde{z})^{b-\tilde{z}-\tilde{I}}, i-\tilde{z}]$ appears (once) in the decomposition of the direct product of the irreps $\mu \times \nu$ of $U(t)$. Particularly

$$\mu_s + \nu_{t+1-s} = \tilde{I}-\tilde{z} \quad \text{for } i = \tilde{I},$$
$$\mu_s + \nu_{t-s} = \tilde{I}-\tilde{z} \quad \text{for } i = \tilde{z}, \tag{3.2}$$
$$\tilde{I}-\tilde{z}-\mu_{t-s+1} \geqslant \nu_s \geqslant \tilde{I}-\tilde{z}-\mu_{t-s},$$

with $\Sigma_{s=1}^{t}(\tilde{I}-\tilde{z}-\mu_s-\nu_{t-s+1}) = \tilde{I}-i$ (each term is non-negative in the lhs) for arbitrary i.

The symbol $M(\lambda)$ denotes the measure of the Young frame λ,

$$M(\lambda) = \frac{N!}{d_\lambda} = \frac{\Pi_{s=1}^{t}(\lambda_s+t-s)!}{\prod_{s<r}(\lambda_s-\lambda_r-s+r)}, \tag{3.3}$$

where d_λ is dimension of the irrep λ of S_N. The partitions λ'^*, μ^*, ν^* and Eq. (3.1b) are obtained from λ', μ, ν and Eq. (3.1a) after interchange of the rows and columns.

The total degree of polynomial (3.1), the (truncated if necessary) weight space of zeros, the reduction formulas (2.7b) and (2.7c), and the invariance with respect to substitution (2.6c) may be checked straightforward. The S_4 symmetry of the function[4,11,12]

$$A_{\lambda'}\begin{pmatrix} b'+v+1, a'+a''+b-v+3, a''+b-v+2, b'+b''+v+2 \\ v+t-a'-a''-b-b'-4 \end{pmatrix} \tag{3.4}$$

[see Eq. (5.8) of Ref. 4] with respect to the permutations of the upper four parameters allows one to prove the remaining symmetries and reduction formulas of polynomials $g_{\tilde{I},i}$.

The substitutions (2.6a), (2.6b) and their compositions with (2.6c) applied to Eq. (3.1) allow one to obtain the other versions of expression for $g_{\tilde{I},i}$.[13]

IV. EXPLICIT STRUCTURE OF THE INVERSE ORTHOGONALIZATION COEFFICIENTS

The dual Gram–Schmidt processes [see Eqs. (4.1a) and (4.1b) of Paper I] allow one to obtain also the expression of the inverse (dual) orthogonalization coefficients

$$A_{j\tilde{I}}^{(a'b'a''b'';ab)} = \frac{(2j+1)\mathbb{K}_{j\tilde{I}}g^{j,\tilde{I}}}{[\dim(ab)g_{\tilde{I},\tilde{I}}g_{\tilde{I}+1,\tilde{I}+1}]^{1/2}}\left[\frac{a''^{(-1)(\tilde{I}-\tilde{z}+1)}(a''+b''+1)^{(-1)(\tilde{I}-\tilde{z}+1)}(j+z)^{(-1)(\tilde{I}-\tilde{z})}}{b'^{(b-\tilde{z}-\tilde{I})}(a'+b'+1)^{(b-\tilde{z}-\tilde{I})}(b-\tilde{z}+j+1)^{(b-\tilde{z}-\tilde{I}+1)}}\right]^{1/2} \tag{4.1}$$

for $\min(b-\tilde{z},a+\tilde{z},i_0'+\tilde{i}_0'')\geqslant j\geqslant\tilde{I}\geqslant\max(|\tilde{i}_z|,|\tilde{z}|)$ which vanish otherwise. The functions $g^{j,\tilde{I}}(q_{\alpha\beta})$ (where $b-\tilde{z}\geqslant j\geqslant\tilde{I}\geqslant\tilde{z}$ and parameters q_{11}, $b-\tilde{z}-\tilde{I}$, and $j-\tilde{I}$ are fixed) are the polynomials with rational coefficients in five free [see Eqs. (1.14)] parameters of the total degree

$$3(b-\tilde{z}-\tilde{I})(\tilde{I}-\tilde{z}+1)-3(j-\tilde{I}). \tag{4.2}$$

Particularly,

$$g^{\tilde{I},\tilde{I}}=g_{\tilde{I}+1,\tilde{I}+1},\quad g^{\tilde{I}+1,\tilde{I}}=g_{\tilde{I}+1,\tilde{I}}. \tag{4.3}$$

Similarly to Eqs. (4.6a) and (4.6b) of Paper I, polynomials $g^{j,\tilde{I}}$ may be expressed in terms of the determinants of the overlaps $E_{\tilde{I},\tilde{J}}$ or $F^{\tilde{I},\tilde{J}}$ (in notation of Paper I):

$$g^{j\tilde{I}}=\det\begin{bmatrix} E_{b-\tilde{z},b-\tilde{z}}\cdots E_{b-\tilde{z},\tilde{I}+1} & 0\\ \cdots\cdots\cdots & \cdots\\ E_{j,b-\tilde{z}}\cdots E_{j,\tilde{I}+1} & 1\\ \cdots\cdots\cdots & \cdots\\ E_{\tilde{I},b-\tilde{z}}\cdots E_{\tilde{I},\tilde{I}+1} & 0\end{bmatrix}(b-\tilde{z}+j+1)^{(b-\tilde{z}-\tilde{I}+1)}\Big[(2j+1)(b-\tilde{z}+\tilde{I}+1)^{(b-\tilde{z}-\tilde{I})}$$

$$\times\mathbb{K}_{j,\tilde{I}}\prod_{j'=\tilde{I}+1}^{b-\tilde{z}-\tilde{I}}e_{j'}\Big]^{-1} \tag{4.4a}$$

$$=\det\begin{bmatrix}F^{\tilde{z},\tilde{z}}\cdots F^{\tilde{z},\tilde{I}-1} & F^{\tilde{z},j}\\ \cdots\cdots\cdots & \cdots\\ F^{\tilde{I},\tilde{z}}\cdots F^{\tilde{I},\tilde{I}-1} & F^{\tilde{I},j}\end{bmatrix}(\tilde{I}+\tilde{z})^{(-1)(\tilde{I}-\tilde{z})}\Big[(j+\tilde{z})^{(-1)(\tilde{I}-\tilde{z})}\mathbb{K}_{j\tilde{I}}\prod_{j'=\tilde{z}+1}^{\tilde{I}}f_{j'}\Big]^{-1}. \tag{4.4b}$$

Particularly,

$$g^{b-\tilde{z},b-\tilde{z}}=g^{\tilde{z}-1,\tilde{z}-1}=g^{b-\tilde{z},\tilde{z}}=1,$$
$$g^{b-\tilde{z}-1,b-\tilde{z}-1}=E_{b-\tilde{z},b-\tilde{z}},$$
$$g^{b-\tilde{z}-1,b-\tilde{z}-2}=-E_{b-\tilde{z},b-\tilde{z}-1}/\mathbb{K}_{b-\tilde{z},b-\tilde{z}-1},$$
$$g^{j,\tilde{z}}=F^{j,\tilde{z}}/\mathbb{K}_{j\tilde{z}}. \tag{4.4c}$$

Analogical analysis as in Sec. II allows us to check the symmetries of polynomial $g^{j\tilde{I}}$ and the reduction formulas (respectively, for $q_{11}-b+\tilde{z}+j\leqslant q_{21}\leqslant q_{11}$, for $-\tilde{I}+\tilde{z}\leqslant v\leqslant 0$, or for $q_{11}-\tilde{I}+\tilde{z}\leqslant q_{13}\leqslant q_{11}$):

$$g^{j\tilde{I}}(q_{\alpha\beta})=(b-\tilde{z}+j+1)^{(q_{11}-q_{21})}\prod_{s=1}^{\tilde{I}-\tilde{z}}(b+s)^{(q_{11}-q_{21})}$$
$$\times\prod_{s=1}^{\tilde{I}-\tilde{z}+1}(b'+v+s)^{(q_{11}-q_{21})}(b'+b''+v+1+s)^{(q_{11}-q_{21})}g^{j,\tilde{I}}(q_{1\beta}\leftrightarrow q_{2\beta}), \tag{4.5a}$$
$$=\prod_{s=1}^{b-\tilde{z}-\tilde{I}}(b'+1-s)^{(-v)}(b'+b''+a+v+3-s)^{(-v)}\prod_{s=1}^{b-\tilde{z}-j}(b-v+1-s)^{(-v)}$$
$$\times\prod_{s=b-\tilde{z}-j+1}^{b-\tilde{z}-\tilde{I}}(b-v-s)^{(-v)}g^{j,\tilde{I}}(q_{\alpha1}\leftrightarrow q_{\alpha2},\ q_{\alpha3}\leftrightarrow q_{\alpha4},\ q_{\alpha5}\leftrightarrow q_{\alpha6}), \tag{4.5b}$$
$$=\prod_{s=1}^{b-\tilde{z}-\tilde{I}}(a'+b'+2-s)^{(q_{11}-q_{13})}(a+b'+v+2-s)^{(q_{11}-q_{13})}$$
$$\times\prod_{s=1}^{b-\tilde{z}-j}(q_{12}+1-s)^{(q_{11}-q_{13})}\prod_{s=b-\tilde{z}-j+1}^{b-\tilde{z}-\tilde{I}}(q_{12}-s)^{(q_{11}-q_{13})}g^{j,\tilde{I}}(q_{\alpha1}\leftrightarrow q_{\alpha3},\ q_{\alpha2}\leftrightarrow q_{\alpha4}). \tag{4.5c}$$

The zeros of the polynomial $g^{j,\tilde{I}}$ form at least the truncated weight space $\mathbb{W}^{b-\tilde{z}-\tilde{I}}_{q_{11}(j-\tilde{I})}(b',v,-b'-v)$ which is obtained from $\mathbb{W}^{b-\tilde{z}-\tilde{I}}_{q_{11}}(b',v-b',v)$ [cf. condition (2.8) and Eq. (2.9)] with the decreased by unity multiplicities of the lattice points $(b'v)$ corresponding to the solutions of the equation

$$(v-\tilde{z}+j)^{(j-\tilde{I})}=0. \tag{4.6}$$

The uniqueness of the polynomial $g^{j,\tilde{I}}$ of the total degree (4.2) with the S_4 symmetry (2.6), the truncated weight space of zeros $\mathbb{W}^{b-\tilde{z}-\tilde{I}}_{q_{11}(j-\tilde{I})}(b',v-b',v)$, and the reduction formulas (4.5) may also be proved.

The extrapolation of the particular cases given by Eqs. (4.3) and (4.4) allows one to write the following expressions for polynomial $g^{j,\tilde{I}}$:

$$g^{j,\tilde{I}}(q_{\alpha\beta}) = \sum_{\mu\nu} \frac{M[(\tilde{I}-\tilde{z}+1)^{b-\tilde{z}-j}(\tilde{I}-\tilde{z})^{j-\tilde{I}}]}{M(\mu)M(\nu)}$$

$$\times \prod_{s=1}^{b-\tilde{z}-\tilde{I}} (b+a'-b''-v+\overline{\Delta}_s-s)^{(\tilde{I}-\tilde{z}+\overline{\Delta}_s-\mu_s)}(b-v+\overline{\Delta}_s-s)^{\tilde{I}-\tilde{z}+\overline{\Delta}_s-\nu_s)}(b'+v+1-s)^{(-1)\mu_s}$$

$$\times(a'+a''+b-v+3-s)^{(-1)\mu_s}(a'+b'-b+\tilde{z}+\tilde{I}+s+1)^{(\nu_s)}(a''-\tilde{z}+\tilde{I}+s+1)^{(\nu_s)} \tag{4.7a}$$

$$= \sum_{\mu^*\nu^*} \frac{M[(b-\tilde{z}-\tilde{I})^{\tilde{I}-\tilde{z}},b-\tilde{z}-j]}{M(\mu^*)M(\nu^*)}(v-\tilde{z}+j)^{(-1)(b-\tilde{z}-j-\mu^*_{t'})}(b''-a'+v+j-\tilde{z})^{(-1)(b-\tilde{z}-j-\nu^*_{t'})}$$

$$\times \prod_{s=1}^{t'-1}(v+s-1)^{(-1)(b-\tilde{z}-\tilde{I}-\mu^*_s)}(b''-a'+v+s-1)^{(-1)b-\tilde{z}-\tilde{I}-\nu^*_s)} \prod_{s=1}^{t'}(b'+v+s)^{(\mu^*_s)}$$

$$\times(a'+a''+b-v+2+s)^{(\mu^*_s)}(a'+b'-b+\tilde{z}+\tilde{I}-s+2)^{(-1)\nu^*_s}(a''+\tilde{I}-\tilde{z}-s+2)^{(-1)\nu^*_s}. \tag{4.7b}$$

Here $\overline{\Delta}_s = 1$ for $1 \leqslant s \leqslant b-\tilde{z}-j$ and 0 otherwise;

$$t' = \tilde{I}-\tilde{z}+1;$$

$$b-\tilde{z}-\tilde{I}-\mu^*_{t'+1-s} \geqslant \nu^*_s \geqslant b-\tilde{z}-\tilde{I}-\mu^*_{t'-s} \tag{4.8}$$

with $\Sigma_{s=1}^{t'+1}(b-\tilde{z}-\tilde{I}-\mu^*_s-\nu^*_{t'-s+1}) = j-\tilde{I}$ (each term is non-negative in the lhs). Partitions μ and ν are such that irrep $[(\tilde{I}-\tilde{z}+1)^{b-\tilde{z}-j}(\tilde{I}-\tilde{z})^{j-\tilde{I}}]$ appears (once) in the decomposition of the direct product of irreps $\mu\times\nu$.

Polynomial (4.7) may be also expressed in terms of function A_λ defined according Eq. (5.8) of Ref. 4.

V. CONCLUSION

In this paper, the explicit constructions of the orthonormal coupling coefficients of SU(3) are completed for the arbitrary multiplicity of the irreducible representations but for the most simple (paracanonical) labeling scheme which, unfortunately, is not invariant with respect to the contragrediency transformation of the tensor operators and states. This construction includes the following three steps: the expansion of the general isoscalar factors in terms of the boundary ones, the expression of the boundary orthonormal isofactors as a ratio of the numerator and denominator (under square root) polynomials (these two steps are solved or discussed in Paper I), and the explicit expression of the indecomposable numerator–denominator polynomials. (An alternative construction includes the definite bilinear combinations of isofactors and the inverse orthogonalization coefficients. In the different situations the direct or inverse approach may be more convenient.) Here the last problem is solved with the numerator–denominator functions that appear being related to the auxiliary functions $A_\lambda\left(\begin{smallmatrix}abde\\c\end{smallmatrix}\right)$ of Ref. 4, where partitions λ may be represented as rectangles (with possible deficient squares in the last row or column) measured by the current numbers of the multiplicity labels from the beginning and from the end. Contrary to the case of the canonical G-functions,[4,5] there is no problem associated with the multiplicity of the irrep λ in the decomposition of the direct product of irreps $\mu\times\nu$ of U(t). However, the new classes of sums over partitions μ and ν appearing here and in Refs. 4 and 5 still wait for their interpretation.

The construction of the orthogonalization coefficients for the canonical tensor operators of SU(3) (respectively, the boundary canonical isofactors depending also on eight parameters) may be the last step in the explicit construction of the canonical orthonormal isofactors of SU(3). In a subsequent paper we hope to present also the explicit orthogonalization coefficients for the Elliott–Draayer states of the noncanonical chains of subgroups with the one-dimensional multiplicity labels and the final overlap coefficients related to the Saalschutzian ${}_4F_3(1)$ series.

[1]S. Ališauskas, "On biorthogonal and orthonormal Clebsch–Gordan coefficients of SU(3): Analytical and algebraic approaches," J. Math. Phys. 29, 2351 (1988). Note that the symbol z should be corrected to r in the $9j$ coefficient of the rhs of Eq. (3.2a), the symbol $\tilde{\tau}_0'$ should be deleted in the explanation of Eq. (4.2), the symbols I″, − should appear (twice) as subscripts of $\eta_{-,I',-}$ on the top of p. 2363, and Ref. 15 is published in 1982.

[2]K. T Hecht, "SU_3 recoupling and fractional parentage in the $2s-1d$ shell," Nucl. Phys. 62, 1 (1965).

[3]L. C. Biedenharn, M. A. Lohe, and J. D. Louck, "On the denominator function for canonical SU(3) tensor operators," J. Math. Phys. 26, 1458 (1985).

[4]J. D. Louck, L. C. Biedenharn, and M. A. Lohe, "On the denominator function for canonical SU(3) tensor operators. II. Explicit polynomial form," J. Math. Phys. 29, 1106 (1988).

[5]L. C. Biedenharn, R. A. Gustafson, M. S. Lohe, J. D. Louck, and S. C. Milne, "Special functions and group theory in theoretical physics," in *Special Functions: Group Theoretical Aspects and Applications*, edited by R. A. Askey, T. H. Koornwinder, and W. Schempp (Reidel, Dordrecht, 1984), p. 129.

[6]Z. Pluhař, Yu. F. Smirnov, and V. N. Tolstoy, "Clebsch–Gordan coefficients of SU(3) with simple symmetry properties," J. Phys. A: Math. Gen. 19, 21 (1986).

[7]R. Le Blanc and D. J. Rowe, "A resolution of the SU(3) outer multiplicity problem and computation of Wigner coefficients for SU(3)," J. Phys. A: Math. Gen. 19, 19 (1986).

[8]The triple sums of the type $G_q^1(\Delta;x)$ and $G_q^q(qqq;x)$ of Ref. 4 depending on 11 parameters also appeared originally in special isofactors of Sp(4)⊃SU(2) ×SU(2) and SO(n)⊃SO(n') + SO(n''), see S. J. Ališauskas and A. P. Jucys, J. Math. Phys. 12, 594 (1971); E. Z. Norvaišas and S. J. Ališauskas, Liet. Fiz. Rink 14, 443 (1974); Sov. Phys.—Collection 14, (3) 17 (1974); S. J. Ališauskas, Fiz. Elem. Chastits At. Yadra 14, 1336 (1983); Sov. J. Part. Nucl. 14, 563 (1983).

[9]L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U.P., Cambridge, 1966).
[10]A. P. Jucys and A. A. Bandzaitis, *Theory of Angular Momentum as Quantum Mechanics* (Mokslas, Vilnius, 1977), 2nd ed. (in Russian).
[11]L. C. Biedenharn and J. D. Louck, "Some properties of generalized hypergeometric coefficients," Adv. Appl. Math. **9**, 447 (1988).
[12]L. C. Biedenharn, A. M. Bincer, M. A. Lohe, and J. D. Louck, "New relations and identities for generalized hypergeometric coefficients," Adv. Appl. Math. **10** (1990).
[13]The alternative form of A_λ functions (4.18) of Ref. 12 gives another class of expressions for $g_{\tilde{I}i}$ which structure is sufficiently simple only in $i = \tilde{I}$ or $t = 1$ cases.

The matrix representation of U_4 in the $U_2 \times U_2$ basis and some isoscalar factors for $U_{p+q} \supset U_p \times U_q$

Feng Pan
Department of Physics, Liaoning Normal University, Dalian 116022, People's Republic of China

(Received 12 September 1989; accepted for publication 20 December 1989)

Vector coherent state theory is applied to the noncanonical chain $U_{p+q} \downarrow U_p \times U_q$. Some matrix elements of U_4 generators in the $U_4 \supset U_2 \times U_2$ basis are derived by using K-matrix theory. Some transformation coefficients between $U_4 \supset U_2 \times U_2 \supset U_1 \times U_1$ and $U_4 \supset U_3 \supset U_2 \supset U_1$ basis vectors are obtained. Finally, analytical expressions of isoscalar factors for $U_{p+q} \supset U_p \times U_q$ for coupling $\{M_1\dot{0}\} \times \{M_2\dot{0}\}$ to $\{M'_1 M'_2 \dot{0}\}$ are derived by using these coefficients and isoscalar factors for $U_n \supset U_{n-1}$.

I. INTRODUCTION

The matrix representation of the canonical basis spanning finite-dimensional irreps of U_n has been thoroughly discussed in the study of many-body problems.[1-4] In many applications, however, it is necessary to choose a noncanonical basis for the unitary group U_{p+q} adapted to the subgroup $U_p \times U_q$. The usefulness of such a basis has been discussed by many authors.[5-7]

In Refs. 5 and 6, Klimyk *et al.* used the theory of the principal nonunitary series representations of an appropriately chosen semisimple Lie group to calculate matrix elements of U_{p+q} generators and some simple coupling coefficients. However, these matrix elements obtained in this way are nonunitarized. The unitarization can only be carried out for each individual case separately. In their series of papers only the irrep $\{M_1, \dot{0}, M_2\}$ of U_{p+q} with $M_1 \geqslant \dot{0} \geqslant M_2$, where $\dot{0} = (0,...,0)$, were considered.

Recently, a vector coherent state (VCS) theory and a simple K-matrix technique have been established by Deenen, Quesen, Rowe, and many others.[8-11] The combination of these two advances has proved to be a powerful tool in deriving matrix elements of generators of semisimple Lie groups from ladder representations of specific subgroups.

In this paper, we will briefly discuss the VCS representation of U_{p+q} in the $U_p \times U_q$ basis. As a simple example some reduced matrix elements of U_4 generators in the $U_4 \supset U_2 \times U_2$ basis are obtained. Then, we will derive some transformation coefficients between $U_4 \supset U_2 \times U_2 \supset U_1 \times U_1$ and $U_4 \supset U_3 \supset U_2 \supset U_1$ basis vectors. Finally, analytical expressions of isoscalar factors for $U_{p+q} \supset U_p \times U_q$ for coupling $\{M_1\dot{0}\} \times \{M_2\dot{0}\}$ to $\{M'_1 M'_2 \dot{0}\}$ will be derived by using these coefficients and isoscalar factors for $U_n \supset U_{n-1}$.

II. VCS REPRESENTATION OF U_{p+q} IN $U_p \times U_q$ BASIS

The u_{p+q} Lie algebra can be written abstractly as

$$\{E_{ij}; 1 \leqslant i, j \leqslant p+q\}, \tag{2.1}$$

with the commutation relation

$$[E_{ij}, E_{lk}] = E_{ik}\delta_{jl} - E_{lj}\delta_{ik}. \tag{2.2}$$

We decompose it into the $u_p \oplus u_q$ subalgebra

$$\{C^{p}_{\alpha\beta} = E_{\alpha\beta}; 1 \leqslant \alpha, \beta \leqslant p\} \tag{2.3a}$$

and

$$\{C^{q}_{\mu\nu} = E_{\mu\nu}; p+1 \leqslant \mu, \nu \leqslant p+q\}, \tag{2.3b}$$

with a set of raising (lowering) operators

$$\{A_{ij} = E_{ij}; 1 \leqslant i \leqslant p, p+1 \leqslant j \leqslant p+q\}, \tag{2.3c}$$

and a set of lowering (raising) operators

$$\{B_{ji} = E_{ji}; 1 \leqslant i \leqslant p, p+1 \leqslant j \leqslant p+q\}, \tag{2.3d}$$

with respect to the $u_p(u_q)$ subalgebra. Evidently, A_{ij} with $1 \leqslant i \leqslant p$ and $p+1 \leqslant j \leqslant p+q$ forms an Abelian algebra, and satisfies the condition

$$A^{\dagger}_{ij} = B_{ji}. \tag{2.4}$$

From Eq. (2.2) we obtain the following set of commutation relations:

$$\begin{aligned}
&[C^{i}_{\alpha\beta}, C^{i}_{\alpha'\beta'}] = \delta_{\beta\alpha'} C^{i}_{\alpha\beta'} - \delta_{\alpha\beta'} C^{i}_{\alpha'\beta}, \quad i = p \text{ or } q,\\
&[C^{p}_{\alpha\beta}, C^{q}_{\mu\nu}] = [A_{ij}, A_{i'j'}] = [B_{ji}, B_{j'i'}] = 0,\\
&[C^{p}_{\alpha\beta}, A_{ij}] = \delta_{\beta i} A_{\alpha j}, \quad [C^{p}_{\alpha\beta}, B_{ji}] = -\delta_{\alpha i} B_{j\beta},\\
&[C^{q}_{\mu\nu}, A_{ij}] = -\delta_{\mu j} A_{i\nu}, \quad [C^{q}_{\mu\nu}, B_{ji}] = \delta_{\nu j} B_{\mu i},\\
&[A_{ij}, B_{j'i'}] = \delta_{jj'} C^{p}_{ii'} - \delta_{ii'} C^{q}_{j'j}.
\end{aligned} \tag{2.5}$$

Given a generic irreducible ladder representation $\{M\}_{p+q}$, where $\{M\}_{p+q} = \{M_{1_{p+q}} M_{2_{p+q}} \cdots M_{p+q\,p+q}\}$, and satisfies the betweenness condition $M_{1_{p+q}} \geqslant M_{2_{p+q}} \geqslant \cdots \geqslant M_{p+q\,p+q}$, a u_p highest (u_q lowest) weight space is defined by all the states belonging to the maximal u_p (minimal u_q) subrepresentation $\{m\}_p$ ($\{m'\}_q$),

$$M_{ip+q} = m_{ip} \quad (1 \leqslant i \leqslant p), \tag{2.6a}$$

$$M_{p+j\,p+q} = m'_{jq} \quad (1 \leqslant j \leqslant q). \tag{2.6b}$$

Let $\{|\{m\}_p \{m'\}_q; \eta\rangle\}$ denotes an orthonormal basis for this representation. The u_{p+q} VCS wave function is defined by[11]

$$\Psi(z) = \sum_{\eta} |\{m\}_p \{m'\}_q; \eta\rangle \langle \{m\}_p \{m'\}_q; \eta| e^{z \cdot A} |\Psi\rangle, \tag{2.7}$$

where $|\Psi\rangle$ is any state of the unirrep $\{M\}_{p+q}$,

$$z \cdot A = z_{ji} A_{ij}, \tag{2.8}$$

where (and in the following equations) the repeated subscripts should be summed, and z_{ij} are $p \times q$ complex variables used as coordinates for the factor space $U_{p+q}/U_p \times U_q$.

Using the results given by Ref. 11, we readily obtain the

 0022-2488/90/061333-07$03.00

following VCS representation of the u_{p+q} algebra:

$$\Gamma(A_{ij})=\frac{\partial}{\partial z_{ji}},$$

$$\Gamma(C^{p}_{\alpha\beta})=\mathscr{E}^{(p)}_{\alpha\beta}-z_{j\beta}\frac{\partial}{\partial z_{j\alpha}},$$

$$\Gamma(C^{q}_{\mu\nu})=\mathscr{E}^{(q)}_{\mu\nu}+z_{\mu i}\frac{\partial}{\partial z_{\nu i}}, \tag{2.9}$$

$$\Gamma(B_{ji})=z_{ji'}\mathscr{E}^{(p)}_{i'i}-z_{j'i}\mathscr{E}^{(q)}_{jj'}-z_{ji'}z_{\mu i}\frac{\partial}{\partial z_{\mu i'}},$$

where $\mathscr{E}^{(p)}_{ij}(1\leqslant i,j\leqslant p)$ $(\mathscr{E}^{(q)}_{\mu\nu}(1+p\leqslant\mu,\nu\leqslant p+q))$ spans an intrinsic u_p (u_q) algebra only acting on the u_p highest (u_q lowest) weight space $|\{m\}_p\{m'\}_q;\eta\rangle$.

In order to construct an orthonormal Bargmann basis $|\{M\}_p\{M'\}_q;\eta\rangle$ that reduces the stability subalgebra $u_p\oplus u_q$, we need first to construct orthonormal polynomials $Z^{\{n\}}_{\mu\nu}(z)$ in (z_{ji}), which transforms as the components of an irreducible tensor $\{-n\}\equiv\{-n_\alpha,-n_{\alpha-1},\ldots,-n_1\}$ under u_p and $\{n\}\equiv\{n_1,n_2,\ldots,n_\alpha\}$ under u_q with $\alpha=\min(p,q)$. Then, the orthonormal Bargmann basis is defined by[11]

$$\langle z|\xi\{M\}_p\{M'\}_q;\eta\rangle = [Z^{\{n\}}(z)\times|\{m\}_p\{m'\}_q\rangle]^{\rho\{M\}_p\{M'\}_q}_\eta, \tag{2.10}$$

where $\xi\sim(\{n\}\rho)$ indicates the two kinds of $u_p\oplus u_q$ multiplicity that can arise in the decomposition $u_{p+q}\downarrow u_p\oplus u_q$.

The branching rule for $U_{p+q}\downarrow U_p\times U_q$ has been given by Ref. 12 using the tensor method; the more simplified form given in Ref. 7 can be stated as follows: Given a fixed irrep $\{\lambda\}$ of U_{p+q} and a fixed irrep $\{\mu\}$ of U_p contained entirely in $\{\lambda\}$, the subduction series are

$$\{\lambda\}\downarrow\{\mu\}\otimes(\{\lambda\}-\{\mu\})=\sum_{\{\nu\}}\Gamma_{\lambda\mu\nu}\{\mu\}\otimes\{\nu\}, \tag{2.11}$$

where the deleted portion $\{\lambda\}-\{\mu\}$, which is a "skew tableau" in the sense defined by Robinson,[13] forms a irrep for the subgroup U_q, and $\Gamma_{\lambda\mu\nu}$ is the multiplicity of occurrence of $\{\nu\}$ in $\{\lambda\}-\{\mu\}$.

III. APPLICATION TO THE $U_4\supset U_2\times U_2\supset U_1\times U_1$ CHAIN

As a simple nontrivial example, we will discuss the $U_4\supset U_2\times U_2\supset U_1\times U_1$ chain in this section. We set the $u_2\oplus u_2$ and $u_1\oplus u_1$ subalgebras as follows:

$$u_2^{(1)}:E_{ij}\quad(i,j=1,2),$$
$$u_2^{(2)}:E_{\mu\nu}\quad(\mu,\nu=3,4), \tag{3.1a}$$
$$u_1^{(1)}:E_{11},\quad u_1^{(2)}:E_{33}.$$

The remaining generators of u_4 can be put in the form of tensor operators:

$$T^{\{10\}\{0-1\}}_{\mu\ \ \nu}:\{E_{ij};i=1,2,j=3,4\}, \tag{3.1b}$$

$$T^{\{0-1\}\{10\}}_{\alpha\ \ \beta}:\{E_{ji};i=1,2,j=3,4\}. \tag{3.1c}$$

A. VCS representation

The $u_4\downarrow u_2\oplus u_2\downarrow u_1\oplus u_1$ VCS wave function can be written as

$$\Psi(z)=\sum_{mm'}\left|\begin{matrix}\{M_1M_2\} & \{M_3M_4\}\\ m & m'\end{matrix}\right\rangle \times\left\langle\begin{matrix}\{M_1M_2\} & \{M_3M_4\}\\ m & m'\end{matrix}\right|e^X|\Psi\rangle, \tag{3.2}$$

where

$$X=z_{31}E_{13}+z_{41}E_{14}+z_{32}E_{23}+z_{42}E_{24}$$
$$\equiv z_1E_{13}+z_2E_{14}+z_3E_{23}+z_4E_{24},$$

the state

$$\left|\begin{matrix}\{M_1M_2\} & \{M_3M_4\}\\ m & m'\end{matrix}\right\rangle\equiv\left|\begin{matrix} & \{M_1M_2M_3M_4\} & \\ \{M_1M_2\} & & \{M_3M_4\}\\ m & & m'\end{matrix}\right\rangle,$$

which satisfies

$$E_{ij}\left|\begin{matrix}\{M_1M_2\} & \{M_3M_4\}\\ m & m'\end{matrix}\right\rangle=0,\quad \text{for } i=1,2,\quad j=3,4. \tag{3.3}$$

The VCS representation of $u_4\downarrow u_2\oplus u_2$ can be obtained by using Eq. (2.9):

$$\Gamma(E_{31})=z_1(\mathscr{E}_{11}-\mathscr{E}_{33})+z_3\mathscr{E}_{21}-z_2\mathscr{E}_{34}-z_1z\cdot\partial+\begin{vmatrix}z_1 & z_2\\ z_3 & z_4\end{vmatrix}\partial_4,$$

$$\Gamma(E_{32})=z_1\mathscr{E}_{12}+z_3(\mathscr{E}_{22}-\mathscr{E}_{33})-z_4\mathscr{E}_{34}-z_3z\cdot\partial-\begin{vmatrix}z_1 & z_2\\ z_3 & z_4\end{vmatrix}\partial_2,$$

$$\Gamma(E_{41})=z_2(\mathscr{E}_{11}-\mathscr{E}_{44})-z_1\mathscr{E}_{43}+z_4\mathscr{E}_{21}-z_2z\cdot\partial-\begin{vmatrix}z_1 & z_2\\ z_3 & z_4\end{vmatrix}\partial_3, \tag{3.4a}$$

$$\Gamma(E_{42})=z_2\mathscr{E}_{12}-z_3\mathscr{E}_{43}+z_4(\mathscr{E}_{22}-\mathscr{E}_{44})-z_4z\cdot\partial+\begin{vmatrix}z_1 & z_2\\ z_3 & z_4\end{vmatrix}\partial_1,$$

$$\Gamma(E_{13})=\partial_1,\quad\Gamma(E_{23})=\partial_3,\quad\Gamma(E_{14})=\partial_2,$$
$$\Gamma(E_{24})=\partial_4,$$

and

$$\begin{aligned}
\Gamma(E_{11})&=\mathscr{E}_{11}-z_1\partial_1-z_2\partial_2=\mathscr{E}_{11}+\mathscr{E}^{\rm col}_{11},\\
\Gamma(E_{22})&=\mathscr{E}_{22}-z_3\partial_3-z_3\partial_4=\mathscr{E}_{22}+\mathscr{E}^{\rm col}_{22},\\
\Gamma(E_{21})&=\mathscr{E}_{21}-z_1\partial_3-z_2\partial_4=\mathscr{E}_{21}+\mathscr{E}^{\rm col}_{21},\\
\Gamma(E_{12})&=\mathscr{E}_{12}-z_3\partial_1-z_4\partial_2=\mathscr{E}_{12}+\mathscr{E}^{\rm col}_{12},\\
\Gamma(E_{33})&=\mathscr{E}_{33}+z_1\partial_1+z_3\partial_3=\mathscr{E}_{33}+\mathscr{E}^{\rm col}_{33},\\
\Gamma(E_{44})&=\mathscr{E}_{44}+z_2\partial_2+z_4\partial_4=\mathscr{E}_{44}+\mathscr{E}^{\rm col}_{44},\\
\Gamma(E_{34})&=\mathscr{E}_{34}+z_1\partial_2+z_3\partial_4=\mathscr{E}_{34}+\mathscr{E}^{\rm col}_{34},\\
\Gamma(E_{43})&=\mathscr{E}_{43}+z_2\partial_1+z_4\partial_3=\mathscr{E}_{43}+\mathscr{E}^{\rm col}_{43}.
\end{aligned} \tag{3.4b}$$

The orthonormal Bargmann basis can be written as

$$\langle z|\{n_1n_2\}\{M'_1M'_2\}\{M'_3M'_4\};mm'\rangle = [Z^{\{n_1n_2\}}(z)\times|\{M_1M_2\}\{M_3M_4\}\rangle]^{\{M'_1M'_2\}\{M'_3M'_4\}}_{mm'}. \quad (3.5)$$

The orthonormal polynomials

$$Z^{\{n_1n_2\}}_{-m_1m_2}(z)\equiv Z^{\{-n_2-n_1\}\{n_1n_2\}}_{-m_1m_2}(z)$$

can be constructed as follows:

$$Z^{\{n_1n_2\}}_{-m_1m_2}(z) = \mathcal{N}(\mathcal{C}^{\text{col}}_{12})^{n_1-m_1}z_1^{m_2-n_2}z_2^{n_1-m_2}\begin{vmatrix} z_1 & z_2\\ z_3 & z_4\end{vmatrix}^{n_2}, \quad (3.6)$$

where $\mathcal{N}$ is the normalization factor. Using the explicit form of $\mathcal{C}^{\text{col}}_{12}$ given by Eq. (3.4b), we finally obtain

$$Z^{\{n_1n_2\}}_{-m_1m_2}(z) = \left[\frac{(n_1-n_2+1)!(m_2-n_2)!(n_1-m_2)!(m_1-n_2)!}{(n_1+1)!n_2!(n_1-m_1)!(n_1-n_2)!}\right]^{1/2} \times\sum_k \binom{n_1-m_1}{k}\frac{z_1^{m_2-n_2+m_1-n_1+k}z_2^{n_1-m_2-k}z_3^{n_1-m_1-k}z_4^{k}}{(n_1-m_2-k)!(m_2-n_2+k)!}\begin{vmatrix} z_1 & z_2\\ z_3 & z_4\end{vmatrix}^{n_2}. \quad (3.7)$$

It can easily be proved that

$$\begin{pmatrix} \sum_{i,j=1}^{2}\mathcal{C}^{\text{col}}_{ij}\mathcal{C}^{\text{col}}_{ji}\\ \sum_{i,j=3}^{4}\mathcal{C}^{\text{col}}_{ij}\mathcal{C}^{\text{col}}_{ji}\\ \mathcal{C}^{\text{col}}_{11}\\ \mathcal{C}^{\text{col}}_{33}\end{pmatrix} Z^{\{n_1n_2\}}_{-m_1m_2}(z) = \begin{pmatrix} n_1(n_1+1)+n_2(n_2-1)\\ n_1(n_1+1)+n_2(n_2-1)\\ -m_1\\ m_2\end{pmatrix} Z^{\{n_1n_2\}}_{-m_1m_2}(z). \quad (3.8)$$

B. The matrix representations

In order to obtain the matrix representations of this chain, we use the simple K-matrix theory summarized in Ref. 11. It is the transformation K that maps the VCS representations of $u_4\downarrow u_2\oplus u_2$ to an equivalent representation $\gamma=K^{-1}\Gamma K$ that is unitary with respect to the Bargmann inner product. Because the set of operators (A_{ij}) forms an Abelian algebra, we can set $K=K^\dagger$. Thus, the K^2 matrix elements can be obtained by using the simple equation of Rowe,[11]

$$K^2z_i = [\hat{\Lambda},z_i]K^2, \quad (3.9)$$

where $\hat{\Lambda}$ is the combination of Casimir invariants

$$\hat{\Lambda} = -\sum_{i,j=3}^{4}\mathcal{C}_{ij}\mathcal{C}^{\text{col}}_{ji} - \sum_{i,j=1}^{2}\mathcal{C}_{ij}\mathcal{C}^{\text{col}}_{ji} - \tfrac{1}{2}z\cdot\partial z\cdot\partial + \tfrac{1}{2}z\cdot\partial + \begin{vmatrix} z_1 & z_2\\ z_3 & z_4\end{vmatrix}\begin{vmatrix} \partial_1 & \partial_2\\ \partial_3 & \partial_4\end{vmatrix}, \quad (3.10a)$$

with the eigenvalue

$$\Lambda = \tfrac{1}{2}[M_1(M_1+1)+M_2(M_2-1)+M_3(M_3+1)+M_4(M_4-1)] - \tfrac{1}{2}[M'_1(M'_1+1)+M'_2(M'_2-1)+M'_3(M'_3+1)+M'_4(M'_4-1)] + n_1(n_1+1)+n_2(n_2-1)+n_2(n_1+1)-\tfrac{1}{2}(n_1+n_2)(n_1+n_2-1). \quad (3.10b)$$

When $M_1=M_2$ or $M_3=M_4$, the reduction $U_4\downarrow U_2\times U_2$ is multiplicity-free. In this case the $K^2(\{M'_1M'_2\}\{M'_3M'_4\})$ submatrix is one-dimensional and can be obtained recursively by using Eq. (3.22) of Ref. 11:

$$K^2(\{M'_1M'_2\}\{M'_3M'_4\}) = \frac{(M_1-M+1)!(M_2-M)!}{(M'_1-M+1)!(M'_2-M)!},\quad \text{for } M_3=M_4=M, \quad (3.11a)$$

$$K^2(\{M'_1M'_2\}\{M'_3M'_4\}) = \frac{(M-M_3)!(M-M_4+1)!}{(M-M'_3)!(M-M'_4+1)!},\quad \text{for } M_1=M_2=M. \quad (3.11b)$$

For the generic irrep $\{M_1M_2M_3M_4\}$ of U_4, the nonmultiplicity-free case will occur. By solving the recursion relation of Eq. (3.25) given by Ref. 11 with the initial value

$$K^2_{\{0\}\{0\}}(\{M_1M_2\}\{M_3M_4\}) = 1, \quad (3.12a)$$

We can successively obtain the K^2 submatrices

$$K^2_{\{10\}\{10\}}(\{M_1-1M_2\}\{M_3+1M_4\}) = M_1-M_3+1,$$
$$K^2_{\{10\}\{10\}}(\{M_1-1M_2\}\{M_3M_4+1\}) = M_1-M_4+2,$$
$$K^2_{\{10\}\{10\}}(\{M_1M_2-1\}\{M_3+1M_4\}) = M_2-M_3,$$
$$K^2_{\{10\}\{10\}}(\{M_1M_2-1\}\{M_3M_4+1\}) = M_2-M_4+1, \quad (3.12b)$$

and

$$K^2_{\{11\}\{11\}}(\{M_1-1M_2-1\}\{M_3+1M_4+1\}$$
$$=\tfrac14[(M_2-M_4)(M_1-M_2)(M_3-M_4+2)(M_1-M_3+1)$$
$$+(M_1-M_3)(M_1-M_2+2)(M_3-M_4)(M_2-M_4+1)$$
$$+(M_1-M_2)(M_3-M_4)(M_1-M_4+2)(M_2-M_3-1)$$
$$+(M_1-M_2+2)(M_3-M_4+2)(M_2-M_3)(M_1-M_4+1)]/(M_3-M_4+1)(M_1-M_2+1).$$
$$K^2_{\{20\}\{20\}}(\{M_1-1M_2-1\}\{M_3+1M_4+1\})$$
$$=\tfrac14[(M_1-M_2)(M_3-M_4)(M_1-M_4+3)(M_2-M_3)+(M_1-M_2)(M_3-M_4+2)(M_1-M_3+2)$$
$$\times(M_2-M_4+1)+(M_1-M_2+2)(M_3-M_4)(M_2-M_4+2)(M_1-M_3+1)+(M_1-M_2+2)$$
$$\times(M_3-M_4+2)(M_2-M_3+1)(M_1-M_4+2)]/(M_3-M_4+1)(M_1-M_2+1),$$
$$K^2_{\{20\}\{11\}}(\{M_1-1M_2-1\}\{M_3+1M_4+1\})=K^2_{\{11\}\{20\}}(\{M_1-1M_2-1\}\{M_3+1M_4+1\})$$
$$=-(M_1-M_2+M_3-M_4)[(M_1-M_2+2)(M_3-M_4+2)$$
$$\times(M_1-M_2)(M_3-M_4)]^{1/2}/2(M_1-M_2+1)(M_3-M_4+1). \tag{3.12c}$$

The results will be more complicated with the increase of n_1 and n_2.

In the following we will concentrate on multiplicity-free cases. The reduced matrix elements of E_{ji} with $j=3,4$ and $i=1,2$ can be derived by using the following equations[11]:

$$\left\langle \begin{matrix}\{M_1M_2MM\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|E_{ji}\left\|\begin{matrix}\{M_1M_2MM\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle$$
$$=\frac{K(\{M_1''M_2''\}\{M_3''M_4''\})}{K(\{M_1'M_2'\}\{M_3'M_4'\})}$$
$$\times\left\langle \begin{matrix}\{M_1M_2MM\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|z\left\|\begin{matrix}\{M_1M_2MM\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle, \tag{3.13a}$$

$$\left\langle \begin{matrix}\{MMM_3M_4\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|E_{ji}\left\|\begin{matrix}\{MMM_3M_4\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle$$
$$=\frac{K(\{M_1''M_2''\}\{M_3''M_4''\})}{K(\{M_1'M_2'\}\{M_3'M_4'\})}$$
$$\times\left\langle \begin{matrix}\{MMM_3M_4\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|z\left\|\begin{matrix}\{MMM_3M_4\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle, \tag{3.13b}$$

and

$$\left\langle \begin{matrix}\{M_1M_2MM\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|z\left\|\begin{matrix}\{M_1M_2MM\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle$$
$$=\langle\{n_1'n_2'\}\|z\|\{n_1n_2\}\rangle\, U(\tfrac12,(M_1-M_2),\tfrac12(n_1-n_2),$$
$$\times\tfrac12(M_1''-M_2''),\tfrac12;\tfrac12(M_1'-M_2'),$$
$$\times\tfrac12(n_1'-n_2'))U(0,\tfrac12(n_1-n_2),$$
$$\times\tfrac12(M_3''-M_4''),\tfrac12;\tfrac12(M_3'-M_4'),\tfrac12(n_1'-n_2')), \tag{3.14a}$$

$$\left\langle \begin{matrix}\{MMM_3M_4\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|z\left\|\begin{matrix}\{MMM_3M_4\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle$$
$$=\langle\{n_1'n_2'\}\|z\|\{n_1n_2\}\rangle\, U(0,\tfrac12(n_1-n_2),\tfrac12(M_1''M_2''),\tfrac12;$$
$$\tfrac12(M_1'-M_2'),\tfrac12(n_1'-n_2'))$$
$$\times U(\tfrac12(M_3-M_4),\tfrac12(n_1-n_2),\tfrac12(M_3''-M_4''),\tfrac12;\tfrac12(M_3'$$
$$-M_4'),\tfrac12(n_1'-n_2')), \tag{3.14b}$$

where the U-coefficient is a Racah W-coefficient in unitary form, and[14]

$$\langle\{n_1'n_2'\}\|z\|\{n_1n_2\}\rangle$$
$$=\delta_{n_1+1n_1'}\delta_{n_2'n_2}\left[\frac{(n_1+2)(n_1-n_2+1)}{(n_1-n_2+2)}\right]^{1/2}$$
$$+\delta_{n_1n_1'}\delta_{n_2+1n_2'}\left[\frac{(n_2+1)(n_1-n_2+1)}{(n_1-n_2)}\right]^{1/2}. \tag{3.15}$$

The reduced matrix elements of E_{ij} with $i=1,2$ and $j=3,4$ follow from Hermitian conjugation

$$\left\langle \begin{matrix}\{M_1M_2M_3M_4\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\|E_{ij}\left\|\begin{matrix}\{M_1M_2M_3M_4\}\\ \{M_1''M_2''\}\{M_3''\}\{M_4''\}\end{matrix}\right\rangle$$
$$=(-)^{1+1/2(M_1'-M_2'+M_2''-M_1''+M_3'-M_4'+M_4''-M_3'')}\left[\frac{(M_1''-M_2''+1)(M_3''-M_4''+1)}{(M_1'-M_2'+1)(M_3'-M_4'+1)}\right]^{1/2}$$
$$\times\left\langle \begin{matrix}\{M_1M_2M_3M_4\}\\ \{M_1''M_2''\}\{M_3''M_4''\}\end{matrix}\right\|E_{ij}\left\|\begin{matrix}\{M_1M_2M_3M_4\}\\ \{M_1'M_2'\}\{M_3'M_4'\}\end{matrix}\right\rangle. \tag{3.16}$$

The Eq. (3.14) can be written more explicitly by using the results given by Eq. (3.15):

$$\left\langle \begin{matrix} \{M_1M_2MM\} \\ \{M_1''M_2''\}\{M_3''M_4''\} \end{matrix} \middle\| z \middle\| \begin{matrix} \{M_1M_2MM\} \\ \{M_1'M_2'\}\{M_3'M_4'\} \end{matrix} \right\rangle$$

$$= \begin{cases} \left[\dfrac{(M_1-M_2+M_1'-M_2'-n_1+n_2)(M_1-M_2-M_1'+M_2'+n_1-n_2+2)(n_1+2)}{4(M_1'-M_2'(n_1-n_2+2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'-1M_2'\},\quad \{M_3''M_4''\}=\{M_3'+1M_4'\}, \\ \left[\dfrac{(M_1'-M_2'+n_1-n_2+M_1-M_2+4)(M_1'-M_2'+n_1-n_2-M_1+M_2+2)(n_1+2)}{4(M_1'-M_2'+2)(n_1-n_2+2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'M_2'-1\},\quad \{M_3''M_4''\}=\{M_3'+1M_4'\}, \\ \left[\dfrac{(M_1'-M_2'+n_1-n_2+M_1-M_2+2)(M_1'-M_2'+n_1-n_2-M_1+M_2)(n_2+1)}{4(M_1'-M_2')(n_1-n_2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'-1M_2'\},\quad \{M_3''M_4''\}=\{M_3'M_4+1\}, \\ -\left[\dfrac{(M_1-M_2-M_1'+M_2'+n_1-n_2)(M_1-M_2+M_1'-M_2'-n_1+n_2+2)(n_2+1)}{4(M_1'-M_2'+2)(n_1-n_2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'M_2'-1\},\quad \{M_3''M_4''\}=\{M_3'M_4'+1\} \end{cases} \tag{3.17a}$$

$$\left\langle \begin{matrix} \{MMM_3M_4\} \\ \{M_1''M_2''\}\{M_3''M_4''\} \end{matrix} \middle\| z \middle\| \begin{matrix} \{MMM_3M_4\} \\ \{M_1'M_2'\}\{M_3'M_4'\} \end{matrix} \right\rangle$$

$$= \begin{cases} -\left[\dfrac{(M_3-M_4-n_1+n_2+M_3'-M_4'+2)(M_3-M_4+n_1-n_2-M_3'+M_4')(n_2+1)}{4(M_3'-M_4'+2)(n_1-n_2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'-1M_2'\},\quad \{M_3''M_4''\}=\{M_3'+1M_4'\}, \\ \left[\dfrac{(M_3-M_4+n_1-n_2-M_3'+M_4'+2)(M_3-M_4-n_1+n_2+M_3'-M_4')(n_1+2)}{4(M_3'-M_4')(n_1-n_2+2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'M_2'-1\},\quad \{M_3''M_4''\}=\{M_3'M_4'\}+1\}, \\ \left[\dfrac{(M_3'-M_4'+M_3-M_4+n_1-n_2+2)(n_1-n_2+M_3'-M_4'-M_3+M_4)(n_2+1)}{4(n_1-n_2)(M_3'-M_4')}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'-1M_2'\},\quad \{M_3''M_4''\}=\{M_3'M_4'+1\}, \\ -\left[\dfrac{(n_1-n_2+M_3'-M_4'+M_3-M_4+4)(n_1-n_2+M_3'-M_4'-M_3+M_4+2)(n_1+2)}{4(n_1-n_2+2)(M_3'-M_4'+2)}\right]^{1/2}, \\ \quad \text{if } \{M_1''M_2''\}=\{M_1'M_2'-1\},\quad \{M_3''M_4''\}=\{M_3'+1M_4'\}, \end{cases} \tag{3.17b}$$

where $\{M_3'M_4'\}=\{n_1n_2\}$ in Eq. (3.17a) and $\{M_1'M_2'\}=\{n_1n_2\}$ in Eq. (3.17b).

IV. SOME ISOSCALAR FACTORS FOR $U_{p+q} \supset U_p \times U_q$

For two rowed irreps $\{M_1M_2 00\}$ of U_4, the reduction for $U_4 \downarrow U_2 \times U_2$ is multiplicity-free. The basis vectors of $U_4 \supset U_2 \times U_2 \supset U_1 \times U_1$ can be expanded in terms of $U_4 \supset U_3 \supset U_2 \supset U_1$ canonical basis vectors with one of the two sets of the $U_2 \supset U_1$ labels fixed:

$$\left| \begin{matrix} \{M_1M_2 00\} \\ \{M_1-rM_2-s-t\}\{r+ts\} \\ m \quad m'=r+t \end{matrix} \right\rangle = \sum_k A_k(rst,M_1M_2;m'=r+t) \left| \begin{matrix} \{M_1M_2 00\} \\ M_1-s+kM_2-k\,0\} \\ M_1-rM_2-s-t\} \\ m \end{matrix} \right\rangle, \tag{4.1}$$

where $A_k(rst,M_1M_2;m')$ is the transformation coefficient; the U_3 labels can be determined by acting E_{33} and E_{44} on Eq. (4.1). Furthermore, Acting E_{34} on Eq. (4.1), we have

$$A_k(rst,M_1M_2;r+t) = (-)^k \mathscr{N}\left[\binom{s}{k}\frac{(\alpha-s+2k+1)(\alpha-s+k)!}{(\alpha-r+k)!(\alpha+k+1)!}(r-s+k)!(\alpha+k+t+1)!(s-k+t)!\right]^{1/2}, \tag{4.2}$$

where $\alpha = M_1 - M_2$, and $\mathscr{N}$ is a normalization factor. Using the identity

$$\sum_{k=0}^{s}\binom{s}{k}\frac{(\alpha-s+2k+1)(\alpha-s+k)!(r-s+k)!(\alpha+k+t+1)!(s-k+t)!}{(\alpha-r+k)!(\alpha+k+1)!} = \frac{(r-s)!(\alpha+t+1)!(r+t+1)!t!}{(\alpha-r+s)!(r+t-s+1)!}, \tag{4.3}$$

we obtain the special transformation coefficients with $m' = r+t$.

$$A_k(rst,M_1M_2;r+t) = (-)^{k+\Delta(rst;M_1M_2)} \times\left[\binom{s}{k}\frac{(\alpha-s+2k+1)(\alpha-s+k)!(r-s+k)!(\alpha+k+t+1)!(s-k+t)!(\alpha-r+s)!(r+t-s+1)!}{(\alpha-r+k)!(\alpha+k+1)!(r-s)!(\alpha+t+1)!(r+t+1)!t!}\right]^{1/2}, \tag{4.4}$$

where $\Delta(rst;M_1M_2)$ is an appropriate phase factor. Acting E_{32} on Eq. (4.1), we get

$$A_k(r+1st,M_1M_2;r+t+1) = -A_k(rst,M_1M_2;r+t)\left[\frac{(r-s+k+1)(r+t-s+2)(\alpha-r+k)}{(r+1-s)(r+t+2)(\alpha-r+s)}\right]^{1/2}. \tag{4.5}$$

Similarly, acting E_{42} on Eq. (4.1), we obtain

$$A_k(rs+1t,M_1M_2;r+t) = -A_k(rst,M_1M_2;r+t)\left[\frac{(s+1)(\alpha-s+2k)(s-k+t+1)(\alpha-r+s+1)(r-s+1)}{(\alpha-s+2k+1)(\alpha-s+k)(r-s+k)(r+t-s)(s-k+1)}\right]^{1/2}, \tag{4.6a}$$

$$A_k(rst+1,M_1M_2;r+t+1) = A_k(rst,M_1M_2;r+t)\left[\frac{(\alpha+k+t+2)(s-k+t+1)(r+t-s+2)}{(\alpha+t+2)(r+t+2)(t+1)}\right]^{1/2}, \tag{4.6b}$$

where the results obtained from Eqs. (3.13)–(3.17) are used in deriving Eqs. (4.5) and (4.6). Thus, the overall phase factor $\Delta(rst,M_1M_2;r+t)$ can be chosen as $(-)^{s-t}$.

Next, coupling two symmetric basis vectors given by Eq. (4.1), and using the Racah factorization lemma, we have

$$\left[\begin{array}{c|cc} U_4 & \{M_10\} & \{M_20\} \\ U_2\times U_2 & \{M_1-p_10\}\{p_10\} & \{M_2-p_20\}\{p_20\}\end{array}\right]\left[\begin{array}{cc|c} p_1 & p_2 & \{m_3m_4\} \\ q_1 & q_2 & m_3\end{array}\right]A_0(p_100,M_10;q_1)A_0(p_200,M_20;q_2)$$
$$=\sum_k A_k(m_1m_2m_3m_4,M_1'M_2';m_3)\left[\begin{array}{cc} M_1 & M_2 \\ M_1-p_1+q_1 & M_2-p_2+q_2\end{array}\right]\left|\begin{array}{c}\{M_1'M_2'\} \\ \{m_1+m_2+m_3-M_2'+kM_2'-k\}\end{array}\right.$$
$$\times\left[\begin{array}{cc|c} M_1-p_1+q_1 & M_2-p_2+q_2 & \{m_1+m_2+m_3-M_2'+kM_2'-k\} \\ M_1-p_1 & M_2-p_2 & \{m_1m_2\}\end{array}\right], \tag{4.7}$$

where

$$A_k(m_1m_2m_3m_4,M_1'M_2';m_3) = \left\langle\begin{array}{c}\{M_1'M_2'00\} \\ \{m_1m_2\}\{m_3m_4\} \\ m\quad m_3\end{array}\right|\left.\begin{array}{c}\{M_1'M_2'00\} \\ \{m_1+m_2+m_3-M_2'+kM_2'k\} \\ \{m_1m_2\} \\ m\end{array}\right\rangle, \tag{4.8}$$

and

$$\left[\begin{array}{cc|c} p_1 & p_2 & \{m_3m_4\} \\ q_1 & q_2 & m_3\end{array}\right], \text{etc.}$$

are isoscalar factors for $U_n \supset U_{n-1}$.

According to the above phase choice, we have

$$A_0(r00,M0;q) = (-1)^r. \tag{4.9}$$

Because the isoscalar factors for $U_{p+q} \supset U_p\times U_q$ and $U_n \supset U_{n-1}$ are p, q, and n independent, using the analytical continuation and the known isoscalar factors for $U_n \supset U_{n-1}$,[15] from Eq. (4.7) we obtain the isoscalar factors for $U_{p+q} \supset U_p \times U_q$ of the following type:

$$\begin{aligned}
&\left[\begin{array}{c|cc|c} U_{p+q} & \{M_1\dot{0}\} & \{M_2\dot{0}\} & \{M_1'M_2'\dot{0}\} \\ U_p\times U_q & \{M_1-p_1\dot{0}\}\{p_1\dot{0}\} & \{M_2-p_2\dot{0}\}\{p_2\dot{0}\} & \{m_1m_2\dot{0}\}\{m_3m_4\dot{0}\} \end{array}\right] \\
&= (-)^{M_1'-m_1-m_4+p_2}\, 2^{m_3-m_4}(m_3-m_4+1)!\left[\frac{(2M_1'-2M_2'+2)(M_1-M_2')!(m_1-M_1+p_1)!}{(p_1+p_2+m_3-m_4+2)!!(m_3-m_4+p_1-p_2)!!}\right. \\
&\quad\times\left.\frac{((m_1+m_3-M_1')!(M_1'-m_1-m_4)!)^{-1}(p_1+p_2-m_3-m_4)!!(m_1-M_2'+m_4)!}{(m_3-m_4+p_2-p_1)!!(M_1'-M_1)!(M_1-p_1-m_2)!(m_1+m_3-M_2'+1)!(m_3+1)!}\right]^{1/2} \\
&\quad\times\sum_{kxyz}(-)^{k+x+y+z}\frac{z!(m_3-z)!(M_1'-M_1+p_1-z-x)!(M_1-M_2'-p_1+k+z+x)!}{x!y!(p_1-z-x)!(z-y)!(m_1+m_2+m_3-M_1-M_2'+p_1+k-z-x)!} \\
&\quad\times\frac{(m_1+m_2+m_3-M_1-M_2'+p_1+k-y)!(M_1-p_1-m_2+y)!}{(M_1-p_1-M_2'+z+x)!(m_1-M_1+p_1-y)!(M_1-p_1-M_2'+k+y)!(M_1'-M_2'+k+1)!} \\
&\quad\times\left[\binom{m_4}{k}\frac{(m_1+m_2+m_3-2M_2'-2k+1)(M_1'-M_2'-m_4+2k+1)(M_1'-M_2'-m_4+k)!}{k!(m_2+m_3-M_2'+k)!(M_1-M_2'+k-p_1+z)!(M_1'-m_1-m_2-m_3+M_2'-k)!}\right. \\
&\quad\times\left.(M_1-p_1-M_2'+z+k)!(M_1'-m_1-m_4+k)!(m_1+m_2+m_3-2M_2'+k)!\right]^{1/2} \\
&\quad\times\delta_{M_1+M_2,M_1'+M_2'}\,\delta_{M_1+M_2-p_1-p_2,m_1+m_2}\,\delta_{p_1+p_2,m_3+m_4}. \qquad (4.10)
\end{aligned}$$

V. CONCLUDING REMARKS

Using the VCS theory and K-matrix technique, we obtain some matrix elements for $U_4 \supset U_2 \times U_2$ when this reduction is multiplicity-free. However, for the general $U_{p+q} \supset U_p \times U_q$ basis, the results will be very complicated. Firstly, the complication will arise in the construction of orthonormal polynomials $Z_{\mu}^{\{-n\}\{n\}}{}_{\nu}(z)$. Another difficulty will occur when the reduction $U_{p+q} \downarrow U_p \times U_q$ is nonmultiplicity-free. In this case, the analytical forms of matrix elements of raising (lowering) operators cannot be achieved; and these matrix elements can only be calculated numerically.

[1]E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 463 (1963).
[2]M. Moshinsky, *Group Theory and the Many-body Problem* (Gordon and Breach, London, 1968).
[3]J. D. Louck, Am. J. Phys. **38**, 3 (1970).
[4]M. Ciftan, J. Math. Phys. **10**, 1635 (1969).
[5]A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 1995, 2011 (1979); **23**, 1399 (1982).
[6]A. U. Klimyk and I. I. Kacharik, J. Math. Phys. **29**, 2377 (1988).
[7]C. R. Sarma, J. Math. Phys. **23**, 1235 (1982).
[8]J. Deenen and C. Quesen, J. Math. Phys. **25**, 1638, 2354 (1984); **26**, 2705 (1985); C. Quesen, *ibid.* **27**, 478, 869 (1986).
[9]D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).
[10]D. J. Rowe, G. Rosensteel, and R. Gilmore, J. Math. Phys. **26**, 2787 (1985).
[11]D. J. Rowe, R. Le Blanc, and K. T. Hecht, J. Math. Phys. **29**, 287 (1988).
[12]R. C. King, J. Phys. A: Math. Gen. **8**, 429 (1975).
[13]G. de B. Robinson, *Representation Theory of the Symmetric Group* (University of Toronto, Toronto, 1961).
[14]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968).
[15]E. Chacon, M. Ciftan, and L. C. Biedenharn, J. Math. Phys. **13**, 577 (1972).

Deformations of some infinite-dimensional Lie algebras

Alice Fialowski
Department of Mathematics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6395

(Received 21 February 1989; accepted for publication 17 January 1990)

The concept of a versal deformation of a Lie algebra is investigated and obstructions to extending an infinitesimal deformation to a higher-order one are described. The rigidity of the Witt algebra and the Virasoro algebra is deduced from cohomology computations for certain Lie algebras of vector fields on the real line. The Lie algebra of vector fields on the line that vanish at the origin also turns out to be rigid. All the affine Lie algebras are rigid; this is derived from the cohomology of their maximal nilpotent subalgebra. On the other hand, the maximal nilpotent subalgebras in both the Virasoro and affine cases are not rigid and have interesting nontrivial deformations (in fact, most vector field Lie algebras are not rigid).

I. INTRODUCTION

It is known that in characteristic zero a semisimple Lie algebra has no nontrivial deformations.[1] The same is true for infinite-dimensional classical Lie algebras from the "Cartan series." We say that those Lie algebras are rigid. However, in nonzero characteristics, both semisimple and Cartan type Lie algebras have nontrivial deformations, as was shown by Rudakov (1971)[2] and Dzumadil'daev (1980).[3] For some other modular Lie algebras there are also interesting results.[4] It is known that solvable and nilpotent Lie algebras have very many nontrivial deformations. All infinitesimal deformations (defined later) of the maximal nilpotent subalgebra of simple finite-dimensional Lie algebras are known (Leger–Luks[5]), but the infinitesimal deformations for some other of their subalgebras are impossible to classify (Piper, 1971).[6]

In infinite dimension, we consider Lie algebras having a triangular decomposition:[7] affine algebras, Virasoro algebra, and their nilpotent subalgebras.

In this paper I am going to present my results on the deformations of these infinite-dimensional Lie algebras.

II. PRELIMINARIES

For computing deformations, we need cohomology theory. We should mention that although a highly developed general theory existed, there were very few computations. The situation changed in the late 1960s, after the important Russian works of Gel'fand and Fuks on the cohomology with trivial coefficients of Lie algebras of vector fields on a smooth manifold.[8] In 1976 the cohomology of the maximal nilpotent subalgebras of affine algebras with trivial coefficients was also computed (Garland–Lepowsky[9]).

(i) For computing deformations, we have to compute cohomology with coefficients in the adjoint representation.

Let L be a Lie algebra (finite or infinite dimensional). Let us define a Lie superalgebra structure on the cochain complex $C^{\cdot}(L;L)$. For $a \in C^p(L;L)$, $b \in C^q(L;L)$, define $ab \in C^{p+q-1}(L;L)$ by

$$ab(g_1,\ldots,g_{p+q-1})$$

$$:= \sum_{\sigma} \operatorname{sgn}(\sigma) a(b(g_{i_1},\ldots,g_{i_q}),g_{j_1},\ldots,g_{j_{p-1}}),$$

where σ runs over all the shuffle permutations with $i_1 < \cdots < i_q$ and $j_1 < \cdots < j_{p-1}$. Put

$$[a,b] := ab - (-1)^{(p-1)(q-1)} ba.$$

The differential of degree 1 acts on brackets by the rule:

$$d([a,b]) = [da,b] - (-1)^{p-1} [a,db].$$

It is easy to verify that the cochain complex $C^{\cdot}(L;L)$ is a differential Lie superalgebra. The superbracket multiplication can be lifted to the cohomology space:

$$H^p(L;L) \otimes H^q(L;L) \to H^{p+q-1}(L;L)$$

(see Ref. 10 for details).

(ii) Recall the intuitive definition of a deformation of a Lie algebra L_0. It is a family of Lie algebras L_t with the same underlying vector space, and with the bracket

$$\mu_t(x,y) = \mu_0(x,y) + \varphi(t)(x,y)$$
$$= [x,y] + \varphi_1 t + \varphi_2 t^2 + \cdots,$$

where $x,y \in L_0$, $\varphi(t) = \sum_{i=1}^{\infty} t^i \varphi_i$, and $\mu_0(x,y)$ is the original bracket in L_0. Obviously $\varphi_i \in C^2(L_0;L_0)$ and the Jacobi identity means

$$-d\varphi = \tfrac{1}{2}[\varphi,\varphi],$$

or for each k,

$$-2\sum_k d\varphi_k = \sum_k \sum_{i+j=k} [\varphi_i,\varphi_j] \mod(t^{k+1}),$$

where d and [,] where defined in (i). (This is the so-called "deformation equation.")

A *deformation* is said to be *of order k* if the Jacobi identities are satisified mod (t^{k+1}). A deformation of order 1 is called an *infinitesimal deformation*.

III. VERSAL DEFORMATIONS

First we give a general definition of Lie algebra deformations. There are four steps in the generalization (for details on this see Refs. 10,11).

(i) Consider L_t as a Lie algebra over $K((t))$.

(ii) Generalize over $K[[t_1,\ldots,t_r]]$.

(iii) Let the parameter space A be a local finite-dimensional algebra. We say that L_A is a deformation of the Lie algebra L, parameterized by a local finite dimensional algebra A if L_A is a Lie algebra structure over A on $L \otimes_K A$ such that the Lie algebra structure on $L = L_A \otimes_A K = (L \otimes A) \otimes_A K$ is the given one on L.

 0022-2488/90/061340-04$03.00

(iv) Let the parameter space be a complete local algebra A ($A = \lim A/m_A^n$, where A/m_A^n are local finite dimensional for each n). A *deformation* of L parametrized by a complete local algebra A is a projective limit of deformations of L, parametrized by A/m_A^n. Two deformations, L_A and L'_A, parametrized by A are called *equivalent* if there exists a Lie algebra isomorphism over A of L_A on L'_A, inducing the identity of $L_A \otimes_A K = L$ on $L'_A \otimes_A K = L$.

Define a functor from the category $\widehat{C}$ of complete local algebras into the equivalence classes of deformations of L, parametrized by A.

The generalizations (i)–(iv) are necessary in order to define the so-called versal deformations, which induce all the other deformations of a given Lie algebra.

Definition: A deformation L_R of L parametrized by $R \in \widehat{C}$ is a *versal deformation* if for any L_A, parametrized by $A \in \widehat{C}$, there exists a morphism $f: R \to A$ such that (i) $L_R \otimes_R A$ is equivalent to L_A; (ii) if the map $m_R/m_R^2 \to m_A/m_A^2$ induced by f is unique.

Theorem 1.1: If $H^2(L;L)$ is finite dimensional then there exists a versal deformation.

Proof: The statement follows from a general theorem of Schlessinger.[12] See the details in Ref. 10.

Remark: Suppose $A = \mathbb{C}[t_1,...,t_n]/I$. Then the deformation equation is the following (see Ref. 11):

$$2 \sum_{|\alpha| \geq 1} (d\varphi_\alpha) t^\alpha + \sum_{|\alpha| \geq 1,} \sum_{\beta+\gamma=\alpha} [\varphi_\beta, \varphi_\gamma] t^\beta t^\gamma \equiv 0 \quad (\text{mod } I).$$

For $|\alpha| = 1$ we get $d\varphi_\alpha = 0$ for each α, which means that φ_α has to be a cocycle.

Proposition: The elements of $H^2(L;L)$ correspond bijectively to the nonequivalent infinitesimal deformations.

Proof: This well-known fact can be proved by direct computation.

Corollary: The condition $H^2(L;L) = 0$ is sufficient for L to be rigid (but not necessary).

IV. OBSTRUCTIONS

After defining the nontrivial infinitesimal deformations, the next natural question arises: Is it possible to extend an infinitesimal deformation to a deformation of higher order? The answer is "no" in general. To extend it (represented by a cocycle φ_1) to second order, parametrized by $\mathbb{C}[t]/(t^3)$, it is necessary and sufficient that $[\varphi_1, \varphi_1]$ is cohomologous to 0, which means that it must be a coboundary. The Jacobi identity of order 2 is

$$-2\, d\varphi_2 = [\varphi_1, \varphi_1].$$

If φ_2 is a cochain such that this identity is satisfied, then we can define a deformation of order 2 with the bracket

$$\mu_t = \mu_0 + \varphi_1 t + \varphi_2 t^2,$$

where φ_2 is well defined up to a two-cocycle. The cohomology class of $[\varphi_1, \varphi_1]$ is the *first obstruction* to forming a one-parameter family of deformations whose first term is cohomologous to φ_1. If it vanishes, another obstruction may show up at the next level. To extend it to a third-order deformation parametrized by $\mathbb{C}[t]/(t^4)$, it is necessary and sufficient that $[\varphi_1, \varphi_2]$ is also cohomologous to zero. If φ_3 is a cochain such that

$$-2\, d\varphi_3 = [\varphi_1, \varphi_2] + [\varphi_2, \varphi_1],$$

(the class $[\varphi_1, \varphi_1]$ is zero), then we can define a deformation of order three with the bracket

$$\mu_t = \mu_0 + \varphi_1 t + \varphi_2 t^2 + \varphi_3 t^3.$$

Here, φ_3 is also defined up to a two-cocycle. The cohomology class of $[\varphi_1, \varphi_2]$ is the *second obstruction* to forming a one-parameter family of deformations, whose first term is cohomologous to φ_1.

In general, let us define in $H^{\cdot}(L;L)$ higher operations, called *Massey operations*. These operations of order n are partially defined and they are well-defined modulo those of order $(n-1)$. It is enough to define them on the homogeneous elements. For $y_1 \in H^2(L;L)$ and $y_2 \in H^2(L;L)$ the Massey operation of order 2 is the superbracket. Suppose that for $y_1 \in H^2(L;L)$, $y_2 \in H^2(L;L)$, and $y_3 \in H^2(L;L)$, $[y_i, y_j] = 0$. Then for the cocycles x_i representing y_i, $[x_i x_j] = dx_{ij}$, where x_{ij} are one-cochains. Then the Massey operation of order 3, $[y_1, y_2, y_3]$ takes value in the factor space

$$\frac{H^3(L;L)}{[y_1, H^2(L;L)] + [y_2, H^2(L;L)] + [y_3, H^2(L;L)]},$$

This cocycle is not well defined and depends on the choice of x_{ij}, but its image in the factor space is well defined.

In general, Massey operations are defined on n classes of any cohomology space $H^{k_1}(L;L)$, $H^{k_2}(L;L)$,... such that the operations of smaller order are all cohomologous to zero.[10,13] More generally, Massey operations can be defined in $H^{\cdot}(L;A)$, where A is any L module. In the case $A = L$, the Massey operations take value in $H^3(L;L)$, and they are closely connected with the obstructions to "expanding" an infinitesimal deformation of the Lie algebra.

Theorem 1.2: If all the Massey products from $H^2(L;L)$ of an infinitesimal deformation are cohomologous to zero then there exists a formal deformation of the Lie algebra L, continuing the given infinitesimal deformation.[13] (The converse is obviously true.)

Remark: Whether this series converges or not remains a question.

After finding the obstructions, we can compute a versal deformation for the given Lie algebra step by step. A versal deformation of order one is given with the help of infinitesimal deformations and is parametrized by $K[t_1,...,t_n]/(m^2)$, where m is the maximal ideal in $K[t_1,...,t_n]$:

$$\mu_t = \mu_0 + \varphi_1 t_1 + \cdots + \varphi_n t_n.$$

Let us try to extend this deformation to a versal deformation of order 2 parametrized by $K[t_1,...,t_n]/I$ where I contains m^3. The bracket should be of the form

$$\mu_t = \mu_0 + \sum_{i=1}^{n} \varphi_i t_i + \sum \varphi_{ij} t_i t_j,$$

with the conditions that

$$-2 \sum d\varphi_{ij}\, t_i t_j = \sum [\varphi_i, \varphi_j] t_i t_j \quad (\text{mod } I)$$

(see Refs. 10 and 11). The conditions for the coefficients of a versal deformation can be obtained from the deformation equation step by step.

V. VIRASORO ALGEBRA AND ITS SUBALGEBRAS

Consider the complexification $\mathscr{L}$ of the Lie algebra of polynomial vector fields on the circle:

$$e_k \to (\exp ik\varphi) \frac{d}{d\varphi},$$

where φ is the angular parameter. The Lie algebra $\mathscr{L}$ is called the *Witt algebra*. It is well known[14] that $\mathscr{L}$ has a unique nontrivial one-dimensional central extension. The extended Lie algebra $\widehat{\mathscr{L}}$ is called the *Virasoro algebra*.

Theorem 2.1: For the maximal nilpotent subalgebra L_1 of the Virasoro algebra,

$$\dim H^q(L_1;L_1) = 2q - 1,$$

and the space $H^q(L_1;L_1)$ is generated by elements of weight $-(3q^2 - q)/2 + i$, where $q > 0$, $i = 1,2,...,2q - 1$.

Proof: There are two alternate proofs. The first proof[10] uses Feigin–Fuks spectral sequences[15] and Goncharowa's result[16] on the cohomology with trivial coefficients (see also Ref. 17). The second proof is similar to the procedure to determine the cohomology of maximal nilpotent subalgebra of a complex semisimple Lie algebra with coefficients in an irreducible representation, see Ref. 18 for details.

Corollary: The Lie algebra L_1 has three nonequivalent infinitesimal deformations. Denote the cocycles representing the different cohomology classes by α, β, and γ, where α is of weight -2, β of weight -3, and γ of weight -4. Such cocycles are given explicitly in Ref. 10.

Theorem 2.2: The Witt algebra and the Virasoro algebra are rigid.

Proof: This follows from Ref. 16 and Theorem 2.1.

Theorem 2.3: For the Lie algebra L_1, the infinitesimal deformation of weight -2 can be extended to a real deformation. The one with weight -3 can be extended to a deformation of order 2, but not of higher order, and the infinitesimal deformation of weight -4 cannot be extended at all.

Proof: It follows from computing the possible Massey operations.[10] A nice realization of the extended deformation of weight -2 is the following. Denote by $L_1(t)$ the Lie algebra of vector fields $(x^2 + t)\varphi(t)(d/dx)$. Define a linear isomorphism $\epsilon_t: L_1 \to L_1(t)$ by the formula

$$\epsilon_t(e_i) = (x^2 + t)x^{i-1}\left(\frac{d}{dx}\right) = e_i + te_{i-2}.$$

Then

$$[e_i,e_j]_t = \epsilon_t^{-1}[\epsilon_t(e_i),\epsilon_t(e_j)]$$

defines a deformation of L_1 of weight -2.

Theorem 2.4: Let L_k denote the subalgebra of the Witt algebra with the basis $e_j = x^{j+1}(d/dx)$, $j = k, k+1,...$. The cohomology spaces $H^q(L_k;L_s)$ are finite dimensional for each positive integer k and s:

$$\dim H^q(L_k;L_s) \leqslant k \dim H^q(L_{k+1}:\mathbb{C}) + (k+1-s)\dim H^q(L_k:\mathbb{C}).$$

Proof: The cohomology $H^{\cdot}(L_k;L_s)$ can be computed with the help of the spectral sequence, associated to the filtration

$$L_s \supset L_{s+1} \supset L_{s+2} \supset \cdots$$

in the coefficient module.[19]

Corollary: Each of the Lie algebras L_K, $k \geqslant 1$ has a finite number of nonequivalent deformations.

Remark: The upper bound seems to be rather crude because for $q = 1$ and $k = s$ it gives

$$\dim H^1(L_k;L_k) \leqslant k^2 + 3k + 1,$$

while a direct computation shows that the precise dimension of this cohomology space is k.

Theorem 2.5: For the Lie algebra L_0 of vector fields on the line, vanishing at the origin,

$$H^q(L_0;L_0) = 0, \quad \text{for each } q \geqslant 1.$$

Particularly, the Lie algebra L_0 is rigid.

Proof: It follows from constructing the corresponding spectral sequence and the results for trivial coefficient cohomology:

$$H^q(L_0) = \begin{cases} \mathbb{C}, & \text{for } q = 0,1, \\ 0, & \text{for } q > 1. \end{cases}$$

VI. AFFINE ALGEBRAS AND THEIR SUBALGEBRAS

Let $\mathfrak{g}$ denote an affine Kac–Moody Lie algebra (twisted or untwisted) and $\mathfrak{g}_+$ denote its maximal nilpotent subalgebra ($\mathfrak{g} = \mathfrak{g}_- \oplus h \oplus g_+$, see e.g., Ref. 20). Recall that the one-dimensional cohomology space with coefficients in the adjoint representation corresponds to the exterior derivations of the given Lie algebra.

Theorem 3.1: A basis in the space of exterior derivations of the Lie algebra $\mathfrak{g}_+$ is the following:

$$\mathbf{h}_i: \mathfrak{g} \to [h_i,\mathfrak{g}], \quad i = 1,...,n-1,$$

$$\boldsymbol{\tau}_i: t^{is+1}\left(\frac{d}{dt}\right), \quad i = 0,1,2...,$$

where s is the order of the exterior automorphism of the corresponding finite-dimensional simple Lie algebra.

Proof: Let $\mathfrak{g} = \oplus_{i>0}\mathfrak{g}_i$ be a nilpotent graded Lie algebra and $B = \oplus B_i$ a graded $\mathfrak{g}$ module. The space of k chains $C_k^{(m)}(\mathfrak{g};B)$ is spanned by monomials of the form

$$\mathfrak{g}_1 \wedge \cdots \wedge \mathfrak{g}_k \otimes b,$$

where $\mathfrak{g}_s \in g_{i_s}$, $b \in B_j$, $i_1 + \cdots + i_k + j = m$. Denote by $F_p C_k^{(m)}(\mathfrak{g};B)$ the subspace of $C_k^{(m)}(\mathfrak{g};B)$ generated by monomials with $i_1 + \cdots + i_k \leqslant p$. Obviously, $\{F_p\}$ is a decreasing filtration. Let us apply the spectral sequence corresponding to this filtration to the computation of the homology of $\mathfrak{g}_+$ with coefficients in the coadjoint representation $\mathfrak{g}_+^*$ (which is equivalent to the computation of the cohomology of $\mathfrak{g}_+$ with coefficients in the adjoint representation). For each of the affine Lie algebras the terms and differentials of this spectral sequence can be explicitly determined.[21]

Examples of infinitesimal deformations of the Lie algebra $\mathfrak{g}_+$:

(i) If $\alpha \in H^1(\mathfrak{g}_+;\mathfrak{g}_+)$ and $\beta \in H^1(\mathfrak{g}_+;\mathbb{C})$ then $\alpha\beta \in H^2(\mathfrak{g}_+;\mathfrak{g}_+)$. The number of such deformations is $\dim H^1(\mathfrak{g}_+;\mathfrak{g}_+) \cdot \dim H^1(\mathfrak{g}_+;\mathbb{C})$. (In Theorem 3.1 we saw that the first factor, and hence the product, is infinite.)

(ii) Let $1 \leqslant i \leqslant n$, where n is the rank of $\mathfrak{g}$. Define a deformation of $\mathfrak{g}_+$ inside $\mathfrak{g}$: e_i deforms into $e_i + tf_i$, the other additive generators of $\mathfrak{g}_+$ do not change. The number of such deformations is n.

(iii) Let $1 \leqslant i, j \leqslant n$, such that in the Cartan matrix $a_{ij} = -1$. If $a_{ij} = a_{ji}$, choose $i < j$. The Lie algebra $\mathfrak{g}_+$ again deforms inside $\mathfrak{g}$, with e_i deforming into $e_i + tf_j$, and $[e_i, e_j]$ into $[e_i, e_j] - th_j$, while the other additive generators do not change. The number of such deformations is the number of nonzero pairs (a_{ij}, a_{ji}) with $i \neq j$.

Theorem 3.2: For all affine Lie algebras $\mathfrak{g}$ except $\widetilde{A}_1$: (i) All the homogeneous infinitesimal deformations of $\mathfrak{g}_+$ may be extended to real deformations.

(ii) The space of infinitesimal deformations, $H^2(\mathfrak{g}_+;\mathfrak{g}_+)$, is spanned by deformations described in (i), (ii), and (iii). Thus all deformations arise from these infinitesimal ones.

Proof: There are two methods. The first proof uses filtration in the cochain complex and the corresponding spectral sequence.[21] The other method[22] uses results for the maximal nilpotent algebra n_+ of finite-dimensional semisimple Lie algebras. We know $H^{\cdot}(n_+;s)$ where s is the adjoint representation of the finite-dimensional semisimple algebra s. Consider the exact sequences of n_+ modules:

$$0 \rightarrow n_+ \rightarrow s \rightarrow s/n_+ \rightarrow 0, \quad 0 \rightarrow h \rightarrow s/n_+ \rightarrow n_+^* \rightarrow 0$$

(h is the Cartan subalgebra of s). These sequences allow us to reduce the computation of $H^2(n_+;n_+)$ to that of $H^1(n_+;n_+^*)$ which can be computed directly. Generalizing this method for the infinite dimensional affine algebras, we get the statement of Theorem 3.2.

Theorem 3.3: The case of $\widetilde{A}_1$ is an exceptional case because there are two additional infinitesimal deformations not listed in (i), (ii) [type (iii) does not exist]. These two infinitesimal deformations cannot be extended to real deformations of $\mathfrak{g}_+$, even not to order 2 (because their Massey square is nonzero).

Proof: By direct computation with the help of the described spectral sequence.[21] The generalization of the finite dimensional method[22] does not work for this particular case.

Theorem 3.4: All the affine Lie algebras are rigid.

Proof: It follows from Theorem 3.2 for $\mathfrak{g} \neq \widetilde{A}_1$ and by direct computation for $\mathfrak{g} = \widetilde{A}_1$. For an independent proof see Ref. 23.

[1]C. Chevalley and S. Eilenberg, Trans. Am. Math. Soc. **63**, 85 (1948).
[2]A. N. Rudakov, Izv. Acad. USSR, Math. **35**, 1113 (1971).
[3]A. S. Dzumadil'daev, Sov. Math. Dokl. **21**, 605 (1980).
[4]A. S. Dzumadil'daev and A. I. Kostrikin, Tr. Mat. Inst. Steklov **148**, 141 (1978).
[5]G. Leger and E. M. Luks, Trans. Am. Math. Soc. **195**, 305 (1974).
[6]W. S. Piper, J. Diff. Geom. **5**, 437 (1977).
[7]R. Moody, Lecture given at the Conference "Kac–Moody Lie algebras and physics," North Carolina, 13–17 December, 1988.
[8]I. M. Gel'fand and D. B. Fuks, Funct. Anal. Appl. **3**, 194 (1969), **4**, 110 (1970).
[9]H. Garland and J. Lepowsky, Invent. Math. **34**, 37 (1976).
[10]A. Fialowski, in *Proceedings of the NATO–ASI Conference on Deformation Theory of Algebras and Applications*, 1986 (Kluwer, Dordrecht, 1988), p. 384.
[11]A. Fialowski, Math. USSR Sbornik **55**, 467 (1986).
[12]M. Schlessinger, Trans. Am. Math. Soc. **130**, 208 (1968).
[13]V. S. Retach, Funct. Anal. Appl. **11**, 319 (1977).
[14]I. M. Gel'fand and D. B. Fuks, Funct. Anal. Appl. **2**, 342 (1968).
[15]B. L. Feigin and D. B. Fuks, Funct. Anal. Appl. **14**, 45 (1980).
[16]L. V. Goncharowa, Funct. Anal. Appl. **7**, 91 (1973).
[17]D. B. Fuks, *Cohomology of Infinite Dimensional Lie Algebras*, Contemporary Soviet Math. (Consultants Bureau, New York, 1986).
[18]B. L. Feigin and A. Fialowski, Bull. Am. Math. Soc. **17**, 333 (1987).
[19]A. Fialowski, "On the cohomology $H^*(L_k;L_S)$," Preprint, 1988.
[20]V. G. Kac, *Infinite-dimensional Lie Algebras* (Cambridge U.P., London, 1986).
[21]A. Fialowski, Stud. Sci. Math. Hung. **19**, 467 (1986).
[22]B. L. Feigin and A. Fialowski, Stud. Sci. Math. Hung. **23**, 477 (1989).
[23]P. Lecomte and C. Roger, "Rigidity of current Lie algebras of complex simple type," MSRI Preprint No. 15311-85, Berkeley, 1986.

Solvable Lie algebras of dimension six

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

(Received 19 September 1989; accepted for publication 10 January 1990)

Six-dimensional solvable Lie algebras over the field of real numbers that possess nilradicals of dimension four are classified into equivalence classes. This completes Mubarakzyanov's classification of the real six-dimensional solvable Lie algebras.

I. INTRODUCTION

On the strength of the Levi theorem, the Lie algebras fall into the following three categories: The semisimple algebras, the solvable algebras, and the semidirect sums of solvable and semisimple algebras. A list is available of the semisimple Lie algebras of a finite dimension; this result is due to Cartan. Recently, mainly because of the cosmological applications, one observes a growing interest in the field of the real Lie algebras that also possess a nonsemisimple structure. Jacobson's monograph[1] offers the enumeration of the solvable algebras of low dimensions. However, it is not an easy task to find a classification method to extend directly to the case of higher-dimensional solvable algebras. In 1962, Mubarakzyanov attempted to obtain such a method.[2,3] Then he successfully applied it for algebras of dimension less than five.[3] Over the next few months he completely solved the classification problem for five-dimensional solvable algebras[4] and partially for six-dimensional solvable algebras.[5]

In this paper we consider the problem of finding all equivalence classes for the six-dimensional solvable Lie algebras N_6 over the field $\mathbb{R}$ of real numbers. Since the dimension of a nilradical NR of N is $\dim NR \geqslant \frac{1}{2} \dim N$ (see Ref. 3), there are four possibilities to consider: Nilpotent six-dimensional algebras and solvable algebras that contain five-, four-, or three-dimensional nilradicals. In a little-known dissertation,[6] on nilpotent Lie algebras, Umlauf classified nilpotent algebras of dimension six over the field of complex numbers. A similar classification of nilpotent six-dimensional algebras was performed by Morozov[7] (over any field of characteristic 0) and, independently, by Skjelbred and Sund[8] (over $\mathbb{R}$). Algebras N_6 that contain five-dimensional nilradicals were classified in the paper mentioned above by Mubarakzyanov into 99 equivalence classes.[5] Algebras N_6 that contain three-dimensional nilradical are decomposable.[2,5] Therefore, all we have to do to complete the classification of six-dimensional solvable algebras, is to classify the algebras that contain nilradicals of dimension four.

In the following, a method to obtain the solvable algebras is presented. We notice[4] that Mubarakzyanov solved the problem of obtaining N_5 by a similar method. However, he did not use the notions of *derivation* and *semidirect sum* explicitly; they are convenient in performing the classification of Lie algebras.[9–11] In Sec. III an application to six-dimensional algebras is given. There are 27 algebras of dimension six that contain Abelian nilradicals of dimension four; they are summarized in Tables I and II. Tables III, IV, and V provide algebras of dimension six that contain non-Abelian nilradicals of dimension four. We use the results of Sec. III in another paper in which nine-dimensional algebras that admit a nontrivial Levi decomposition are expressed in terms of six-dimensional solvable algebras and three-dimensional simple algebras.[11]

II. A METHOD TO OBTAIN THE SOLVABLE ALGEBRAS

A solvable Lie algebra N has a decomposition of the form

$$N = NR \dotplus X, \tag{1}$$

satisfying

$$[NR,NR] \subset NR, \quad [NR,X] \subseteq NR, \quad [X,X] \subset NR, \tag{2}$$

where NR denotes *nilradical* of N, the vector space X is spanned by remaining generators and $\dotplus$ denotes the direct sum of vector spaces. Let us define the commutators of the $[NR,X]$ type by

$$[x,n] = D(x)\star n, \tag{3}$$

where $x \in X$, $n \in N$, and $D(x)$ is a linear mapping, $D(x)$: $NR \ni n \to D(x)\star n \in NR$.

The Jacobi identity for the triples $\{x, n_1, n_2\}$ implies that $D(x)$ are *derivations* in NR, $D(x) \in \mathrm{Der}(NR)$,

$$D(x)\star[n_1,n_2] = [D(x)\star n_1,n_2] + [n_1,D(x)\star n_2], \tag{4}$$

for every x in X, and n_1, n_2 in NR. Since $x \notin NR$, these derivations are not nilpotent and, as a consequence, they are *outer* ($\equiv$noninner) in NR. Every nilpotent algebra has an outer derivation.[12] However, nilpotent algebras are known, called *characteristically nilpotent*, that possess only nilpotent derivations;[13,14] these are not nilradicals of any solvable Lie algebra.

The Jacobi identity involving two x's and one n implies

$$[D(x_1),D(x_2)] = \mathrm{ad}_{NR}([x_1,x_2]), \tag{5}$$

where ad_{NR} is the restriction of the adjoint representation to NR: $\mathrm{ad}_{NR}(n_i)\star n \equiv [n_i,n]_{NR}$. Therefore, we have $[D(x_i),D(x_j)] \in \mathrm{Inn}(NR)$, for every $x_i \in X, x_j \in X$ where $\mathrm{Inn}(NR)$ denotes the algebra of inner derivations in NR. For the Abelian nilradical, it follows that $[D(x_i),D(x_j)] = 0$.

What changes can be made in the linear mapping $D(x) \in \mathrm{Der}(NR)$? Let $\{n_1,n_2,\ldots;x_1,x_2,\ldots\}$ be a basis of $NR \dotplus X$, where $n_i \in NR$ and $x_i \in X$. Consider $\bar{x}_k = \Sigma \alpha_{ik} x_i + \Sigma \beta_{ik} n_i$, where $A = (\alpha_{ik})$ is nonsingular. This gives the new basis $\{\bar{x}_k\}$ of X and changes $D(x_i) \in \mathrm{Der}(NR)$ to the following $D(\bar{x}_k) \in \mathrm{Der}(NR)$:

$$D(\bar{x}_k) = \sum \alpha_{ik} D(x_i) + \sum \beta_{ik}\, \mathrm{ad}_{NR}(n_i). \tag{6}$$

A change of basis in the nilradical $\bar{n}_k = \Sigma \tau_{ik} n_i$, where

 0022-2488/90/061344-07$03.00

$T=(\tau_{ik})$ is the automorphism, will change every $D(x_i)$ to a similar matrix $TD(x_i)T^{-1}$.

If the nilradical NR of a solvable algebra is given, the classification problem reduces to the one of finding the derivations of the nilradical that are not nilpotent and that satisfy Eq. (5). These derivations define the commutators of the $[NR,X]$ type. Finally, the structure of the subspace X, i.e., commutation relations $[X,X]$ can be found.

III. A CLASS OF SOLVABLE ALGEBRAS

Now, we shall determine all real solvable algebras $N=NR\dotplus X$ such that $\dim NR=4$ and $\dim X=2$. The dimension of the center of such the algebra is

$$\dim Z(N)\leqslant 2\dim NR-\dim N=2 \qquad (7)$$

(see Ref. 15). The algebras that possess a two-dimensional center are decomposable into a direct sum of lower-dimensional algebras.[15] Therefore, in the following, the classification problem is solved for the cases $\dim Z(N)=0$ and $\dim Z(N)=1$, respectively.

We start with a nilpotnet algebra that forms the nilradical NR. There are the following three nilpotent algebras of dimension four to be considered: The Abelian algebra $4A_1$, $A_{4,1}$, and $A_{3,1}\oplus A_1$; in our notation we follow the scheme of Patera *et al.*: the term $A_{r,j}$ dentoes an r-dimensional algebra of jth type and the term nA_1 denotes the n-dimensional Abelian algebra.[16]

To specify the commutators of the $[NR,X]$ type of the pair $\{D(x_1), D(x_2)\}$ of the following type has to be given: (a) $D(x_1)$ and $D(x_2)$—derivations in the nilradical. (b) Every linear combination $\alpha D(x_1)+\beta D(x_2)$; where α, $\beta\in\mathbb{R}$ and $\alpha^2+\beta^2\neq 0$, are non-nilpotent. The operators (matrices) $D(x_1)$ and $D(x_2)$ that possess the property (b) shall be called, after Mubarakzyanov, *nil-independent* matrices.

If the structure of the nilradical is given, then we have as many nonisomorphic algebras $N=NR\dotplus X$ as there are equivalence classes of $\{D(x_1), D(x_2)\}$. The expression "equivalence classes" should be understood in the sence that:

(A) The pair

$$\{\alpha_{11}D(x_1)+\alpha_{21}D(x_2)+\mathrm{ad}_{NR}(n),\alpha_{12}D(x_1)+\alpha_{22}D(x_2)+\mathrm{ad}_{NR}(n')\},$$

where $\alpha_{11}\alpha_{22}-\alpha_{21}\alpha_{12}\neq 0, n\in NR$, and $n'\in NR$, is equivalent to $\{D(x_1), D(x_2)\}$.

(B) The pair $\{TD(x_1)T^{-1}, TD(x_2)T^{-1}\}$, where T is the automorphism of NR, is equivalent to $\{D(x_1), D(x_2)\}$.

Let us consider the Abelian nilradical $NR=4A_1$. Then every linear transformation sending NR into NR is an outer derivation in NR. We are allowed to choose nil-independent pairs of $\{D(x_1), D(x_2)\}$ in the following form:

$$\begin{pmatrix}\alpha&&&\\&\gamma&&\\&&0&\\&&&1\end{pmatrix},\ \begin{pmatrix}\beta&&&\\&\delta&&\\&&1&\\&&&0\end{pmatrix},\quad \alpha\beta\neq 0,\ \gamma^2+\delta^2\neq 0, \qquad (8)$$

$$\begin{pmatrix}\alpha&&&\\&1&&\\&&0&\\&&1&0\end{pmatrix},\ \begin{pmatrix}\beta&&&\\&\gamma&&\\&&1&\\&&&1\end{pmatrix},\quad \alpha^2+\beta^2\neq 0, \qquad (9)$$

$$\begin{pmatrix}1&&&\\&1&&\\&&0&\\&&1&0\end{pmatrix},\ \begin{pmatrix}\alpha&&&\\1&\alpha&&\\&&1&\\&&&1\end{pmatrix}, \qquad (10)$$

$$\begin{pmatrix}1&&&\\&1&&\\&&0&\\&&1&0\end{pmatrix},\ \begin{pmatrix}&-1&&\\1&&&\\&&\alpha&\\&&\beta&\alpha\end{pmatrix},\quad \alpha\neq 0, \qquad (11)$$

$$\begin{pmatrix}\alpha&&&\\&0&&\\&&1&\\&&1&1\end{pmatrix},\ \begin{pmatrix}\beta&&&\\&1&&\\&&0&\\&&&0\end{pmatrix},\quad \alpha\beta\neq 0, \qquad (12)$$

$$\begin{pmatrix}\alpha&&&\\&\alpha&&\\&&1&\\&&1&1\end{pmatrix},\ \begin{pmatrix}1&&&\\1&1&&\\&&0&\\&&\beta&0\end{pmatrix},\quad \alpha^2+\beta^2\neq 0, \qquad (13)$$

$$\begin{pmatrix}\alpha&&&\\&\alpha&&\\&&1&\\&&1&1\end{pmatrix},\ \begin{pmatrix}\gamma&-1&&\\1&\gamma&&\\&&0&\\&&\beta&0\end{pmatrix},\quad \alpha^2+\beta^2\neq 0, \qquad (14)$$

$$\begin{pmatrix}1&&&\\&0&&\\&&0&\\&1&&0\end{pmatrix},\ \begin{pmatrix}0&&&\\&1&&\\&&1&\\&&1&1\end{pmatrix}, \qquad (15)$$

$$\begin{pmatrix}1&&&\\&0&&\\&&0&\\&1&&0\end{pmatrix},\ \begin{pmatrix}0&&&\\&1&&\\&1&1&\\&&\alpha&1\end{pmatrix}, \qquad (16)$$

$$\begin{pmatrix}\alpha&&&\\&1&&\\&&1&\\&\beta&&1\end{pmatrix},\ \begin{pmatrix}1&&&\\&0&&\\&1&0&\\&&1&0\end{pmatrix}, \qquad (17)$$

$$\begin{pmatrix}0&&&\\1&0&&\\&&1&\\&&1&1\end{pmatrix},\ \begin{pmatrix}1&&&\\&1&&\\&&\alpha&\\&&&\alpha\end{pmatrix}, \qquad (18)$$

$$\begin{pmatrix}1&&&\\1&1&&\\&&1&\\&&1&1\end{pmatrix},\ \begin{pmatrix}&&-1&\\\alpha&&\beta&-1\\1&&&\\-\beta&1&\alpha&\end{pmatrix}, \qquad (19)$$

$$\begin{pmatrix}\alpha&&&\\&\gamma&&\\&&&-1\\&&1&\end{pmatrix},\quad\begin{pmatrix}\beta&&&\\&\delta&&\\&&1&\\&&&1\end{pmatrix},\quad \alpha^2+\beta^2\neq 0,\ \gamma^2+\delta^2\neq 0, \tag{20}$$

$$\begin{pmatrix}\alpha&&&\\&0&&\\&&\gamma&-1\\&&1&\gamma\end{pmatrix},\quad\begin{pmatrix}\beta&&&\\&1&&\\&&0&\\&&&0\end{pmatrix},\quad \alpha\beta\neq 0, \tag{21}$$

$$\begin{pmatrix}1&&&\\&1&&\\&&\alpha&-\beta\\&&\beta&\alpha\end{pmatrix},\quad\begin{pmatrix}\gamma&-1&&\\1&\gamma&&\\&&\delta&\\&&&\delta\end{pmatrix},\quad \beta\neq 0, \tag{22}$$

$$\begin{pmatrix}0&&&\\1&0&&\\&&\alpha&-1\\&&1&\alpha\end{pmatrix},\quad\begin{pmatrix}1&&&\\&1&&\\&&\beta&\\&&&\beta\end{pmatrix}, \tag{23}$$

$$\begin{pmatrix}\alpha&&&\\1&\alpha&&\\&&&-1\\&&1&\end{pmatrix},\quad\begin{pmatrix}0&&&\\&0&&\\&&1&\\&&&1\end{pmatrix}, \tag{24}$$

$$\begin{pmatrix}&-1&&\\1&&&\\&&\alpha&-\beta\\&&\beta&\alpha\end{pmatrix},\quad\begin{pmatrix}1&&&\\&1&&\\&&\gamma&\\&&&\gamma\end{pmatrix},\quad \beta\neq 0, \tag{25}$$

$$\begin{pmatrix}&-1&&\\1&&&\\1&&&-1\\&1&1&\end{pmatrix},\quad\begin{pmatrix}1&&&\\&1&&\\&&1&\\&&&1\end{pmatrix}, \tag{26}$$

for algebras that contain the center of dimension zero, and

$$\begin{pmatrix}0&&&\\&\alpha&&\\&&0&\\&&&1\end{pmatrix},\quad\begin{pmatrix}0&&&\\&\beta&&\\&&1&\\&&&0\end{pmatrix},\quad \alpha^2+\beta^2\neq 0, \tag{27}$$

$$\begin{pmatrix}0&&&\\&1&&\\&&0&\\&&1&0\end{pmatrix},\quad\begin{pmatrix}0&&&\\&\alpha&&\\&&1&\\&&&1\end{pmatrix}, \tag{28}$$

$$\begin{pmatrix}1&&&\\&0&&\\&&0&\\&&1&0\end{pmatrix},\quad\begin{pmatrix}\alpha&&&\\&1&&\\&&0&\\&&&0\end{pmatrix}, \tag{29}$$

$$\begin{pmatrix}1&&&\\&1&&\\&&0&\\&&1&0\end{pmatrix},\quad\begin{pmatrix}&-1&&\\1&&&\\&&0&\\&&\alpha&0\end{pmatrix}, \tag{30}$$

$$\begin{pmatrix}0&&&\\&0&&\\&&1&\\&&1&1\end{pmatrix},\quad\begin{pmatrix}0&&&\\&1&&\\&&0&\\&&&0\end{pmatrix}, \tag{31}$$

$$\begin{pmatrix}0&&&\\&\alpha&&\\&&&-1\\&&1&\end{pmatrix},\quad\begin{pmatrix}0&&&\\&\beta&&\\&&1&\\&&&1\end{pmatrix},\quad \alpha^2+\beta^2\neq 0, \tag{32}$$

$$\begin{pmatrix}0&&&\\&0&&\\&&\alpha&-1\\&&1&\alpha\end{pmatrix},\quad\begin{pmatrix}0&&&\\&1&&\\&&0&\\&&&0\end{pmatrix}, \tag{33}$$

$$\begin{pmatrix}0&&&\\1&0&&\\&&&-1\\&&1&\end{pmatrix},\quad\begin{pmatrix}0&&&\\&0&&\\&&1&\\&&&1\end{pmatrix}, \tag{34}$$

for algebras that contain one-dimensional centers. The restrictions on the parameters in the matrices given above are made both to avoid decomposability into a direct sum of lower-dimensional algebras, and to obtain the dimension of the center of the algebra equal to 0 or 1, as required. Every pair of matrices given above defines a different algebra.

If one assumes $\dim Z(N)=0$ then a basis of X can be found such that $[X,X]=0$. Therefore, $X=2A_1$ is the complement of the nilradical in N_6, and decomposition (1) becomes the semidirect sum of the ideal NR and the subalgebra X: $N=NR\oplus 2A_1$. For algebras that have a nonzero center, there is $[X,X]\neq 0$, in general. However, some simplifications of the commutators $[x_1,x_2]$ can be made (see below).

Our results are summarized in Tables I and II where 27 algebras of dimension six are given that contain an Abelian nilradical and a center of dimension of 0 and 1, respectively. In the tables are given the names of the corresponding algebra and all the nonzero commutators. We use the names $N_{6,j}^{\alpha\beta\ldots}$, where $j=1,\ldots,27$ denotes the consecutive number of the type and the superscripts, if any, give the parameters on which the algebra depends. The basis of nilradical is denoted by $\{n_1,\ldots,n_4\}$, and the additional basis elements are denoted by x_1 and x_2. We write commutators in the shortened form $[n_1x_2]$ instead of $[n_1,x_2]$.

Let us consider algebras that contain a non-Abelian nilradical $A_{4,1}$. The algebra $A_{4,1}$ is defined by the following nonvanishing commutators: $[n_2,n_4]=n_1$ and $[n_3,n_4]=n_2$, where $\{n_1,\ldots,n_4\}$ form the basis of $A_{4,1}$. Suppose $D(x)\star n_i=\Sigma\delta_{ki}n_k$, where $D(x)\in\mathrm{Der}(A_{4,1})$. Therefore, the matrix of $D(x)$ is

TABLE I. Real solvable Lie algebras of dimension six that contain the Abelian nilradical of dimension four and the center of dimension zero. $\{n_1,...,n_4\}$ form a basis for the nilradical and $\{x_1, x_2\}$ are the remaining basis elements.

Name	Nonzero commutation relations		
$N_{6,1}^{\alpha\beta\gamma\delta}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_2]=\gamma n_2$,	$[x_1n_4]=n_4$,
$\alpha\beta\neq 0$,	$[x_2n_1]=\beta n_1$,	$[x_2n_2]=\delta n_2$,	$[x_2n_3]=n_3$
$\gamma^2+\delta^2\neq 0$			
$N_{6,2}^{\alpha\beta\gamma}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_2]=n_2$,	$[x_1n_3]=n_4$,
$\alpha^2+\beta^2\neq 0$	$[x_2n_1]=\beta n_1$,	$[x_2n_2]=\gamma n_2$,	$[x_2n_3]=n_3$,
	$[x_2n_4]=n_4$		
$N_{6,3}^{\alpha}$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,	$[x_1n_3]=n_4$,
	$[x_2n_1]=\alpha n_1+n_2$,	$[x_2n_2]=\alpha n_2$,	$[x_2n_3]=n_3$,
	$[x_2n_4]=n_4$		
$N_{6,4}^{\alpha\beta}$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,	$[x_1n_3]=n_4$,
$\alpha\neq 0$	$[x_2n_1]=n_2$,	$[x_2n_2]=-n_1$,	
	$[x_2n_3]=\alpha n_3+\beta n_4$,	$[x_2n_4]=\alpha n_4$	
$N_{6,5}^{\alpha\beta}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_3]=n_3+n_4$,	
$\alpha\beta\neq 0$	$[x_1n_4]=n_4$,	$[x_2n_1]=\beta n_1$,	$[x_2n_2]=n_2$
$N_{6,6}^{\alpha\beta}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_2]=\alpha n_2$,	
$\alpha^2+\beta^2\neq 0$	$[x_1n_3]=n_3+n_4$,	$[x_1n_4]=n_4$,	
	$[x_2n_1]=n_1+n_2$,	$[x_2n_2]=n_2$,	$[x_2n_3]=\beta n_4$
$N_{6,7}^{\alpha\beta\gamma}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_2]=\alpha n_2$,	
$\alpha^2+\beta^2\neq 0$	$[x_1n_3]=n_3+n_4$,	$[x_1n_4]=n_4$,	
	$[x_2n_1]=\gamma n_1+n_2$,	$[x_2n_2]=-n_1+\gamma n_2$,	$[x_2n_3]=\beta n_4$
$N_{6,8}$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_4$,	
	$[x_2n_2]=n_2$,	$[x_2n_3]=n_3+n_4$,	$[x_2n_4]=n_4$
$N_{6,9}^{\alpha}$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_4$,	
	$[x_2n_2]=n_2+n_3$,	$[x_2n_3]=n_3+\alpha n_4$,	$[x_2n_4]=n_4$
$N_{6,10}^{\alpha\beta}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_2]=n_1+\beta n_4$,	$[x_1n_3]=n_3$,
	$[x_1n_4]=n_4$,	$[x_2n_1]=n_1$,	$[x_2n_2]=n_3$,
	$[x_2n_3]=n_4$		
$N_{6,11}^{\alpha}$	$[x_1n_1]=n_2$,	$[x_1n_3]=n_3+n_4$,	$[x_1n_4]=n_4$,
	$[x_2n_1]=n_1$,	$[x_2n_2]=n_2$,	$[x_2n_3]=\alpha n_3$,
	$[x_2n_4]=\alpha n_4$		
$N_{6,12}^{\alpha\beta}$	$[x_1n_1]=n_1+n_2$,	$[x_1n_2]=n_2$,	
	$[x_1n_3]=n_3+n_4$,	$[x_1n_4]=n_4$,	
	$[x_2n_1]=\alpha n_2+n_3-\beta n_4$,	$[x_2n_2]=n_4$,	
	$[x_2n_3]=-n_1+\beta n_2+\alpha n_4$,	$[x_2n_4]=-n_2$	
$N_{6,13}^{\alpha\beta\gamma\delta}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_2]=\gamma n_2$,	$[x_1n_3]=n_4$,
$\alpha^2+\beta^2\neq 0$	$[x_1n_4]=-n_3$,	$[x_2n_1]=\beta n_1$,	$[x_2n_2]=\delta n_2$,
$\gamma^2+\delta^2\neq 0$	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$	
$N_{6,14}^{\alpha\beta\gamma}$	$[x_1n_1]=\alpha n_1$,	$[x_1n_3]=\gamma n_3+n_4$,	
$\alpha\beta\neq 0$	$[x_1n_4]=-n_3+\gamma n_4$,	$[x_2n_1]=\beta n_1$,	$[x_2n_2]=n_2$
$N_{6,15}^{\alpha\beta\gamma\delta}$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,	
$\beta\neq 0$	$[x_1n_3]=\alpha n_3+\beta n_4$,	$[x_1n_4]=-\beta n_3+\alpha n_4$,	
	$[x_2n_1]=\gamma n_1+n_2$,	$[x_2n_2]=-n_1+\gamma n_2$,	
	$[x_2n_3]=\delta n_3$,	$[x_2n_4]=\delta n_4$	
$N_{6,16}^{\alpha\beta}$	$[x_1n_1]=n_2$,	$[x_1n_3]=\alpha n_3+n_4$,	
	$[x_1n_4]=-n_3+\alpha n_4$,	$[x_2n_1]=n_1$,	$[x_2n_2]=n_2$,
	$[x_2n_3]=\beta n_3$,	$[x_2n_4]=\beta n_4$	
$N_{6,17}^{\alpha}$	$[x_1n_1]=\alpha n_1+n_2$,	$[x_1n_2]=\alpha n_2$,	$[x_1n_3]=n_4$,
	$[x_1n_4]=-n_3$,	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$
$N_{6,18}^{\alpha\beta\gamma}$	$[x_1n_1]=n_2$,	$[x_1n_2]=-n_1$,	
$\beta\neq 0$	$[x_1n_3]=\alpha n_3+\beta n_4$,	$[x_1n_4]=-\beta n_3+\alpha n_4$,	
	$[x_2n_1]=n_1$,	$[x_2n_2]=n_2$,	
	$[x_2n_3]=\gamma n_3$,	$[x_2n_4]=\gamma n_4$	

TABLE I. (*Continued.*)

Name	Nonzero commutation relations		
$N_{6,19}$	$[x_1n_1]=n_2+n_3$,	$[x_1n_2]=-n_1+n_4$,	
	$[x_1n_3]=n_4$,	$[x_1n_4]=-n_3$,	$[x_2n_1]=n_1$,
	$[x_2n_2]=n_2$,	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ 0 & \delta_{22} & \delta_{12} & \delta_{24} \\ 0 & 0 & 2\delta_{22}-\delta_{11} & \delta_{34} \\ 0 & 0 & 0 & \delta_{11}-\delta_{22} \end{pmatrix}. \tag{35}$$

Let us take $D(x_1)=(\delta_{ik})$ and $D(x_2)=(\delta'_{ik})$ in the form given by (35). The matrices $D(x_1)$, $D(x_2)$ should be nil-independent. Therefore, let us assume $\delta_{11}=\delta'_{22}=1$ and $\delta_{22}=\delta'_{11}=0$. Furthermore, we may add an inner derivation to $D(x)$; more precisely, let us define

$$x_1\rightarrow x_1+(\mu-\delta_{14})n_2-\delta_{24}n_3+\delta_{12}n_4,$$

and

$$x_2\rightarrow x_2-\delta'_{14}n_2-\delta'_{24}n_3+\delta'_{12}n_4,$$

where

$$\mu\equiv\delta_{13}\delta'_{34}-\delta'_{13}\delta_{34},$$

so that the resulting derivations $D(x_1)$, $D(x_2)$ have the form

$$\begin{pmatrix} 1 & 0 & \delta_{13} & \mu \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & \delta_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \delta'_{13} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & \delta'_{34} \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{36}$$

Suppose now that,

$$[D(x_1),D(x_2)]\in\mathrm{Inn}(A_{4,1}). \tag{37}$$

Then one has $\delta'_{13}=-\delta_{13}$ and $-2\delta'_{34}=3\delta_{34}$. Every pair (36) of nil-independent derivations $\{D(x_1),\ D(x_2)\}$ that satisfies (37) is equivalent to the following diagonal pair:

$$\begin{pmatrix} 1 & & & \\ & 0 & & \\ & & -1 & \\ & & & 1 \end{pmatrix}, \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 2 & \\ & & & -1 \end{pmatrix}. \tag{38}$$

TABLE II. Real solvable Lie algebras of dimension six that contain the Abelian nilradical of dimension four and the one-dimensional center.

Name	Nonzero commutation relations		
$N_{6,20}^{\alpha\beta}$	$[x_1n_2]=\alpha n_2$,	$[x_1n_4]=n_4$,	
$\alpha^2+\beta^2\neq0$	$[x_2n_2]=\beta n_2$,	$[x_2n_3]=n_3$,	$[x_1x_2]=n_1$
$N_{6,21}^{\alpha}$	$[x_1n_2]=n_2$,	$[x_1n_3]=n_4$,	$[x_2n_2]=\alpha n_2$,
	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$,	$[x_1x_2]=n_1$
$N_{6,22}^{\alpha\epsilon}$	$[x_1n_1]=n_1$,	$[x_1n_3]=n_4$,	$[x_2n_1]=\alpha n_1$,
$\epsilon=0,1$	$[x_2n_2]=n_2$,	$[x_1x_2]=\epsilon n_3$	
$\alpha^2+\epsilon^2\neq0$			
$N_{6,23}^{\alpha\epsilon}$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,	$[x_1n_3]=n_4$,
$\epsilon=0,1$	$[x_2n_1]=n_2$,	$[x_2n_2]=-n_1$,	$[x_2n_3]=\alpha n_4$,
	$[x_1x_2]=\epsilon n_3$		
$N_{6,24}$	$[x_1n_3]=n_3+n_4$,	$[x_1n_4]=n_4$,	$[x_2n_2]=n_2$,
	$[x_1x_2]=n_1$		
$N_{6,25}^{\alpha\beta}$	$[x_1n_2]=\alpha n_2$,	$[x_1n_3]=n_4$,	$[x_1n_4]=-n_3$,
$\alpha^2+\beta^2\neq0$	$[x_2n_1]=\beta n_2$,	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$,
	$[x_1x_2]=n_1$		
$N_{6,26}^{\alpha}$	$[x_1n_3]=\alpha n_3+n_4$,	$[x_1n_4]=-n_3+\alpha n_4$,	
	$[x_2n_2]=n_2$,	$[x_1x_2]=n_1$	
$N_{6,27}^{\epsilon}$	$[x_1n_1]=n_2$,	$[x_1n_3]=n_4$,	$[x_1n_4]=-n_3$,
$\epsilon=0,1$	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$,	$[x_1x_2]=\epsilon n_1$

TABLE III. Real solvable Lie algebra of dimension six that contains the non-Abelian nilradical $A_{4,1}$.

Name	Nonzero commutation relations		
$N_{6,28}$	$[n_2n_4]=n_1$,	$[n_3n_4]=n_2$,	
	$[x_1n_1]=n_1$,	$[x_1n_3]=-n_3$,	$[x_1n_4]=n_4$,
	$[x_2n_2]=n_2$,	$[x_2n_3]=2n_3$,	$[x_2n_4]=-n_4$

Furthermore, $[D(x_1),D(x_2)]=0$, and this implies $\mathrm{ad}_{NR}([x_1,x_2])=\mathrm{ad}_{NR}(\alpha n_1)$. Hence, we have $[x_1, x_2]=\alpha n_1$, α is an arbitrary constant. Let $\{\bar{x}_1,\bar{x}_2\}$ be another basis of X, where $\bar{x}_1=x_1$ and $\bar{x}_2=x_2-\alpha n_1$. Then $D(\bar{x}_1)=D(x_1), D(\bar{x}_2)=D(x_2)$, and $[\bar{x}_1,\bar{x}_2]=0; X=2A_1$ becomes the complement of the nilradical in the algebra N_6. Therefore, pair (38) defines the semidirect sum of $A_{4,1}$ and the Abelian $2A_1$. This is the only six-dimensional algebra that contains the nilradical $A_{4,1}$; for the commutation relations see Table III.

Finally, let us consider the algebras that contain a nilradical $A_{3,1}\oplus A_1$. The algebra $A_{3,1}\oplus A_1$ has basis $\{n_1,\ldots,n_4\}$ such that $[n_2,n_3]=n_1$ and all other products of base elements are 0. The properties of the derivation algebra $\mathrm{Der}(A_{3,1}\oplus A_1)$ have been discussed in Ref. 17. The matrix $D(x)\in\mathrm{Der}(A_{3,1}\oplus A_1)$ is

$$\begin{pmatrix} \delta_{22}+\delta_{33} & \delta_{12} & \delta_{13} & \delta_{14} \\ 0 & \delta_{22} & \delta_{23} & 0 \\ 0 & \delta_{32} & \delta_{33} & 0 \\ 0 & \delta_{42} & \delta_{43} & \delta_{44} \end{pmatrix}. \tag{39}$$

Then, let us define $\bar{x}=x-\delta_{13}n_2+\delta_{12}n_3$ which, on the strength of Eq. (6), reduces δ_{12} and δ_{13} in $D(\bar{x})$ to zero. The nil-independent derivations $D(x_1)$ and $D(x_2)$ of this form, that satisfy Eq. (5), are equivalent to the following:

$$\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 0 & \\ & & & \alpha \end{pmatrix}, \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 1 & \\ & & & \beta \end{pmatrix}, \quad \alpha^2+\beta^2\neq 0, \tag{40}$$

$$\begin{pmatrix} 2 & & & \\ & 1 & & \\ & & 1 & \\ & & & \alpha \end{pmatrix}, \begin{pmatrix} 0 & & & \\ & 0 & & \\ & 1 & 0 & \\ & & & 1 \end{pmatrix}, \tag{41}$$

TABLE IV. Real solvable Lie algebras of dimension six that contain the non-Abelian nilradical $A_{3,1}\oplus A_1$ and the center of dimension zero.

Name	Nonzero commutation relations		
$N_{6,29}^{\alpha\beta}$	$[n_2n_3]=n_1$,		
$\alpha^2+\beta^2\neq 0$	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,	$[x_1n_4]=\alpha n_4$,
	$[x_2n_1]=n_1$,	$[x_2n_3]=n_3$,	$[x_2n_4]=\beta n_4$
$N_{6,30}^{\alpha}$	$[n_2n_3]=n_1$,		
	$[x_1n_1]=2n_1$,	$[x_1n_2]=n_2$,	$[x_1n_3]=n_3$,
	$[x_1n_4]=\alpha n_4$,	$[x_2n_2]=n_3$,	$[x_2n_4]=n_4$
$N_{6,31}$	$[n_2n_3]=n_1$,		
	$[x_1n_2]=n_2$,	$[x_1n_3]=-n_3$,	$[x_2n_1]=n_1$,
	$[x_2n_3]=n_3$,	$[x_2n_4]=n_1+n_4$	
$N_{6,32}^{\alpha}$	$[n_2n_3]=n_1$,	$[x_1n_2]=n_2$,	$[x_1n_3]=-n_3$,
	$[x_1n_4]=n_1$,	$[x_2n_1]=n_1$,	$[x_2n_2]=\alpha n_2$,
	$[x_2n_3]=(1-\alpha)n_3$,	$[x_2n_4]=n_4$	
$N_{6,33}$	$[n_2n_3]=n_1$,		
	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,	$[x_2n_1]=n_1$,
	$[x_2n_3]=n_3+n_4$,	$[x_2n_4]=n_4$	
$N_{6,34}^{\alpha}$	$[n_2n_3]=n_1$,	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,
	$[x_1n_3]=n_4$,	$[x_2n_1]=(1+\alpha)n_1$,	
	$[x_2n_2]=\alpha n_2$,	$[x_2n_3]=n_3$,	$[x_2n_4]=n_4$
$N_{6,35}^{\alpha\beta}$	$[n_2n_3]=n_1$,	$[x_1n_2]=n_3$,	$[x_1n_3]=-n_2$,
$\alpha^2+\beta^2\neq 0$	$[x_1n_4]=\alpha n_4$,	$[x_2n_1]=2n_1$,	$[x_2n_2]=n_2$,
	$[x_2n_3]=n_3$,	$[x_2n_4]=\beta n_4$	
$N_{6,36}$	$[n_2n_3]=n_1$,	$[x_1n_2]=n_3$,	$[x_1n_3]=-n_2$,
	$[x_2n_1]=2n_1$,	$[x_2n_2]=n_2$,	$[x_2n_3]=n_3$,
	$[x_2n_4]=n_1+2n_4$		
$N_{6,37}^{\alpha}$	$[n_2n_3]=n_1$,		
	$[x_1n_2]=n_3$,	$[x_1n_3]=-n_2$,	$[x_1n_4]=n_1$,
	$[x_2n_1]=2n_1$,	$[x_2n_2]=n_2+\alpha n_3$,	
	$[x_2n_3]=-\alpha n_2+n_3$,	$[x_2n_4]=2n_4$	

TABLE V. Real solvable Lie algebras of dimension six that contain the non-Abelian nilradical $A_{3,1} \oplus A_1$ and the one-dimensional center.

Name	Nonzero commutation relations		
$N_{6,38}$	$[n_2n_3]=n_1$,	$[x_1n_1]=n_1$,	$[x_1n_2]=n_2$,
	$[x_2n_1]=n_1$,	$[x_2n_3]=n_3$,	$[x_1x_2]=n_4$
$N_{6,39}$	$[n_2n_3]=n_1$,	$[x_1n_2]=n_3$,	$[x_1n_3]=-n_2$,
	$[x_2n_1]=2n_1$,	$[x_2n_2]=n_2$,	$[x_2n_3]=n_3$,
	$[x_1x_2]=n_4$		
$N_{6,40}$	$[n_2n_3]=n_1$,	$[x_1n_2]=n_3$,	$[x_1n_3]=-n_2$,
	$[x_2n_4]=n_4$,	$[x_1x_2]=n_1$	

$$\begin{pmatrix}0&&&\\&1&&\\&&-1&\\&&&0\end{pmatrix},\quad\begin{pmatrix}1&&&1\\&0&&\\&&1&\\&&&1\end{pmatrix},\tag{42}$$

$$\begin{pmatrix}0&&&1\\&1&&\\&&-1&\\&&&0\end{pmatrix},\quad\begin{pmatrix}1&&&\\&\alpha&&\\&&1-\alpha&\\&&&1\end{pmatrix},\tag{43}$$

$$\begin{pmatrix}1&&&\\&1&&\\&&0&\\&&&0\end{pmatrix},\quad\begin{pmatrix}1&&&\\&0&&\\&&1&\\&&1&1\end{pmatrix},\tag{44}$$

$$\begin{pmatrix}1&&&\\&1&&\\&&0&\\&&1&0\end{pmatrix},\quad\begin{pmatrix}1+\alpha&&&\\&\alpha&&\\&&1&\\&&&1\end{pmatrix},\tag{45}$$

$$\begin{pmatrix}0&&&\\&&-1&\\&1&&\\&&&\alpha\end{pmatrix},\quad\begin{pmatrix}2&&&\\&1&&\\&&1&\\&&&\beta\end{pmatrix},\quad \alpha^2+\beta^2\neq 0,\tag{46}$$

$$\begin{pmatrix}0&&&\\&&-1&\\&1&&\\&&&0\end{pmatrix},\quad\begin{pmatrix}2&&&1\\&1&&\\&&1&\\&&&2\end{pmatrix},\tag{47}$$

$$\begin{pmatrix}0&&&1\\&&-1&\\&1&&\\&&&0\end{pmatrix},\quad\begin{pmatrix}2&&&\\&1&-\alpha&\\&\alpha&1&\\&&&2\end{pmatrix},\tag{48}$$

for algebras that contain the center of dimension zero, and

$$\begin{pmatrix}1&&&\\&1&&\\&&0&\\&&&0\end{pmatrix},\quad\begin{pmatrix}1&&&\\&0&&\\&&1&\\&&&0\end{pmatrix},\tag{49}$$

$$\begin{pmatrix}0&&&\\&&-1&\\&1&&\\&&&0\end{pmatrix},\quad\begin{pmatrix}2&&&\\&1&&\\&&1&\\&&&0\end{pmatrix},\tag{50}$$

$$\begin{pmatrix}0&&&\\&&-1&\\&1&&\\&&&0\end{pmatrix},\quad\begin{pmatrix}0&&&\\&0&&\\&&0&\\&&&1\end{pmatrix},\tag{51}$$

for algebras that contain one-dimensional centers. Our results are summarized in Tables IV and V in which the algebras that have nilradical $A_{3,1} \oplus A_1$ and the center of dimension 0 and 1, respectively, are given.

ACKNOWLEDGMENTS

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

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

[1]N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[2]G. M. Mubarakzyanov, Sbornik Aspirantskikh Rabot. Mat. Mech. Phys. (Izdat. Kazan. Univ., Kazan, 1962), p. 103.
[3]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **32**, 114 (1963).
[4]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **34**, 99 (1963).
[5]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **35**, 104 (1963).
[6]K. A. Umlauf, *Über die Zusammensetzung der Endlichen Continuierlichen Transformationsgruppen, Insbesondre der Gruppen vom Range Null* (Breitkopf and Härtel, Leipzig, 1891).
[7]V. V. Morozov, Izv. Vyssh. Uchebn. Zaved. Mat. **5**, 161 (1958).
[8]T. Skjelbred and T. Sund, *On the Classification of Nilpotent Lie Algebras*, preprint, Univ. Oslo Mathematics No. 8 (1977).
[9]P. Turkowski, J. Geom. Phys. **4**, 119 (1987).
[10]P. Turkowski, J. Math. Phys. **29**, 2139 (1988).
[11]P. Turkowski, "Structure of Real Lie Algebras," submitted to J. Math. Phys.
[12]N. Jacobson, Proc. Am. Math. Soc. **6**, 281 (1955).
[13]J. Dixmier and W. G. Lister, Proc. Am. Math. Soc. **8**, 155 (1957).
[14]G. Favre, C. R. Acad. Sci. Paris Ser. A-B **274**, A1338 (1972).
[15]G. M. Mubarakzyanov, Izv. Vyssh. Uchebn. Zaved. Mat. **55**, 95 (1966).
[16]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[17]G. F. Leger, Proc. Am. Math. Soc. **4**, 511 (1953).

Nonexistence and existence of various order integrals for two- and three-dimensional polynomial potentials

Paul W. Cleary
Division of Mathematics and Statistics, Commonwealth Scientific and Industrial Research Organization, Private Bag 10, Clayton, Victoria 3168, Australia

(Received 9 May 1989; accepted for publication 7 January 1990)

The nonexistence of integrals of the motion, which are sixth and fourth degree polynomials in the velocities, was established for a range of polynomial potentials. The first known systematic search for sixth-order integrals was performed for cubic and quartic polynomial potentials. It revealed that there exist no nondegenerate cases with such integrals for either potential. Similarly, there exists no nondegenerate fifth or sixth degree polynomial potentials possessing quartic invariants. A complete list of three-dimensional cubic potentials with quadratic integrals is given. All these integrable three-dimensional potentials can be interpreted as orthogonal superpositions of known integrable two-dimensional potentials possessing quadratic integrals. They correspond to 3 of 11 possible coordinate systems in which three-dimensional potentials separate.

I. INTRODUCTION

In many dynamical problems, it is useful to know if the system is integrable or possesses integrals of the motion independent of the Hamiltonian. Direct methods are used for the explicit calculation of the invariants. This enables the compilation of collections of standard form potentials that possess extra integrals. These can then be referred to when studying a particular system to provide useful information about its dynamics. This problem has been examined for two-dimensional polynomial potentials of degree 4 or less possessing integrals that are quartic or less in the velocities.[1–4] These and other results are summarized in the review article by Hietarinta.[5]

There are only three independent integrable cubic potentials. They are $x^3 + 3xy^2 + \alpha y^3$, $2x^3 + xy^2$, and $16x^3 + 3xy^2$. All other integrable cubic potentials can be obtained from these by combinations of rotations, scalings, and reflection. The first pair have quadratic second integrals while the last one has a quartic integral.

In this paper, we give a range of higher-order results for two-dimensional potentials possessing second integrals and a number of null results concerning the nonexistence of certain order integrals for a range of different polynomial potentials. We also present a complete list of all three-dimensional cubic potentials with one or two quadratic integrals. The integrable three-dimensional cubic potentials were found to be closely connected to the above integrable two-dimensional cubic potentials.

II. INTEGRALS FOR TWO-DIMENSIONAL POLYNOMIAL POTENTIALS

Polynomial potentials arise in many problems, particularly when truncated Taylor series expansions are used to facilitate analytic study. It is therefore useful to examine the integrability of such potentials. We will look for integrable cases of the homogeneous discrete-symmetric polynomial potentials:

$$V_n = \sum_{k=0}^{[n/2]} B_k x^{n-2k} y^{2k}. \tag{1}$$

The calculation of integrals of the motion by direct methods requires a particular form to be chosen for the integral. We look for integrals that are mth order in the velocities. That is, of the form

$$I_m = \sum_{p=0}^{m} \sum_{q=0}^{m-p} f_{pq}(x,y)\dot{x}^p \dot{y}^q, \tag{2}$$

where f_{pq} are arbitrary functions of x and y. Calculations of integrals quartic or less in the velocities ($m \leqslant 4$) have been carried out by various authors[1–5] for quartic and lower-order polynomial potentials ($n \leqslant 4$).

In this paper we give the results of a continuing search for new integrable cases with either higher-order integrals or potentials ($n > 4$ or $m > 4$). All known integrable cases of polynomial potentials of the form (1) have integrals of order 2 or 4 but none of order 3. We consider it more likely that any higher-order integrals will be sixth order rather than fifth and therefore concentrate our search on even-order integrals.

A. Quadratic integrals

The conditions required to be satisfied for the existence of quadratic integrals are given in Dorizzi *et al.*[1] A systematic search revealed that all potentials of the form (1) of any degree, possessing integrals quadratic in the velocities, belonged to one of three classes.

(i) When n is even, $V = (x^2 + y^2)^{n/2} = r^n$. This is axially symmetric and possesses the obvious angular momentum integral $L_z^2 = (y\dot{x} - x\dot{y})^2$.

(ii) The combinatorial potentials

$$U_n = \sum_{k=0}^{[n/2]} 2^{n-2k} C_k^{n-k} x^{n-2k} y^{2k}$$

have integrals

$$I_n = \dot{y}(y\dot{x} - x\dot{y}) + y^2 U_{n-1}.$$

 0022-2488/90/061351-05$03.00

This class of potentials was found by Ramani *et al.*[6]

(iii) The potentials

$$W_n = \sum_{k=0}^{[n/2]} C^n_{n-2k}\, x^{n-2k} y^{2k}$$

have integrals

$$J_n = n\dot{x}\dot{y} + (x+y)^n - W_n.$$

These potentials are separable in coordinates $u = x + y$ and $v = x - y$, giving $W_n = \alpha(u^n + v^n)$. The first two nontrivial members W_3 and W_4 were found by Aizawa and Saitô[7] and Bountis *et al.*[8]

The three classes of potentials given above are the only homogeneous discrete-symmetric polynomial potentials of any degree to possess quadratic second integrals independent of the Hamiltonian. For any even degree potential there exist three integrable cases with quadratic integrals. For any odd degree potential there are only two.

B. Quartic integrals

There exist a number of integrable cubic V_3 and quartic V_4 potentials with quartic integrals.[5,6] The next two general homogeneous potentials V_5 and V_6 were examined for the existence of integrable cases with quartic integrals. The required conditions for the existence of such an integral are given in Grammaticos *et al.*[2]

A complete systematic search, using the computer algebra package REDUCE, was performed. For this order integral there are two sets of compatibility relations. Both produce large numbers of equations that need to be satisfied by the coefficients in the potential and those in the integral. Each such equation can usually be solved, leading to a situation where either some condition A is true or some other condition B is true. The resulting multiply branched set of possible solutions has a tree structure, every branch of which must be explored for a possible solution.

After systematically examining all possible solutions of the equations produced by the compatibilty relations, we report that there exist no nondegenerate integrable quintic or sixtic polynomial potentials. All the degenerate subcases, where the integral is the square of well-known quadratic integral, were recovered.

C. Sixtic integrals

The conditions that need to be satisfied to ensure the existence of a sixth-order integral are given in Appendix A. The degree of complexity involved in finding higher-order integrals increases exponentially with the order m of the integral. This case involves solving a complex hierarchy of four successive tiers of coupled partial differential equations.

Any solution must satisfy three different sets of compatibility relations. These lead to an even more complex tree structure than the one for the fourth-order integral case. Again the search was performed on a computer using REDUCE. The coefficient functions in (A3) were used in their full generality. A partial study, by fixing the degree of these functions or using lower degree polynomials, would not be particularly revealing. To our knowledge this is the first brute force search to be carried out for sixth-order integrals for any nonlinear system.

The complete systematic search for sixth-order integrals was carried out for the general cubic and quartic polynomial potentials V_3 and V_4. We report that no such integrals exist for either class of potential aside from the known lower-order results, all of which were again recovered as degenerate subcases.

III. INTEGRALS FOR THREE-DIMENSIONAL CUBIC POLYNOMIAL POTENTIALS

The simplest nontrivial three-dimensional potential is a cubic polynomial potential that is discrete symmetric in z. Most of the rotational degrees of freedom are removed from the system by rotating the potential so that the coefficient λ_4 of the x^2y term is zero. The potential is scaled so that the coefficient of x^3 is unity. The reduced potential is then of the form

$$V = x^3 + \lambda_1 xz^2 + \lambda_2 y^3 + \lambda_3 yz^2 + \lambda_5 xy^2, \qquad (3)$$

where the coefficients λ_i are all real. By choosing $\lambda_4 = 0$ we eliminate dozens of potentials that are merely rotations of the more fundamental potentials which remain. Since the potential is homogeneous it cannot be simplified by translation.

The conditions that need to be satisfied for a three-dimensional potential to possess an additional integral quadratic in the velocities are given in Appendix B. We find that there exist only five distinct classes of potentials of the form (3) with one or two additional quadratic invariants.

(i) The potential

$$V = x^3 + 3xy^2 + \lambda_1 xz^2 + \lambda_3 yz^2 + [(\lambda_3^2 - \lambda_1^2)/\lambda_1\lambda_3]\, y^3, \qquad (4)$$

has one integral independent of the Hamiltonian

$$I_1 = \lambda_3(\lambda_3\dot{x} - \lambda_1\dot{y})^2 + 2(\lambda_3 x - \lambda_1 y)^3. \qquad (5)$$

This potential can be reduced to the separable form

$$V = \sqrt{\lambda_1^2 + \lambda_3^2}\,[(1/\lambda_1)u^3 + (1/\lambda_3)v^3 + uz^2], \qquad (6)$$

by the rotation

$$u = (1/\sqrt{\lambda_1^2 + \lambda_3^2})(\lambda_1 x + \lambda_3 y),$$

$$v = (1/\sqrt{\lambda_1^2 + \lambda_3^2})(\lambda_3 x - \lambda_1 y).$$

The integral becomes $I_1 = \lambda_3\dot{v}^2 + 2\sqrt{\lambda_1^2 + \lambda_3^2}\,v^3$.

(ii) The subcase of (4) with $\lambda_1 = 3$,

$$V = x^3 + 3xy^2 + 3xz^2 + \lambda_3 yz^2 + [(\lambda_3^2 - 9)/3\lambda_3]\, y^3, \qquad (7)$$

is integrable, conserving the integral

$$I_2 = 3\lambda_3\dot{y}\dot{z} + 9\dot{x}\dot{z} + 27x^2z + 18\lambda_3 xyz + 3\lambda_3^2 y^2 z + (\lambda_3^2 + 9)z^3, \qquad (8)$$

in addition to the case (i) integral I_1 in (5). After separating out the v dependence from (6) the remainder of the potential corresponds to the two-dimensional potential $\frac{1}{3}u^3 + uz^2$, which is already known to be separable in coordinates $u \pm z$. Both the $y = 0$ and $z = 0$ projections of the potential (7) belong to the class of integrable two-dimensional potentials $\xi^3 + 3\xi\eta^2 + \alpha\eta^3$ (Ref. 8). The three-dimensional potential (7) can therefore be regarded as an orthogonal superposition of two different copies of this two-dimensional potential.

(iii) The subcase of potential (i) with $\lambda_1 = 1/2$

$$V = x^3 + 3xy^2 + \tfrac{1}{2}xz^2 + \lambda_3 yz^2 + (2\lambda_3 - 1/2\lambda_3)\,y^3, \tag{9}$$

is also integrable conserving the integral

$$I_2 = 8z\dot{z}(\dot{x} + 2\lambda_3\dot{y}) - 8\dot{z}^2(x + 2\lambda_3 y) + (4\lambda_3^2 + 1)z^4 + 4z^2(x + 2\lambda_3 y)^2, \tag{10}$$

in addition to the case (i) integral I_1 in (5). After separating out the v dependence in (6) the remaining part of this potential corresponds to the well-known integrable two-dimensional potential $2u^3 + uz^2$, which separates in parabolic coordinates. The $y = 0$ projection of the potential (9) is the integrable two-dimensional potential $2\xi^3 + \xi\eta^2$ and the $z = 0$ projection belongs to the class of integrable two-dimensional potentials $\xi^3 + 3\xi\eta^2 + \alpha\eta^3$. The three-dimensional potential (9) can again be regarded as a superposition of these two integrable two-dimensional potentials.

(iv) The potentials with $\lambda_1 = \lambda_5$ and $\lambda_2 = \lambda_3 = 0$ are of the form

$$V = x^3 + \lambda_1 x(y^2 + z^2) \tag{11}$$

and are axisymmetric. One coordinate therefore separates out in cylindrical polar coordinates and the angular momentum integral $L_x = y\dot{z} - z\dot{y}$ is conserved. One subcase of this potential possesses another quadratic integral.

(v) This integrable potential, with $\lambda_1 = 1/2$, was first found by Grammaticos *et al.*[9] by setting $\eta^2 = y^2 + z^2$ in the corresponding two-dimensional potential $2x^3 + x\eta^2$. The third invariant is

$$I_3 = y\dot{x}\dot{y} + z\dot{x}\dot{z} - x(\dot{y}^2 + \dot{z}^2) + (y^2 + z^2)(4x^2 + y^2 + z^2). \tag{12}$$

The above collection is the complete set of all real potentials of the form (3) possessing one or two quadratic integrals independent of the Hamiltonian. Makarov *et al.*,[10] using a quantum mechanics formalism, found that after suitable rotations all three-dimensional potentials possessing two additional integrals separated in 1 of 11 different coordinate systems. Three of the cases given above, (ii), (iii), and (v), correspond to rotations of systems that separate in rectangular, parabolic cylindrical, and parabolic rotational coordinates, respectively. There are no integrable potentials of the form (3) that separate in any of the other eight coordinate systems.

Furthermore, we find that an integral exists only when one or both of the $y = 0$ or $z = 0$ projections of the potential has the form of one of the two known integrable two-dimensional potentials with quadratic second integrals, $\xi^3 + 3\xi\eta^2 + \alpha\eta^3$ and $2\xi^3 + \xi\eta^2$, or is axisymmetric. The integrable three-dimensional potentials can then be regarded as orthogonal superpositions of these integrable two-dimensional potentials. Not all superpositions, however, are integrable. There are four possible combinations of the above two-dimensional potentials. The fourth one $x^3 + xy^2/2 + 3xz^2 + \lambda_2 y^3 + \lambda_3 yz^2$ conserves no integrals for any values of λ_2 or λ_3 except for the Hamiltonian. It also does not separate in any of the 11 coordinate systems of Makarov *et al.*[10] Dorizzi *et al.*[11] used the reverse approach of superimposing integrable two-dimensional even quartic potentials, in conjunction with singularity analysis, to obtain a new integrable three-dimensional even quartic potential.

The only other integrable two-dimensional cubic potential $16\xi^3 + 3\xi\eta^2$ has a quartic second integral. That only superpositions of integrable two-dimensional potentials appear to give integrable three-dimensional potentials suggests that such potentials as

$$x^3 + 3xy^2 + \tfrac{3}{16}xz^2 + \lambda_2 y^3 + \lambda_3 yz^2,$$

$$x^3 + \tfrac{1}{2}xy^2 + \tfrac{3}{16}xz^2 + \lambda_2 y^3 + \lambda_3 yz^2,$$

$$x^3 + \tfrac{3}{16}xy^2 + 3xz^2 + \lambda_2 y^3 + \lambda_3 yz^2,$$

$$x^3 + \tfrac{3}{16}xy^2 + \tfrac{1}{2}xz^2 + \lambda_2 y^3 + \lambda_3 yz^2,$$

and

$$x^3 + \tfrac{3}{16}xy^2 + \tfrac{3}{16}xz^2 + \lambda_2 y^3 + \lambda_3 yz^2,$$

may be the only candidates for three-dimensional cubic potentials with quartic integrals.

Consider the nonhomogeneous three-dimensional cubic potential

$$V = \tfrac{1}{2}(x^2 + \sigma^2 y^2 + \omega^2 z^2) - \alpha(2x^3 + xy^2 + xz^2), \tag{13}$$

where α, σ^2, and ω^2 are all arbitrary. The cubic terms belong to case (v). When $\omega^2 = \sigma^2$ the potential is axisymmetric and the angular momentum integral $L_x = y\dot{z} - z\dot{y}$ is conserved. By setting $\eta^2 = y^2 + z^2$ in the two-dimensional integrable Hénon–Heiles potential

$$V = \tfrac{1}{2}(x^2 + \omega^2\eta^2) - \alpha(2x^3 + x\eta^2),$$

Grammaticos *et al.*[9] were able to show that this axisymmetric subcase of (13) was integrable. They obtained the second integral by substituting the expression for η^2 into the two-dimensional integral given by Chang *et al.*[12]

If we remove the axisymmetric requirement by letting $\omega^2 \neq \sigma^2$ then the more general second integral

$$I_2 = (4\sigma^2 - 1)[\dot{y}^2 + \sigma^2 y^2] + (4\omega^2 - 1)[\dot{z}^2 + \omega^2 z^2] + \alpha[\alpha(y^2 + z^2)(4x^2 + y^2 + z^2) - 4y(\dot{x}\dot{y} + \sigma^2 xy) - 4z(\dot{x}\dot{z} + \omega^2 xz) + 4x(\dot{y}^2 + \dot{z}^2)]$$

is conserved for the unrestricted potential (13). This potential is not integrable since L_x is no longer conserved. However, it is interesting that the second integral I_2 exists in more general circumstances than the axisymmetric one that was originally used to generate it.

Finally, consider the potential

$$V = \tfrac{1}{2}(x^2 + y^2 + \tfrac{1}{4}z^2) - \left(\frac{A_1}{3}x^3 + \frac{A_3}{3}y^3 + A_5 x^2 y + A_6 xy^2\right),$$

satisfying the conditions

$$\frac{A_1}{A_5} = \frac{A_5}{A_6} = \frac{A_6}{A_3}.$$

It can be reduced to $V = \alpha\xi^3 + (\xi^2 + \eta^2) + \tfrac{1}{4}z^2$ by rotation, where $\alpha = -2(A_5^2 + A_6^2)^{3/2}/(3A_5A_6)$, and therefore has the three independent energies as integrals

$$K_1 = \dot\eta^2 + \eta^2, \quad K_2 = \dot\xi^2 + \xi^2 + \alpha\xi^3, \quad K_3 = \dot z^2 + \tfrac{1}{4}z^2.$$

The above potential also has a fourth independent integral

$$K_4 = -(A_6x - A_5y)[\dot z^2 - \tfrac{1}{4}z^2] + z\dot z(A_6\dot x - A_5\dot y). \tag{14}$$

Solving the third equation of motion gives $z = \sqrt{2K_2}\cos\frac{1}{2}(t+\alpha)$. The first two equations of motion when combined yield the solution

$$w = A_6x - A_5y = B\sin(t+\beta).$$

When the coefficient of z^2 is 1/4, the x and y motions are locked together in the above way. The fourth integral can then be written as

$$K_4 = K_3(B/2)\sin(\alpha - \beta)$$

and depends only upon the phase difference $\alpha - \beta$, the amplitude B of the harmonic motion of w and the decoupled oscillator energy K_3. For any value of the coefficient of z^2 other than 1/4, the right-hand side of (14) is an explicit function of t and consequently K_4 is not constant. Only when the harmonic oscillations of $A_6x - A_5y$ and z are locked together in a 2:1 resonance is K_4 conserved. The first three integrals all commute pairwise. Here, K_4 commutes with K_2 but not with either K_1 or K_3. So there are three independent commuting invariants.

IV. CONCLUSION

A complete list of all two-dimensional discrete-symmetric polynomial potentials of any degree possessing quadratic integrals is given. They all belong to three distinct previously known classes. There are no other such potentials with quadratic integrals.

The nonexistence of any cubic or quartic polynomial potentials with sixth-order integrals was established by the first known systematic search for such integrals. All previously known results for lower-order integrals were recovered as degenerate subcases. A similar search of fifth and sixth degree polynomial potentials showed that there exist no nondegenerate cases of either potential possessing quartic integrals.

The existence of quadratic integrals for three-dimensional discrete-symmetric cubic polynomial potentials was examined. Five classes of potentials were identified. This list is complete. They are either axisymmetric or can be considered as orthogonal superpositions of known integrable two-dimensional potentials with quadratic second integrals. This behavior suggests that there will at most be five and probably fewer classes of three-dimensional cubic potentials with quartic integrals.

APPENDIX A: SIXTH-ORDER INTEGRALS FOR TWO-DIMENSIONAL POTENTIALS

Direct methods require specific forms for the integrals to be chosen. We will examine integrals of the motion that are sixth-order polynomials in the velocities. In two dimensions they have the form

$$\begin{aligned} K &= a_1\dot x^6 + a_5\dot x^5\dot y + a_2\dot x^4\dot y^2 + a_6\dot x^3\dot y^3 + a_3\dot x^2\dot y^4 + a_7\dot x\dot y^5 \\ &\quad + a_4\dot y^6 + b_1\dot x^4 + b_4\dot x^3\dot y + b_2\dot x^2\dot y^2 + b_5\dot x\dot y^3 + b_3\dot y^4 \\ &\quad + c_1\dot x^2 + c_3\dot x\dot y + c_2\dot y^2 + h, \end{aligned} \tag{A1}$$

where $(a_i, i = 1,...,7)$, $(b_i, i = 1,...,5)$, c_1, c_2, c_3, and h are all arbitrary functions of x and y. Requiring that its time derivative dK/dt vanishes and equating the coefficients of all the velocity terms to zero gives a hierarchy of four tiers of coupled PDEs. We denote the partial derivatives of the a_i's, b_i's, and c_i's with respect to x and y by the subscripts x and y, respectively. Here, V_x and V_y are the x and y partial derivatives of the potential V. The first tier consists of eight coupled linear PDEs:

$$\begin{aligned} &a_{1,x} = 0, \quad a_{2,x} + a_{5,y} = 0, \quad a_{3,x} + a_{6,y} = 0, \\ &a_{5,x} + a_{1,y} = 0, \quad a_{6,x} + a_{2,y} = 0, \\ &a_{7,x} + a_{3,y} = 0, \quad a_{4,x} + a_{7,y} = 0, \quad a_{4,y} = 0, \end{aligned} \tag{A2}$$

with solutions

$$\begin{aligned} a_1 &= \alpha_0 + \alpha_1 y + \alpha_2 y^2 + \alpha_3 y^3 + \alpha_4 y^4 + \alpha_5 y^5 + \alpha_6 y^6, \\ a_2 &= \alpha_2x^2 + 3\alpha_3x^2y + 6\alpha_4x^2y^2 + 10\alpha_5x^2y^3 + 15\alpha_6x^2y^4 \\ &\quad + \alpha_8x + 2\alpha_9xy + 3\alpha_{10}xy^2 + 4\alpha_{11}xy^3 + 5\alpha_{12}xy^4 \\ &\quad + \alpha_{13} + \alpha_{14}y + \alpha_{15}y^2 + \alpha_{16}y^3 + \alpha_{17}y^4, \\ a_3 &= \alpha_4x^4 + 5\alpha_5x^4y + 15\alpha_6x^4y^2 + \alpha_{10}x^3 + 4\alpha_{11}x^3y \\ &\quad + 10\alpha_{12}x^3y^2 + \alpha_{15}x^2 + 3\alpha_{16}x^2y + 6\alpha_{17}x^2y^2 + \alpha_{19}x \\ &\quad + 2\alpha_{20}xy + 3\alpha_{21}xy^2 + \alpha_{22} + \alpha_{23}y + \alpha_{24}y^2, \\ a_4 &= \alpha_{27} + \alpha_{26}x + \alpha_{24}x^2 + \alpha_{21}x^3 + \alpha_{17}x^4 \\ &\quad + \alpha_{12}x^5 + \alpha_6x^6, \\ a_5 &= -\alpha_1x - 2\alpha_2xy - 3\alpha_3xy^2 - 4\alpha_4xy^3 \\ &\quad - 5\alpha_5xy^4 - 6\alpha_6xy^5 - \alpha_7 - \alpha_8y - \alpha_9y^2 - \alpha_{10}y^3 \\ &\quad - \alpha_{11}y^4 - \alpha_{12}y^5, \\ a_6 &= -\alpha_3x^3 - 4\alpha_4x^3y - 10\alpha_5x^3y^2 - 20\alpha_6x^3y^3 - \alpha_9x^2 \\ &\quad - 3\alpha_{10}x^2y - 6\alpha_{11}x^2y^2 - 10\alpha_{12}x^2y^3 - \alpha_{14}x - 2\alpha_{15}xy \\ &\quad - 3\alpha_{16}xy^2 - 4\alpha_{17}xy^3 - \alpha_{18} - \alpha_{19}y - \alpha_{20}y^2 - \alpha_{21}y^3, \\ a_7 &= -\alpha_5x^5 - 6\alpha_6x^5y - \alpha_{11}x^4 - 5\alpha_{12}x^4y - \alpha_{16}x^3 \\ &\quad - 4\alpha_{17}x^3y - \alpha_{20}x^2 - 3\alpha_{21}x^2y - \alpha_{23}x \\ &\quad - 2\alpha_{24}xy - \alpha_{25} - \alpha_{26}y. \end{aligned} \tag{A3}$$

The second tier consists of six coupled PDE's:

$$\begin{aligned} &b_{1,x} = 6a_1V_x + a_5V_y, \quad b_{4,x} + b_{1,y} = 5a_5V_x + 2a_2V_y, \\ &b_{2,x} + b_{4,y} = 4a_2V_x + 3a_6V_y, \\ &b_{5,x} + b_{2,y} = 3a_6V_x + 4a_3V_y, \\ &b_{3,x} + b_{5,y} = 2a_3V_x + 5a_7V_y, \quad b_{3,y} = a_7V_x + 6a_4V_y. \end{aligned} \tag{A4}$$

These lead to the first compatibility relation

$$\begin{aligned} &(a_7V_x + 6a_4V_y)_{xxxxx} - (2a_3V_x + 5a_7V_y)_{xxxxy} \\ &\quad + (3a_6V_x + 4a_3V_y)_{xxxyy} - (4a_2V_x + 3a_6V_y)_{xxyyy} \\ &\quad + (5a_5V_x + 2a_2V_y)_{xyyyy} \\ &\quad - (6a_1V_x + a_5V_y)_{yyyyy} = 0. \end{aligned} \tag{A5}$$

This must be expanded and all the coefficients of $x^m\,y^n$ terms equated to zero, leading to usually about 30–100 relations between the α_i's and the coefficients in the potential. The third tier of four coupled PDEs is

$$c_{1,x} = 4b_1V_x + b_4V_y, \quad c_{2,x} + c_{3,y} = 2b_2V_x + 3b_5V_y,$$
$$c_{3,x} + c_{1,y} = 3b_4V_x + 2b_2V_y, \quad c_{2,y} = b_5V_x + 4b_3V_y. \tag{A6}$$

They lead to the second compatibility relation

$$(b_5V_x + 4b_3V_y)_{xxx} - (2b_2V_x + 3b_5)_{xxy} + (3b_4V_x + 2b_2V_y)_{xyy} - (4b_1V_x + b_4V_y)_{yyy} = 0. \tag{A7}$$

The fourth tier of two coupled PDEs is

$$h_x = 2c_1V_x + c_3V_y, \quad h_y = c_3V_x + 2c_2V_y, \tag{A8}$$

leading to the third compatibility relation

$$(2c_1V_x + c_1V_y)_y - (c_3V_x + 2c_2V_y)_x = 0. \tag{A9}$$

Note that the solutions from the previous tier of PDEs feed back into the rhs's of the next tier of PDEs at each step. This makes the solving of these consecutive groups of equations extremely difficult and tedious. If the potential satisfies all the conditions arising from the expansions of the compatibility relations (A5), (A7), and (A9) then the potential is integrable. The integral is then found by consecutively solving all the remaining equations (A4), (A6), and (A8).

APPENDIX B: QUADRATIC INTEGRALS FOR THREE-DIMENSIONAL POTENTIALS

Direct methods require specific forms for the integrals to be chosen. We will examine integrals of the motion that are quadratic polynomials in the velocities. In three dimensions they have the form

$$K = g_0\dot{x}^2 + g_1\dot{y}^2 + g_2\dot{z}^2 + f_1\dot{x}\dot{y} + f_2\dot{x}\dot{z} + f_3\dot{y}\dot{z} + h, \tag{B1}$$

where $g_0, g_1, g_2, f_1, f_2, f_3$, and h are all functions of x, y, and z. Requiring that its time derivative dK/dt vanishes and equating the coefficients of all the velocity terms to zero gives 13 coupled PDEs. The first ten are

$$\frac{\partial g_0}{\partial x} = 0, \quad \frac{\partial g_1}{\partial y} = 0, \quad \frac{\partial g_2}{\partial z} = 0, \quad \frac{\partial f_3}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_1}{\partial z} = 0,$$
$$\frac{\partial g_1}{\partial x} + \frac{\partial f_1}{\partial y} = 0, \quad \frac{\partial g_2}{\partial x} + \frac{\partial f_2}{\partial z} = 0, \quad \frac{\partial f_1}{\partial x} + \frac{\partial g_0}{\partial y} = 0, \tag{B2}$$
$$\frac{\partial f_2}{\partial x} + \frac{\partial g_0}{\partial z} = 0, \quad \frac{\partial g_2}{\partial y} + \frac{\partial f_3}{\partial z} = 0, \quad \frac{\partial f_3}{\partial y} + \frac{\partial g_1}{\partial z} = 0,$$

and have solutions

$$g_0 = \alpha_1 y^2 + \alpha_2 y + \alpha_3 z^2 + \alpha_4 z + \alpha_5 yz + \alpha_6,$$
$$g_1 = \alpha_1 x^2 + \beta_1 x + \beta_2 z^2 + \beta_3 z + \beta_4 xz + \beta_5,$$
$$g_2 = \beta_2 y^2 + \beta_6 y + \alpha_3 x^2 + \beta_7 x + \beta_8 xy + \beta_9,$$
$$f_1 = -2\alpha_1 xy - \alpha_5 xz - \beta_4 yz - \alpha_2 x - \beta_1 y + \beta_8 z^2 + (\gamma_1 + \gamma_2)z - \gamma_3,$$
$$f_2 = -\alpha_5 xy - 2\alpha_3 xz - \beta_8 yz - \alpha_4 x - \beta_7 z + \beta_4 y^2 - \gamma_2 y - \gamma_4,$$
$$f_3 = -\beta_4 xy - \beta_8 xz - 2\beta_2 yz - \beta_3 y - \beta_6 z + \alpha_5 x^2 - \gamma_1 x - \gamma_5. \tag{B3}$$

The remaining three PDEs are

$$h_x = 2g_0V_x + f_1V_y + f_2V_z,$$
$$h_y = f_1V_x + 2g_1V_y + f_3V_z, \tag{B4}$$
$$h_z = f_2V_x + f_3V_y + 2g_2V_z,$$

where the subscripts x, y, and z denote partial derivatives with respect to x, y, and z, respectively. These equations lead to the following three compatability relations

$$f_1(V_{xx} - V_{yy}) + 2(g_1 - g_0)Vxy + f_3V_{xz} - f_2V_{yz} + \left(\frac{\partial f_1}{\partial x} - 2\frac{\partial g_0}{\partial y}\right)V_x + \left(2\frac{\partial g_1}{\partial x} - \frac{\partial f_1}{\partial y}\right)V_y + \left(\frac{\partial f_3}{\partial x} - \frac{\partial f_2}{\partial y}\right)V_z = 0,$$
$$f_3(V_{yy} - V_{zz}) + f_2V_{xy} - f_1V_{xz} + 2(g_2 - g_1)V_{yz} + \left(\frac{\partial f_2}{\partial y} - \frac{\partial f_1}{\partial z}\right)V_x + \left(\frac{\partial f_3}{\partial y} - 2\frac{\partial g_1}{\partial z}\right)V_y + \left(2\frac{\partial g_2}{\partial y} - \frac{\partial f_3}{\partial z}\right)V_z = 0, \tag{B5}$$
$$f_2(V_{xx} - V_{zz}) + f_3V_{xy} + 2(g_2 - g_0)Vxz - f_1V_{yz} + \left(\frac{\partial f_2}{\partial x} - 2\frac{\partial g_0}{\partial z}\right)V_x + \left(\frac{\partial f_3}{\partial x} - \frac{\partial f_1}{\partial z}\right)V_y + \left(2\frac{\partial g_2}{\partial x} - \frac{\partial f_2}{\partial z}\right)V_z = 0.$$

If the potential V satisfies the conditions in (B5) then other integrals aside from the Hamiltonian exist and the equations (B4) can be solved for $h(x, y, z)$, giving explicit expressions for the integrals.

[1]B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **24**, 2282 (1983).

[2]B. Grammaticos, B. Dorizzi, and A. Ramani, J. Math. Phys. **24**, 2289 (1983).

[3]L. Hall, Physica D **8**, 90 (1983).

[4]J. Hietarinta, Phys. Lett. A **96**, 273 (1983).

[5]J. Hietarinta, Physics Rep. **147**, 87 (1987).

[6]A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 1539 (1982).

[7]Y. Aizawa and N. Saitô, J. Phys. Soc. Jpn. **32**, 1636 (1972).

[8]T. Bountis, H. Segur, and F. Vivaldi, Phys. Rev. A **25**, 1257 (1982).

[9]B. Grammaticos, B. Dorizzi, and A. Ramani, Phys. Lett. A **109**, 81 (1985).

[10]A. A. Makarov, J. A. Smorodinsky, Kh. Valiev, and P. Winternitz, Nuovo Cimento A **LII**, 1061 (1967).

[11]B. Dorizzi, B. Grammaticos, J. Hietarinta, A. Ramani, and F. Schwarz, Phys. Lett. A **116**, 432 (1986).

[12]Y. F. Chang, M. Tabor, and J. Weiss, J. Math. Phys. **23**, 531 (1982).

Dispersion relations for causal Green's functions: Derivations using the Poincaré–Bertrand theorem and its generalizations

K. T. R. Davies
Physics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6373
R. W. Davies
GTE Laboratories, Inc., 40 Sylvan Road, Waltham, Massachusetts 02254
G. D. White
Department of Physics, Northwestern State University, Natchitoches, Louisiana 71457 and Physics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6373

(Received 22 August 1989; accepted for publication 24 January 1990)

A famous theorem by Poincaré and Bertrand formally describes how to interchange the order of integration in a double integral involving two principal-value factors. This theorem has important applications in many-body physics, particularly in the evaluation of response functions (or "loop integrals") at either zero or finite temperatures. Of special interest is the loop containing an integration with respect to the energy of two causal propagators. It is shown that such a response function with two boson or two fermion lines behaves statistically like a boson, while the response function containing a boson and a fermion behaves like a fermion. Examples are given of typical loop integrals occurring in the solution of Dyson's equations for nuclear matter in the presence of delta, nucleon, and pion interactions. A "form factor" that is essential for the convergence of the nucleon–pion loop integral is chosen to have little effect on the analogous nucleon–delta loop integral. The Poincaré–Bertrand (PB) theorem is then generalized to multiple integrals and higher-order poles. From the generalization of the theorem to triple integrals, it is shown that causality is rigorously maintained, at zero temperature, for convolutions with respect to the time of products of Green's functions and thus for Dyson's equations. Also, for finite temperature, the three-propagator loop integral satisfies the statistics appropriate for the loop as a whole, in direct analogy with the result for the two-propagator loop. The intimate connection between the PB theorem and analyticity (or causality) is clearly demonstrated. Although this work considers explicitly only nuclear physics examples, the results are relevant to other fields where many-body theory is used.

I. INTRODUCTION

In the theory of dispersion relations, one frequently has to deal with multiple integrals involving products of Schwartz distributions (e.g., delta functions or principal-value terms). In evaluating such integrals, one must be very careful about reversing the order of integration.[1] For example, in a double integral it is permissible to interchange the order of integration for a product involving two delta functions or for a product containing a delta function and a principal-value singularity. However, as a result of an important theorem by Poincaré and Bertrand (PB), in certain products involving two principal-value factors the order of integration cannot be freely switched; in particular, the PB theorem states[1,2] that

$$\int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)} \int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-x)} f(x,y) = \left[\int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)} \frac{\mathscr{P}}{(y-x)} f(x,y)\right] - \pi^2 f(u,u), \tag{1.1}$$

where $\mathscr{P}$ denotes principal value. Note that the last term on the rhs of Eq. (1.1) results from reversing the order of integration.

The most familiar example from many-body theory that involves a double principal value occurs when computing loop corrections (bubble diagrams) to single-particle propagators.[3-8] Generally, one can write the loop as a frequency integration of two casual Green's functions (Feynman propagators)

$$L_{AB}(\omega) = -i\int_{-\infty}^{\infty} d\omega' \, G_A(\omega') G_B(\omega+\omega'). \tag{1.2}$$

The Feynman propagators are used because they are directly related to the quantum mechanical observables. Frequently, other propagators such as the retarded Green's function are used when calculating particular classical quantities (potentials, for example), which are then used in a quantum or relativistic theory. The correct propagator for quantum fields (requiring holes traveling backward in time) is the Feynman propagator.[9] However, as discussed in Sec. II, the formalism developed here is quite general and also applies to advanced and retarded Green's functions.

 0022-2488/90/061356-18$03.00

Since "free" propagators are usually proportional to functions of the form

$$f^{(\pm)}(\omega) = \frac{1}{\omega^2 - \alpha^2 \pm i\delta} \underset{\delta\to 0+}{\rightarrow} \frac{\mathscr{P}}{(\omega^2 - \alpha^2)} \mp i\pi\delta(\omega^2 - \alpha^2), \quad (1.3)$$

it is clear that the PB theorem is relevant for computing first-order loops (i.e., loops containing only free propagators). However, even for such simple loops, one must be very careful about introducing multiple principal-value factors. In particular, the limit indicated in Eq. (1.3) is not valid if another principal-value term is already present in the integral. In Appendix C, we show how the PB theorem allows one to generalize Eq. (1.3) to first-order loops containing two free propagators. Fortunately, all of this complication is not really necessary for first-order loops since such integrals can be very simple evaluated by other methods (usually complex contour integration).[5,6,8] The most significant application of the PB theorem involves the more general case, where the dressed and/or finite-temperature propagators no longer have the simple form of Eq. (1.3), but from the spectral representation, the real and imaginary parts satisfy a causal dispersion relation.[5] [See, e.g., Eq. (2.5a).]

In this case, the PB theorem can be used to show that $L_{AB}(\omega)$ also satisfies a similar dispersion relation. In Ref. 3, this very important causality relation is implemented in the development, at zero temperature, of a self-consistent nuclear transport theory. Here we concentrate on various generalizations such as the finite-temperature dispersion relation, the case of multiple principal-value integrals, and the case of higher-order poles.

Also, as we shall discuss in subsequent sections, there is an intimate connection between the PB theorem and analyticity. Thus any result obtained using the PB theorem can usually be derived from standard complex variable theory, for which it is especially useful to introduce the spectral representations of the Green's functions.[5] Nevertheless, many of the most important results of this paper are easier to obtain using the PB theorem rather than the complex variable theory. In particular, the statistical properties of finite-temperature loop integrals containing two- and three-particle *causal* propagators follow, in a straightforward way, from the PB theorem. These relations are not as simply derived using complex variable theory, mainly because the spectral representation of a causal Green's function has a rather complicated structure.[5] Moreover, as one tries to evaluate even more sophisticated types of many-body diagrams (e.g., loops containing more than three particles), complex variable manipulations become increasingly unwieldy. In contrast, the advanced and retarded Green's functions have much more tractable spectral representations,[5] thereby allowing one to easily simplify expressions involving these propagators. Again, because we are mainly interested in dispersion relations in quantum field theory, our focus is on *causal* Green's functions; then integrals containing these functions provide a real "showplace" for the applicability of the PB theorem. Thus the PB theorem is seen to be an important tool for formally manipulating complicated many-body diagrams which involve causal propagators.[3] Simple nonrelativistic prototypes of such diagrams are evaluated in Sec. II.

The remainder of this paper is structured as follows. In Sec. II we derive general finite-temperature equations for two-propagator response functions. In this section, we also present some numerical results of calculations in nuclear matter using propagators arising in the solution of Dyson's equations. Then in Sec. III we generalize the PB theorem to the case of more than two integrals. Using the PB theorem involving three integrals, we derive dispersion relations for time-convoluted products of Green's functions, for Dyson's equations, and for the three-propagator loop integral. Next, in Sec. IV we generalize the double-integral PB theorem to the case of higher-order singularities. Appendices A and B contain mathematical details of the derivations contained in Secs. III and IV, respectively, and Appendix C contains a nonrigorous, but intuitive derivation of the PB theorem. Also, as mentioned, Appendix C gives a generalization of Eq. (1.3) to the case in which two free propagators are present in a loop integral.

Finally, we emphasize that while the main thrust of our work pertains to nuclear or particle physics, many of the results should be applicable to other specialties, e.g., condensed matter physics. In particular, the work in Secs. II and III that pertains to finite-temperature Green's functions should be of interest in the general theory of response functions.[5]

II. THE PB THEOREM AND ITS APPLICATION TO FINITE-TEMPERATURE RESPONSE FUNCTIONS

A. General remarks

It is instructive to illustrate the usefulness of Eq. (1.1) with a simple example. Consider two functions related by the Hilbert transform

$$h(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-x)} g(y). \quad (2.1)$$

Next, construct the function

$$\hat{g}(u) = \frac{1}{\pi}\int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)} h(x),$$

which from Eqs. (1.1) and (2.1) becomes

$$\hat{g}(u) = \frac{1}{\pi^2}\int_{-\infty}^{\infty} g(y)dy \int_{-\infty}^{\infty} dx \times \frac{\mathscr{P}}{(x-u)}\frac{\mathscr{P}}{(y-x)} - g(u). \quad (2.2)$$

Since we have

$$\int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-a)}\frac{\mathscr{P}}{(x-b)} = 0, \quad (a \text{ and } b \text{ real}), \quad (2.3)$$

the first term in Eq. (2.2) vanishes and we obtain

$$\hat{g}(u) = -g(u) = \frac{1}{\pi}\int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)} h(x). \quad (2.4)$$

Thus two functions g and h related by Eq. (2.1) must also satisfy the "inverse" relation (2.4).

The well-known results (2.1) and (2.4) for Hilbert transforms[10–12] are usually proved from complex variable theory by letting $x \to z = x + iy$ and requiring that

$\Psi^{(+)}(z)=h(z)+ig(z)$ be an analytic function in the upper-half z plane. The proof of the PB theorem that has been given by Muskhelishvili[2] [hence, also, the reciprocal theorem (2.4) on Hilbert transforms] makes no explicit recourse to complex variable theory. However, by a theorem due to Titchmarsh[13] we know that the existence of the Hilbert transforms (2.1) and (2.4) implies the analyticity and causality of the functions involved. [The causal behavior can be seen by assuming that x is an energy (frequency) variable and Fourier transforming to a time representation.] Another way to view the analyticity of $\Psi^{(+)}(z)$ is to note that it can be expressed as a Cauchy integral:

$$\Psi^{(+)}(z)=\frac{1}{\pi}\int_{-\infty}^{\infty}dx\,\frac{g(x)}{(x-z)}$$

when z is in the upper half-plane or approaches the real axis from above. The above function is manifestly analytic in the upper half-plane. Similarly, the function

$$\Psi^{(-)}(z)=\frac{1}{\pi}\int_{-\infty}^{\infty}dx\,\frac{g(x)}{(x-z)}$$

for z belonging to the lower half-plane or approaching the real axis from below is an analytic function throughout the entire lower half-plane. Thus we see that there is an intimate connection between the PB theorem and analyticity.

B. Finite-temperature response functions

We next use the PB theorem to obtain a dispersion relation for integrals of products of Green's functions.[4,5] Such a relation is useful in evaluating the nuclear response functions occurring, e.g., in pion–nucleon–delta interactions.[3,7,14] We begin with a single finite-temperature, causal propagator satisfying the well-known dispersion relation[5]

$$\mathrm{Re}(G_\alpha(\omega))=\frac{1}{\pi}\int_{-\infty}^{\infty}d\omega'\,\frac{\mathscr{P}}{(\omega'-\omega)}\times\mathrm{Im}(G_\alpha(\omega'))\eta_\alpha(\omega') \tag{2.5a}$$

and

$$\eta_\alpha(\omega)=\begin{cases}\tanh(\omega-\mu_\alpha)/2k_BT, & \text{for bosons,}\\ \coth(\omega-\mu_\alpha)/2k_BT, & \text{for fermions,}\end{cases} \tag{2.6}$$

where k_B is the Boltzmann constant, T is the temperature, and μ_α is the chemical potential. In Eqs. (2.5a) and (2.6), α labels the type of particle (pion, nucleon, delta, etc.). In the zero-temperature limit, we have

$$\eta_\alpha(\omega)\to\mathrm{sgn}(\omega-\mu_\alpha). \tag{2.7}$$

From Eqs. (2.1) and (2.4), we find that

$$\mathrm{Im}(G_\alpha(\omega))\eta_\alpha(\omega)=-\frac{1}{\pi}\int_{-\infty}^{\infty}d\omega'\,\frac{\mathscr{P}}{(\omega'-\omega)}\times\mathrm{Re}[G_\alpha(\omega')]. \tag{2.5b}$$

We note that $G_\alpha(\omega)$ is not an analytic function in either the upper or lower half-ω planes.[5] However, defining

$$g(\omega)=\eta_\alpha(\omega)\mathrm{Im}[G_\alpha(\omega)]\quad(\omega\ \text{real}),$$
$$h(\omega)=\mathrm{Re}[G_\alpha(\omega)]\quad(\omega\ \text{real}),$$

we know from the discussion in Sec. II A[13] that the functions

$$\Psi^{(\pm)}(\omega)=h(\omega)\pm ig(\omega),$$

continued to the complex ω plane, are analytic in the upper (lower) half-planes. Of course, one can always construct the retarded and advanced Green's functions,[15] which are analytic in the upper and lower half-planes, respectively. However, for reasons discussed in Sec. I[9] we prefer to work with the causal Green's function.

The response function is proportional to the integral[5]

$$\Pi_{\alpha\beta}(\omega)=i\int_{-\infty}^{\infty}d\omega'\,G_\alpha(\omega+\omega')G_\beta(\omega') \tag{2.8}$$

where for now we suppress other integrations.

Also, one can Fourier transform to functions of time using

$$G_\alpha(\omega)=(2\pi)^{-1}\int_{-\infty}^{\infty}d\tau\,e^{i\omega\tau}G_\alpha(\tau)$$

and identical equations for G_β and $\Pi_{\alpha\beta}$. Here we assume stationary or steady-state motion with $G_\alpha(t-t')\equiv G_\alpha(t,t')$. We then find that

$$\Pi_{\alpha\beta}(\tau)=iG_\alpha(\tau)G_\beta(-\tau)$$

or

$$\Pi_{\alpha\beta}(t-t')=iG_\alpha(t-t')G_\beta(t'-t),$$

which nicely demonstrates that $\Pi_{\alpha\beta}$ is indeed a "loop" when represented by the Feynman diagram of Fig. 1.

The imaginary part of $\Pi_{\alpha\beta}(\omega)$ is

$$\mathrm{Im}\,\Pi_{\alpha\beta}(\omega)=\int_{-\infty}^{\infty}d\omega'[\mathrm{Re}\,G_\alpha(\omega+\omega')\mathrm{Re}\,G_\beta(\omega')-\mathrm{Im}\,G_\alpha(\omega+\omega')\mathrm{Im}\,G_\beta(\omega')]. \tag{2.9}$$

However, we can evaluate the first term in the integrand of Eq. (2.9) using Eq. (2.5a), giving

$$\begin{aligned}\int_{-\infty}^{\infty}&d\omega'\,\mathrm{Re}[G_\alpha(\omega+\omega')]\mathrm{Re}[G_\beta(\omega')]\\&=\frac{1}{\pi^2}\int_{-\infty}^{\infty}d\omega'\int_{-\infty}^{\infty}d\omega_\alpha\,\frac{\mathscr{P}}{(\omega_\alpha-\omega'-\omega)}\\&\quad\times\mathrm{Im}[G_\alpha(\omega_\alpha)]\eta_\alpha(\omega_\alpha)\int_{-\infty}^{\infty}d\omega_\beta\,\frac{\mathscr{P}}{(\omega_\beta-\omega')}\\&\quad\times\mathrm{Im}[G_\beta(\omega_\beta)]\eta_\beta(\omega_\beta)\\&=\frac{1}{\pi^2}\int_{-\infty}^{\infty}d\omega_\alpha\,\mathrm{Im}[G_\alpha(\omega_\alpha)]\eta_\alpha(\omega_\alpha)\\&\quad\times\int_{-\infty}^{\infty}d\omega'\,\frac{\mathscr{P}}{(\omega_\alpha-\omega'-\omega)}\int_{-\infty}^{\infty}d\omega_\beta\,\frac{\mathscr{P}}{(\omega_\beta-\omega')}\\&\quad\times\mathrm{Im}[G_\beta(\omega_\beta)]\eta_\beta(\omega_\beta),\end{aligned} \tag{2.10}$$

where we have reversed the ω' and ω_α integrations since there is only a single principal-value term.[1] If we apply Eq. (1.1) to the ω' and ω_β integrations in Eq. (2.10), we obtain

$$\int_{-\infty}^{\infty} d\omega' \,\mathrm{Re}[G_\alpha(\omega+\omega')]\mathrm{Re}[G_\beta(\omega')]$$
$$=\frac{1}{\pi^2}\int_{-\infty}^{\infty} d\omega_\alpha \,\mathrm{Im}\, G_\alpha(\omega_\alpha)\eta_\alpha(\omega_\alpha)\int_{-\infty}^{\infty} d\omega_\beta \,\mathrm{Im}\, G_\beta(\omega_\beta)\eta_\beta(\omega_\beta)\int_{-\infty}^{\infty} d\omega' \frac{\mathscr{P}}{(\omega_\alpha-\omega'-\omega)}\frac{\mathscr{P}}{(\omega_\beta-\omega')}$$
$$+\int_{-\infty}^{\infty} d\omega_\alpha \,\mathrm{Im}\, G_\alpha(\omega_\alpha)\eta_\alpha(\omega_\alpha)\mathrm{Im}\, G_\beta(\omega_\alpha-\omega)\eta_\beta(\omega_\alpha-\omega).$$

which from Eq. (2.3) reduces to

$$\int_{-\infty}^{\infty} d\omega' \,\mathrm{Re}[G_\alpha(\omega+\omega')]\mathrm{Re}[G_\beta(\omega')] = \int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta \,\delta(\omega_\alpha-\omega-\omega_\beta)\mathrm{Im}[G_\alpha(\omega_\alpha)] \times\mathrm{Im}[G_\beta(\omega_\beta)]\eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta). \tag{2.11}$$

Then substituting Eq. (2.11) into Eq. (2.9), we find that

$$\mathrm{Im}\,\Pi_{\alpha\beta}(\omega) = \int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta \,\delta(\omega_\alpha-\omega-\omega_\beta) \times[\eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta)-1]\mathrm{Im}[G_\alpha(\omega_\alpha)] \times\mathrm{Im}[G_\beta(\omega_\beta)]. \tag{2.12}$$

In the zero-temperature limit [Eq. (2.7)] we have

$$\mathrm{Im}[\Pi_{\alpha\beta}(\omega)] \underset{T\to 0}{\to} -2\int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta \,\delta(\omega_\alpha-\omega-\omega_\beta) \times[\theta(\omega_\alpha-\mu_\alpha)\theta(\mu_\beta-\omega_\beta) +\theta(\omega_\beta-\mu_\beta)\theta(\mu_\alpha-\omega_\alpha)] \times\mathrm{Im}[G_\alpha(\omega_\alpha)]\mathrm{Im}[G_\beta(\omega_\beta)], \tag{2.13}$$

where

$$\theta(x)=\begin{cases}1, & \text{for } x>0,\\ 0, & \text{for } x<0.\end{cases}$$

In the integrand of Eq. (2.13) the only contributions are from particle–hole terms, not particle–particle or hole–hole terms.[5] (Of course, for finite temperatures, the hole–particle boundaries are smeared, so that this distinction does not apply.)

From Eqs. (2.5a) and (2.8) we then obtain

$$\mathrm{Re}\,\Pi_{\alpha\beta}(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta \frac{\mathscr{P}}{(\omega_\alpha-\omega-\omega_\beta)} \times\mathrm{Im}[G_\alpha(\omega_\alpha)]\mathrm{Im}[G_\beta(\omega_\beta)] \times[\eta_\alpha(\omega_\alpha)-\eta_\beta(\omega_\beta)], \tag{2.14}$$

which has the limit

$$\mathrm{Re}[\Pi_{\alpha\beta}(\omega)] \underset{T\to 0}{\to} -\frac{2}{\pi}\int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta \frac{\mathscr{P}}{(\omega_\alpha-\omega-\omega_\beta)} \times[\theta(\omega_\alpha-\mu_\alpha)\theta(\mu_\beta-\omega_\beta) -\theta(\mu_\alpha-\omega_\alpha)\theta(\omega_\beta-\mu_\beta)] \times\mathrm{Im}[G_\alpha(\omega_\alpha)]\mathrm{Im}[G_\beta(\omega_\beta)]. \tag{2.15}$$

We again note the presence of only particle–hole terms in the integrand as $T\to 0$.

We now construct a dispersion relation for $\Pi_{\alpha\beta}(\omega)$ that is analogous to Eq. (2.5a) for a single Green's function. Thus we need to find a function $\eta_{\alpha\beta}(\omega)$ such that the relation

$$\mathrm{Re}[\Pi_{\alpha\beta}(\omega)] = \frac{1}{\pi}\int_{-\infty}^{\infty} d\omega' \frac{\mathscr{P}}{(\omega'-\omega)} \times\mathrm{Im}[\Pi_{\alpha\beta}(\omega')]\eta_{\alpha\beta}(\omega') \tag{2.16a}$$

is satisfied. From Eqs. (2.12), (2.14), and (2.16a) we find that

$$\eta_{\alpha\beta}(\omega_\alpha-\omega_\beta) = \frac{\eta_\beta(\omega_\beta)-\eta_\alpha(\omega_\alpha)}{\eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta)-1} \tag{2.17}$$

and from Eq. (2.6) it follows that

$$\eta_{\alpha\beta}(\omega) = \begin{cases}\tanh[(\omega-\mu_\alpha+\mu_\beta)/2k_BT], & \text{if } \alpha \text{ and } \beta \text{ are particles with like statistics,}\\ \coth[(\omega-\mu_\alpha+\mu_\beta)/2k_BT], & \text{if } \alpha \text{ and } \beta \text{ are particles with unlike statistics.}\end{cases} \tag{2.18}$$

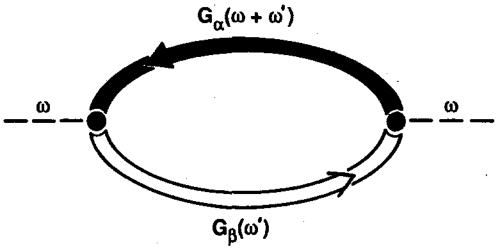


FIG. 1. A Feynman diagram for a two-propagator response function. The dashed line represents the external propagator into which the loop integral ($\times|\mathbf{p}|^2$) is inserted as a "self-energy," e.g., in the solution of Dyson's equation.

The result (2.18) is very satisfying since it demonstrates that statistically two bosons or two fermions behave like a boson, while a boson and a fermion behave like a fermion. Also, we have

$$\eta_{\alpha\beta}(\omega) \underset{T\to 0}{\to} \mathrm{sgn}(\omega-\mu_\alpha+\mu_\beta). \tag{2.19}$$

Notice that Eqs. (2.18) and (2.19) involve only the difference of the chemical potentials. The inverse relation of Eq. (2.16a) can be obtained from Eqs. (2.1) and (2.4), namely,

$$\mathrm{Im}[\Pi_{\alpha\beta}(\omega)]\eta_{\alpha\beta}(\omega) = -\frac{1}{\pi}\int_{-\infty}^{\infty} d\omega' \frac{\mathscr{P}}{(\omega'-\omega)} \times\mathrm{Re}[\Pi_{\alpha\beta}(\omega')]. \tag{2.16b}$$

Again, we know from Titchmarsh's theorem[13] that if we define the functions

$$\chi^{(\pm)}(\omega) = \mathrm{Re}[\Pi_{\alpha\beta}(\omega)] \pm i\eta_{\alpha\beta}(\omega) \times \mathrm{Im}[\Pi_{\alpha\beta}(\omega)] \quad (\text{real } \omega)$$

and then let ω become complex, $\chi^{(\pm)}(\omega)$ are analytic in the upper (lower) half-planes. Also, we observe that in Eqs. (2.14), (2.15), and (2.17)–(2.19) there is an asymmetry between α and β which arises from the asymmetry in the original defining equation (2.8).

The above formalism is very general and applies to propagators other than the causal Green's functions. In particular, for the retarded and advanced Green's functions we have $\eta = +1$ and $\eta = -1$, respectively. Then from Eqs. (2.12) and (2.14), we see that the real and imaginary parts of $\Pi_{\alpha\beta}(\omega)$ vanish if both propagators are advanced or both are retarded. This again is a reflection of the physics, which dictates that only particle–hole combinations are allowed for the response function. However, one could have a nonvanishing retarded–advanced loop and from Eq. (2.17) we find that

$$\eta_{\alpha\beta} = \begin{cases} +1 \text{ for } \alpha \text{ retarded}, & \beta \text{ advanced}, \\ -1 \text{ for } \alpha \text{ advanced}, & \beta \text{ retarded}. \end{cases}$$

C. Calculations in nuclear matter

We now calculate typical response functions in nuclear matter for the case in which we have nucleon–delta–pion interactions.[3,7,14,16,17] The resulting loop integral shown in Fig. 1 is then used as a self-energy insertion into another propagator, as represented by the dashed line. The full lines of the loop can represent a (nucleon, delta) combination, where the dashed, external line is a pion. Similarly, one could have nucleon–pion lines in the loop and a delta particle for the external propagator. Other combinations are, of course, possible and are being considered,[3,14] but the nucleon–delta and nucleon–pion loops are the most important physical processes.[16,17]

Consider now the bare nucleon–delta ($N\Delta$) loop for which we rewrite Eq. (2.8), showing explicitly all the momentum integrations[3]

$$\Pi_{N\Delta}(p) = i\int d^4q\, f^2(2q+p) G_N(p+q) G_\Delta(q), \tag{2.20}$$

where $p \equiv (\omega, \mathbf{p})$ and $q \equiv (\omega', \mathbf{q})$ are four-vectors. A "form factor" f has been introduced in order to assure convergence of the nucleon–pion loop integral,[3] although, as we shall see, its effect on the nucleon–delta loop is minimal. Explicitly, we use the functions[14]

$$G_N(p) = \frac{1 - n_N(\mathbf{p})}{\omega - \epsilon_N(\mathbf{p}) + i\epsilon} + \frac{n_N(\mathbf{p})}{\omega - \epsilon_N(\mathbf{p}) - i\epsilon}, \tag{2.21a}$$

$$G_\Delta(p) = [\omega - \epsilon_\Delta(\mathbf{p}) + \tfrac{1}{2}i\Gamma_\Delta(\omega)]^{-1}, \tag{2.21b}$$

$$f(p) = \beta/(p^4 + \lambda^4), \tag{2.22}$$

where

$$n_N(\mathbf{p}) = \{\exp[(\epsilon_N(p) - \mu_N)/k_B T] + 1\}^{-1} \tag{2.23a}$$

$$\underset{T\to 0}{\to} \theta(\mu_N - \epsilon_N(\mathbf{p})), \tag{2.23b}$$

$$\epsilon_N(\mathbf{p}) = (\hbar^2/2m_N^*)|\mathbf{p}|^2 + m_N + V_N^{(0)}, \tag{2.24}$$

$$\epsilon_\Delta(\mathbf{p}) = (\hbar^2/2m_\Delta^*)|\mathbf{p}|^2 + m_\Delta + V_\Delta^{(0)}, \tag{2.25}$$

and m_N and m_Δ are the free-space masses of the nucleon and delta, respectively. The effective masses m_N^* and m_Δ^*, as well as the potentials $V_N^{(0)}$ and $V_\Delta^{(0)}$, depend upon the nuclear density.[14] The constant β is determined by setting $f = 1.0$ for the on-shell reaction $\Delta \to N + \pi$ in free space, and $\lambda = 1.5 \times m_N$.

Two different forms were assumed for the width $\Gamma_\Delta(\omega)$:

$$\Gamma_\Delta(\omega) = \theta(\omega - \mu_\Delta)\Gamma_\Delta^{(0)}, \tag{2.26}$$

where $\Gamma_\Delta^{(0)} = 115.0$ MeV, the delta width in free space, and

$$\Gamma_\Delta(\omega) = \begin{cases} g_\Delta(\omega), & \text{if } g_\Delta(\omega) \leqslant \Gamma_\Delta^{(0)}, \\ \Gamma_\Delta^{(0)}, & \text{if } g_\Delta(\omega) > \Gamma_\Delta^{(0)}, \end{cases} \tag{2.27}$$

with

$$g_\Delta(\omega) = \frac{\theta(\omega - m_N - m_\pi)(f^*)^2\{2m_N[\omega - (2m_N\omega - m_N^2 + m_\pi^2)^{1/2}]^{3/2}\}m_N}{6\pi m_\pi^2[2m_N\omega - m_N^2 + m_\pi^2]^{1/2}}; \tag{2.28}$$

f^* is the nucleon–delta–pion coupling constant[14,17] and m_π is the free-space mass of the pion. The chemical potential μ_Δ in Eq. (2.26) is the same as that for the nucleon, i.e.,

$$\mu_\Delta = \mu_N = \epsilon_N(k_F), \tag{2.29}$$

where k_F is the Fermi momentum. (For $k_F = 1.333\,\mathrm{fm}^{-1}$, $\mu_\Delta = \mu_N \approx 923$ MeV.) The second form, Eq. (2.27), is obtained by integrating the free-space nucleon–pion self-energy loop and has the correct gradually increasing threshold behavior [in contrast to the step function discontinuity at $\omega = \mu_\Delta$ in Eq. (2.26)].[14] In Eq. (2.27) $\Gamma_\Delta(\omega)$ vanishes for energies below $m_N + m_\pi \approx 1077$ MeV; for energies above this value, it increases monotonically until $\omega \approx 1244$ MeV, after which $\Gamma_\Delta(\omega)$ remains constant at the free-space value $\Gamma_\Delta^{(0)}$. Thus at high energies the two prescriptions are identical.

The full self-energy insertion for the pion propagator is given by $\Sigma(p) = |\mathbf{p}|^2 U_{N\Delta}(p)$:[3,14]

$$U_{N\Delta}(p) = U_{N\Delta}^{(0)}(p)[1 + (8\pi^2/9m_N^* k_F)g'_\Delta U_{N\Delta}^{(0)}(p)]^{-1}, \tag{2.30}$$

where $g'_\Delta = 1.6$ is the Migdal parameter[17] and

$$U_{N\Delta}^{(0)}(p) = \Pi_{N\Delta}(p) + \Pi_{N\Delta}(-p). \tag{2.31}$$

In Eq. (2.20) two different options are used in calculating $\Pi_{N\Delta}(p)$: $f = 1.0$ and f having the cutoff behavior given by Eq. (2.22). However, when $f \neq 1.0$, the PB derivation given in Sec. II B is no longer valid and Eq. (2.16a) is not true. Nevertheless, we impose the following prescription to assure exact causality.[3] First, we calculate the imaginary part of $\Pi_{N\Delta}(p)$ from an equation analogous to Eqs. (2.12) and (2.13), namely

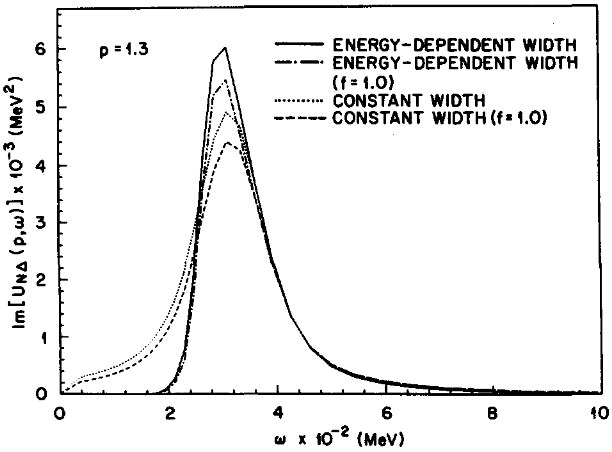


FIG. 2. The bare $\mathrm{Im}[U_{N\Delta}^{(0)}]$ function, calculated from Eqs. (2.31) and (2.33), for $p = 1.3\ \mathrm{fm}^{-1}$ as a function of ω. The form factor is calculated from Eq. (2.22) except for the cases denoted by $f = 1.0$.

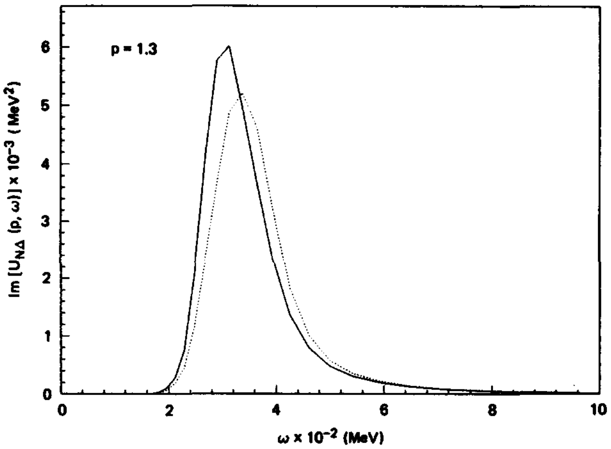


FIG. 4. Comparison of the bare $\mathrm{Im}[U_{N\Delta}^{(0)}]$ function (solid line) with the $\mathrm{Im}[U_{N\Delta}]$ function (dotted line) calculated from Eq. (2.30) for $p = 1.3$ fm^{-1} with $g'_\Delta = 1.6$. We use the energy-dependent width option and $f \neq 1.0$.

$$\mathrm{Im}[\Pi_{N\Delta}(p)] = \int d^4q f^2(2q+p)\mathrm{Im}[G_N(p+q)] \times \mathrm{Im}[G_\Delta(q)][\eta_N(p+q)\eta_\Delta(q) - 1]. \quad (2.32)$$

Then we calculate the real part of $\Pi_{N\Delta}(p)$ from Eq. (2.16a), so that exact causality is recovered. We emphasize that the precise form of Eq. (2.32) is extracted from Eq. (2.12), which was obtained by the derivation using the PB theorem for $f = 1.0$. This prescription, while *ad hoc*, is physically reasonable. It can be shown that Eq. (2.32) reduces to

$$\begin{aligned}\mathrm{Im}[\Pi_N\Delta(\omega,p)] &= 2\pi^2\int_0^\infty q^2dq[2n_N(q) - 1] \\ &\times[\eta_N(\epsilon_N(q))\eta_\Delta(\epsilon_N(q) - \omega) - 1] \\ &\times\int_{-1}^{+1} dx\, \mathrm{Im}[G_\Delta(\epsilon_N(q) - \omega,[q^2+p^2-2qpx]^{1/2})] \\ &\times f^2[2\epsilon_N(q) - \omega, 2q - p]. \quad (2.33)\end{aligned}$$

In Eq. (2.33), note that $q \equiv |\mathbf{q}|$ and $p = |\mathbf{p}|$ are *not* four-vectors, a convention we will maintain for the remainder of this section.

We now present some zero-temperature results. Calculations of $U_{N\Delta}^{(0)}(\omega,p)$ are shown in Figs. 2 and 3 for the constant width prescription (2.26) and the energy-dependent width prescription (2.27). For each of these cases, we also show $f = 1.0$ and f calculated from Eq. (2.22). It is seen that the form factor has little effect on the nucleon–delta loop, which is understandable since it is parametrized to have the value of 1.0 for four-momenta which give the most significant contributions to the integral. Notice, too, that for the energy-dependent width option, the $\mathrm{Im}(U_{N\Delta}^{(0)})$ vanishes for $\omega \lesssim 160.0$ MeV, whereas for the constant width option it goes to zero for small energies only at $\omega = 0$, with a slope given by a well-known phase-space factor.[14] We see, too, that for a given form factor option, the curves from Eqs. (2.26) and (2.27) are identical for large ω. In Figs. 4 and 5 we show comparisons of the bare $U_{N\Delta}^{(0)}$ function, with the full $U_{N\Delta}$

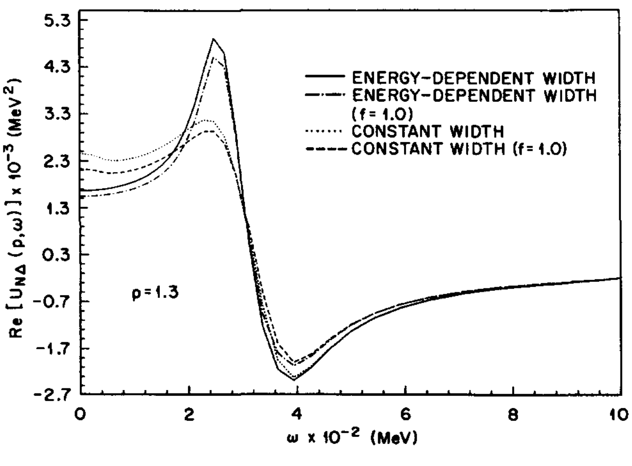


FIG. 3. The bare $\mathrm{Re}[U_{N\Delta}^{(0)}]$ function, calculated using Eqs. (2.16a) and (2.31), for $p = 1.3\ \mathrm{fm}^{-1}$ as a function of ω. The form factor is calculated from Eq. (2.22) except for the cases denoted by $f = 1.0$.

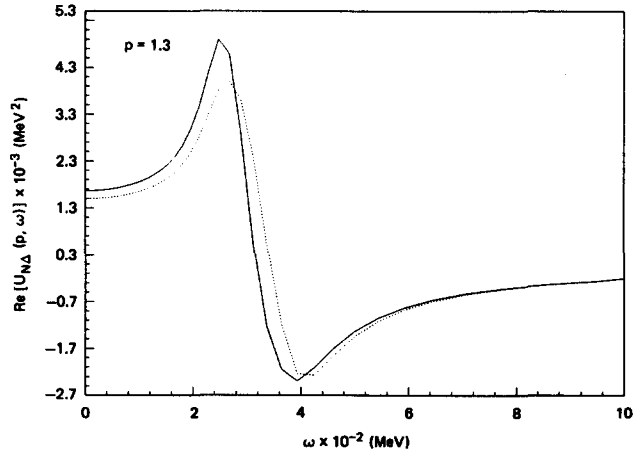


FIG. 5. Comparison of the bare $\mathrm{Re}[U_{N\Delta}^{(0)}]$ function (solid line) with the $\mathrm{Re}[U_{N\Delta}]$ function (dotted line) calculated from Eq. (2.30) for $p = 1.3$ fm^{-1} with $g'_\Delta = 1.6$. We use the energy-dependent width option and $f \neq 1.0$.

function calculated from Eq. (2.30): The main difference between these functions is that the ω value for the peak of the imaginary part of $U_{N\Delta}$ is about 25 MeV higher than that for $U_{N\Delta}^{(0)}$. We see that $U_{N\Delta}$ is somewhat "smoother" that $U_{N\Delta}^{(0)}$, which is expected from the general form of Eq. (2.30).[18] Also, note that the real and imaginary parts of $U_{N\Delta}^{(0)}$ or $U_{N\Delta}$ have the types of shapes one expects for functions which satisfy dispersion relations.[19] Finally, we remark that the functions G_Δ, $\Pi_{N\Delta}$, $U_{N\Delta}^{(0)}$, and $U_{N\Delta}$ have been calculated for only the first iteration of Dyson's equations, using Eqs. (2.20)–(2.28). In the full solution of the equations, one performs a complete self-consistent calculation, iterating until convergence is achieved.[14]

III. GENERALIZATION OF THE PB THEOREM TO MULTIPLE INTEGRALS AND APPLICATIONS

At least two generalizations of the PB theorem exist. One is for triple and higher-order integrals and the other is for higher-order poles. In this section we shall study (i) the generalization to multiple integrals and (ii) the applications of the triple-integral formula to certain nuclear many-body relations and to separable products of sin and cos functions. Then in Sec. IV we shall give a generalization of the double-integral theorem to higher-order poles. In Secs. III and IV it is implicitly assumed that a principal-value product such as $\mathscr{P}/[(x-u)(y-x)]$ may be expanded in partial fractions.[2] This is discussed in more detail in Appendix B. [See, also, the discussion of Eq. (4.15).]

A. Multiple integrals

Consider first the case of triple integrals and define

$$A_3(u)=\int\frac{\mathscr{P}}{(x-u)}dx\int\frac{\mathscr{P}}{(y-x)}dy\times\int\frac{\mathscr{P}}{(z-x)}dz\,f_3(x,y,z), \qquad (3.1)$$

$$B_3(u)=\int dz\int dy\int\frac{\mathscr{P}}{(x-u)}\times\frac{\mathscr{P}}{(y-x)}\frac{\mathscr{P}}{(z-x)}dx\,f_3(x,y,z). \qquad (3.2)$$

In Eq. (3.2) we expand the principal-value terms in partial fractions:[2]

$$B_3(u)=\int dz\int dy\int dx\left\{\frac{\mathscr{P}}{(x-u)}\frac{\mathscr{P}}{(u-y)}\frac{\mathscr{P}}{(u-z)}\times\frac{\mathscr{P}}{(x-y)}\frac{\mathscr{P}}{(y-u)}\frac{\mathscr{P}}{(y-z)}+\frac{\mathscr{P}}{(x-z)}\frac{\mathscr{P}}{(z-u)}\frac{\mathscr{P}}{(z-y)}\right\}f_3(x,y,z), \qquad (3.3)$$

which clearly can be rewritten as

$$\begin{aligned}B_3(u)=&\int\frac{\mathscr{P}}{(u-z)}dz\int\frac{\mathscr{P}}{(u-y)}dy\int\frac{\mathscr{P}}{(x-u)}dx\,f_3(x,y,z)\\&+\int dz\int\frac{\mathscr{P}}{(y-u)}\frac{\mathscr{P}}{(y-z)}dy\int\frac{\mathscr{P}}{(x-y)}dx\,f_3(x,y,z)\\&+\int\frac{\mathscr{P}}{(z-u)}dz\int\frac{\mathscr{P}}{(z-y)}dy\int\frac{\mathscr{P}}{(x-z)}dx\,f_3(x,y,z).\end{aligned} \qquad (3.4)$$

We now interchange the order of integrations in Eq. (3.4), so that the x integration becomes the outermost integration and the z integration becomes the innermost integration. Clearly, for the first term in (3.4) the order can be interchanged directly, with no additional terms, since the denominators of the principal-value terms are independent of one another.[2] Then using Eq. (1.1), we transform the second term of Eq. (3.4) as follows:

$$\begin{aligned}&\int dz\int\frac{\mathscr{P}}{(y-u)}\frac{\mathscr{P}}{(y-z)}dy\int\frac{\mathscr{P}}{(x-y)}dx\,f_3(x,y,z)\\&=\int\frac{\mathscr{P}}{(y-u)}dy\int\frac{\mathscr{P}}{(y-z)}dz\int\frac{\mathscr{P}}{(x-y)}dx\,f_3(x,y,z)-\pi^2\int\frac{\mathscr{P}}{(x-u)}f_3(x,u,u)dx\\&=\int\frac{\mathscr{P}}{(y-u)}dy\int\frac{\mathscr{P}}{(x-y)}dx\int\frac{\mathscr{P}}{(y-z)}dz\,f_3(x,y,z)-\pi^2\int\frac{\mathscr{P}}{(x-u)}dx\,f_3(x,u,u)\\&=\int dx\int\frac{\mathscr{P}}{(y-u)}\frac{\mathscr{P}}{(x-y)}dy\int\frac{\mathscr{P}}{(y-z)}dz\,f_3(x,y,z)-\pi^2\int\frac{\mathscr{P}}{(u-z)}dz\,f_3(u,u,z)-\pi^2\int\frac{\mathscr{P}}{(x-u)}dx\,f_3(x,u,u).\end{aligned} \qquad (3.5)$$

Similarly, the third term of Eq. (3.4) becomes

$$\begin{aligned}&\int\frac{\mathscr{P}}{(z-u)}dz\int\frac{\mathscr{P}}{(z-y)}dy\int\frac{\mathscr{P}}{(x-z)}dx\,f_3(x,y,z)\\&=\int dy\int dz\frac{\mathscr{P}}{(z-u)}\frac{\mathscr{P}}{(z-y)}\int\frac{\mathscr{P}}{(x-z)}dx\,f_3(x,y,z)+\pi^2\int\frac{\mathscr{P}}{(x-u)}dx\,f_3(x,u,u)\\&=\int dy\int dz\frac{\mathscr{P}}{(y-u)}\left[\frac{\mathscr{P}}{(z-y)}-\frac{\mathscr{P}}{(z-u)}\right]\int\frac{\mathscr{P}}{(x-z)}dx\,f_3(x,y,z)+\pi^2\int\frac{\mathscr{P}}{(x-u)}dx\,f_3(x,u,u)\end{aligned}$$

$$= \int \frac{\mathscr{P}}{(y-u)} dy \int dx \int dz \left[\frac{\mathscr{P}}{(z-y)} - \frac{\mathscr{P}}{(z-u)} \right] \frac{\mathscr{P}}{(x-z)} f_3(x,y,z)$$
$$- \pi^2 \int \frac{\mathscr{P}}{(y-u)} dy [f_3(y,y,y) - f_3(u,y,u)] + \pi^2 \int \frac{\mathscr{P}}{(x-u)} dx\, f_3(x,u,u)$$
$$= \int dx \int dy \int dz \frac{\mathscr{P}}{(z-u)} \frac{\mathscr{P}}{(z-y)} \frac{\mathscr{P}}{(x-z)} f_3(x,y,z)$$
$$- \pi^2 \int \frac{\mathscr{P}}{(y-u)} dy [f_3(y,y,y) - f_3(u,y,u)] + \pi^2 \int \frac{\mathscr{P}}{(x-u)} dx\, f_3(x,u,u), \tag{3.6}$$

where we note that in the second step we have made a partial fraction expansion of $[(z-u)(z-y)]^{-1}$, an expression which is again used in the fourth step after reversing the x and y integrations in the first term. Next, we substitute Eqs. (3.5) and (3.6) into Eq. (3.4) to obtain

$$B_3(u) = \int \frac{\mathscr{P}}{(x-u)} dx \int \frac{\mathscr{P}}{(u-y)} dy \int \frac{\mathscr{P}}{(u-z)} dz\, f_3(x,y,z) + \int dx \int \frac{\mathscr{P}}{(y-u)} \frac{\mathscr{P}}{(x-y)} dy \int \frac{\mathscr{P}}{(y-z)} dz\, f_3(x,y,z)$$
$$+ \int dx \int dy \int dz \frac{\mathscr{P}}{(z-u)} \frac{\mathscr{P}}{(z-y)} \frac{\mathscr{P}}{(x-z)} f_3(x,y,z) - \pi^2 \int \frac{\mathscr{P}}{(u-z)} dz\, f_3(u,u,z)$$
$$- \pi^2 \int \frac{\mathscr{P}}{(y-u)} dy [f_3(y,y,y) - f_3(u,y,u)]. \tag{3.7}$$

We observe that in the first three terms of Eq. (3.7), there is no loss in generality in putting all of the principal values into the innermost (z) integration. Then with the same partial fraction expansion that was used in converting Eq. (3.2) to Eq. (3.3), we recombine the principle values in the first two lines of Eqs. (3.7) and find that

$$A_3(u) = B_3(u) + D_3(u), \tag{3.8}$$

where

$$D_3(u) = \pi^2 \int \frac{\mathscr{P}}{(x-u)} dx [f_3(x,x,x) - f_3(u,x,u) - f_3(u,u,x)]. \tag{3.9}$$

Equation (3.8) is the appropriate generalization of the PB theorem to triple integrals. This is a lengthy derivation, but it does demonstrate the two main ingredients necessary for evaluating higher-order integrals: the original PB theorem (1.1) combined with expansions in partial fractions.

In general, we define the functions

$$A_n(u) = \int \frac{\mathscr{P}}{(x_1-u)} dx_1 \int \frac{\mathscr{P}}{(x_2-x_1)} dx_2 \int \frac{\mathscr{P}}{(x_3-x_1)} dx_3 \times \cdots \int \frac{\mathscr{P}}{(x_n-x_1)} dx_n\, f_n(x_1,x_2,\dots x_n), \tag{3.10}$$

$$B_n(u) = \int dx_n \int dx_{n-1} \cdots \int dx_2 \int dx_1 \frac{\mathscr{P}}{(x_1-u)} \frac{\mathscr{P}}{(x_2-x_1)} \times \frac{\mathscr{P}}{(x_3-x_1)} \cdots \frac{\mathscr{P}}{(x_n-x_1)} f_n(x_1,x_2,\dots x_n), \tag{3.11}$$

$$D_n(u) = A_n(u) - B_n(u). \tag{3.12}$$

In Appendix A, we prove that for order $n+1$ if one defines

$$f_n(x_1,x_2,\dots x_n) = \int \frac{\mathscr{P}}{(x_{n+1}-x_1)} dx_{n+1} \times f_{n+1}(x_1,x_2,\dots,x_n,x_{n+1}), \tag{3.13}$$

which is used to construct $D_n(u)$, then the following recursion relation holds:

$$D_{n+1}(u) = D_n(u) + (-)^n \pi^2 \int dx_n \int dx_{n-1} \cdots \int dx_2 \times \Bigg[\frac{\mathscr{P}}{\prod_{j=2}^{n} (u-x_j)} f_{n+1}(u,x_2,x_3,\dots,x_n,u) + \sum_{i=2}^{n} \frac{\mathscr{P}}{(x_i-u) \prod_{\substack{j=2 \\ j\neq i}}^{n} (x_i-x_j)} \times f_{n+1}(x_i,x_2,x_3,\dots,x_n,x_i) \Bigg], \tag{3.14}$$

where a single $\mathscr{P}$ in the numerator now denotes the principal values for all relevant terms. From Eqs. (3.9), (3.13), and (3.14), we obtain an explicit expression for $D_4(u)$:

$$D_4(u) = -\pi^2 \int dx \int dy \frac{\mathscr{P}}{(x-u)} \frac{\mathscr{P}}{(y-u)} [f_4(u,x,y,u) + f_4(u,u,x,y) + f_4(u,x,u,y)] + \pi^2 \int dx \int dy \times \frac{\mathscr{P}}{(x-u)} \frac{\mathscr{P}}{(y-x)} [f_4(x,y,x,x) + f_4(x,x,x,y)] - \pi^2 \int dx \int dy \frac{\mathscr{P}}{(y-u)} \frac{\mathscr{P}}{(y-x)} f_4(y,y,x,y). \tag{3.15}$$

A general formula for $D_n(u)$ can be derived, but it is complicated and of little practical value. The recursion relation (3.14) is considerably more useful, especially in deriving expressions for $D_n(u)$ for relatively small n.

We end this formal subsection with the following comments. One might wonder whether a generalization of the PB theorem exists for integrals of the form

$$A'_n(u) = \int \frac{\mathscr{P}}{(x_1 - u)} dx_1 \int \frac{\mathscr{P}}{(x_2 - x_1)} dx_2 \times \int \frac{\mathscr{P}}{(x_3 - x_2)} dx_3 \cdots \int \frac{\mathscr{P}}{(x_n - x_{n-1})} dx_n \times f_n(x_1, x_2, \ldots, x_n)$$

and we note that for each j from 2 to n we have the principal value $\mathscr{P}/(x_j - x_{j-1})$, whereas the corresponding denominator of Eq. (3.10) we have $\mathscr{P}/(x_j - x_1)$. Unfortunately, it does not seem possible to generalize the PB theorem to principal-value integrals of this type. Moreover, as we showed in Sec. II, perhaps the main advantage of the PB theorem is that it allows one to obtain integrals having all principal-value terms in the same integral. Obviously, this is not going to be possible for $A'_n(u)$; thus there does not seem to be any pressing physical application for integrals of this type.

B. Applications

1. Convolutions with respect to time

It is well known that a convolution with respect to time of two functions has a Fourier transform which is the simple product of the Fourier transforms of the functions, each evaluated at the same frequency or energy.[3] In this subsection we derive dispersion relations for functions of the form $f_1(\omega) f_2(\omega)$.

First, let $h_i(\omega)$ and $g_i(\omega)$ be functions which are Hilbert transforms of one another according to Eq. (2.1). Also, note the relation

$$\int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_\alpha)(\omega - \omega_\beta)(\omega - \omega_\gamma)} d\omega = 0 \quad (\omega_\alpha, \omega_\beta, \omega_\gamma \text{ real}). \tag{3.16}$$

[See Eq. (A4) in Appendix A.] Then evaluate

$$F_2(\omega_0) \equiv \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_0)} h_1(\omega) h_2(\omega) d\omega, \tag{3.17}$$

which from Eqs. (2.1), (3.1), (3.2), (3.8), and (3.9) can be reexpressed as

$$F_2(\omega_0) = \frac{1}{\pi^3} \int_{-\infty}^{\infty} d\omega'' \, g_2(\omega'') \int_{-\infty}^{\infty} d\omega' \, g_1(\omega') \times \int_{-\infty}^{\infty} d\omega \frac{\mathscr{P}}{(\omega - \omega_0)(\omega - \omega')(\omega - \omega'')} + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_0)} d\omega [g_1(\omega) g_2(\omega) - g_1(\omega) g_2(\omega_0) - g_1(\omega_0) g_2(\omega)]. \tag{3.18}$$

Using Eqs. (2.1) and (3.16), we find that

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_0)} d\omega [h_1(\omega) h_2(\omega) - g_1(\omega) g_2(\omega)] = -g_1(\omega_0) h_2(\omega_0) - h_1(\omega_0) g_2(\omega_0), \tag{3.19}$$

a result that is also valid if the functions have other (hidden) coordinates or indices; there is a matrix multiplication or an integration implied in the products $h_1(\omega) h_2(\omega)$, $g_1(\omega) g_2(\omega)$, etc. Here we just focus on the frequency (or energy) dependence.

Next let

$$h_i(\omega) = \mathrm{Re}[G_i(\omega)] \quad (\text{real } \omega), \tag{3.20a}$$

$$g_i(\omega) = \eta_i(\omega) \mathrm{Im}\, G_i(\omega) \quad (\text{real } \omega), \tag{3.20b}$$

so that the real and imaginary part of $G_i(\omega)$ satisfy the dispersion relations (2.5). Substituting Eqs. (3.20) into Eq. (3.19), we obtain

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_0)} d\omega \{\mathrm{Re}[G_1(\omega)] \mathrm{Re}[G_2(\omega)] - \eta_1(\omega) \eta_2(\omega) \mathrm{Im}[G_1(\omega)] \mathrm{Im}[G_2(\omega)]\} = -\{\eta_1(\omega_0) \mathrm{Im}[G_1(\omega_0)] \mathrm{Re}[G_2(\omega_0)] + \eta_2(\omega_0) \mathrm{Im}[G_2(\omega_0)] \mathrm{Re}[G_1(\omega_0)]\} \tag{3.21a}$$

or, from Eqs. (2.1) and (2.4),

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_0)} \{\eta_1(\omega) \mathrm{Im}[G_1(\omega)] \mathrm{Re}[G_2(\omega)] + \eta_2(\omega) \mathrm{Im}[G_2(\omega)] \mathrm{Re}[G_1(\omega)]\} = \mathrm{Re}[G_1(\omega_0)] \mathrm{Re}[G_2(\omega_0)] - \eta_1(\omega_0) \eta_2(\omega_0) \mathrm{Im}[G_1(\omega_0)]\, \mathrm{Im}[G_2(\omega_0)]\}. \tag{3.21b}$$

Equations (3.21) are general results that apply to either zero-temperature and finite-temperature Green's functions.

Also, Eqs. (3.21) can be derived from analyticity considerations. In particular, from Eqs. (2.1) and (3.20) and the Titchmarsh theorem,[13] we know that the function

$$\psi_i(\omega) = h_i(\omega) + i g_i(\omega) \tag{3.22}$$

is analytic in the upper-half complex ω plane. Then a product of such functions

$$\psi_{12}(\omega) = \psi_1(\omega) \psi_2(\omega)$$

will have identical analytic behavior and Eqs. (3.21) follow immediately from complex variable theory.[10] In this section, we have derived Eqs. (3.21) only from Eq. (3.9), the generalization of the PB theorem to triple integrals, without explicit use of the Titchmarsh theorem and the complex variable theory. However, as was emphasized in Sec. II, there is an intimate relationship between the ordinary PB theorem (for double integrals) and analyticity. The present example for triple integrals again underscores the deep connection between the principal-value theorem and the analytic properties of the functions.

We now specialize Eqs. (3.21) to the case

$$\eta(\omega) \equiv \eta_1(\omega) = \eta_2(\omega), \tag{3.23a}$$

$$[\eta(\omega)]^2 = 1, \tag{3.23b}$$

which applies to zero-temperature Green's functions, with the same chemical potential. Equations (3.21) then reduce to

$$\eta(\omega_0) \mathrm{Im}[G_1(\omega_0) G_2(\omega_0)] = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega - \omega_0)} d\omega \times \mathrm{Re}[G_1(\omega) G_2(\omega)], \tag{3.24a}$$

$$\mathrm{Re}[G_1(\omega)G_2(\omega_0)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega-\omega_0)} d\omega\, \eta(\omega) \times \mathrm{Im}[G_1(\omega)G_2(\omega)]. \quad (3.24b)$$

Thus if the zero-temperature causal propagators $G_1(\omega)$ and $G_2(\omega)$ satisfy the dispersion relations (2.5), their product must satisfy the same dispersion relations. Moreover, this result immediately implies the general relations

$$\eta(\omega_0)\mathrm{Im}\left[\prod_{j=1}^{n} G_j(\omega_0)\right] = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega-\omega_0)} d\omega \times \mathrm{Re}\left[\prod_{j=1}^{n} G_j(\omega)\right], \quad (3.25a)$$

$$\mathrm{Re}\left[\prod_{j=1}^{n} G_j(\omega_0)\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathscr{P}}{(\omega-\omega_0)} d\omega \times \eta(\omega)\mathrm{Im}\left[\prod_{j=1}^{n} G_j(\omega)\right]. \quad (3.25b)$$

Note that in Eqs. (3.25) the subscript j can label different Green's functions, or for some or all of the Green's functions it can be the same index.

Also, it should be emphasized again that the products occurring in Eqs. (3.25) are convolutions if one can perform a Fourier transformation to the time variable.[3] For example, define

$$C_n(t-t') = C_n(t,t') = \frac{1}{(2\pi)^{n-1}}\int G_1(t,t_1)dt_1 \times \int G_2(t_1,t_2)dt_2 \int G_3(t_2,t_3)dt_3 \times \cdots \int dt_{n-1}\, G_n(t_{n-1},t'),$$

where we have assumed stationary or steady-state motion with $G_i(t,t') = G_i(t-t')$. It can then be shown that the Fourier transform of C_n is given by

$$C_n(\omega) = (2\pi)^{-1}\int d\tau\, e^{i\omega\tau} C_n(\tau) = \prod_{j=1}^{n} G_j(\omega),$$

which is the product appearing in Eqs. (3.25). In such a convolution one expects that each factor in the product should have the same chemical potential.[3] [See the discussion below regarding Eq. (3.26).]

2. Dyson's equations

We now apply Eqs. (3.25) to the zero-temperature Migdal parametrization of Dyson's equation for the pion propagator,[3,14] namely,

$$D(p) = [1 + \tfrac{16}{9}(f^*/m_\pi)^2|\mathbf{p}|^2 U_{N\Delta}(p) D_0(p)]^{-1} D_0(p), \quad (3.26)$$

where $p \equiv (\omega,\mathbf{p})$ is a four vector, f^* is the nucleon–delta–pion coupling constant, and

$$D_0(p) = (p^2 - m_\pi^2 + i\epsilon)^{-1} \quad (3.27)$$

is the free-particle causal propagator for the pion. As we remarked at the end of Sec. II, $U_{N\Delta}$ is determined by a second Dyson equation involving the exact delta propagator;[3] the iterated solution of these coupled integral equations is presently being investigated.[14] Here, the function $U_{N\Delta}(p)$ is approximated from Eq. (2.30) using the bare function $U^{(0)}_{N\Delta}(p)$, which is calculated from Eqs. (2.31), (2.32), and (2.16a). Also, from Eqs. (2.18) and (2.29) we have

$$\eta_{N\Delta}(\omega) \underset{T\to 0}{\to} \mathrm{sgn}(\omega)$$

and $U^{(0)}_{N\Delta}(p)$ behaves like a Green's function whose chemical potential is zero. Thus $U^{(0)}_{N\Delta}$ and D_0 each have the same chemical potential: $\mu_\pi = 0$. Then note the in each of Eqs. (2.30) and (3.26) there is the product of a function with the inverse of another function, giving rise to an infinite series of products of the form

$$\prod_{j=1}^{n} G_j(\omega),$$

each factor of which has $\mu_\pi = 0$. If we apply Eq. (3.25) to each of these products, we find a satisfying result, namely that $U^{(0)}_{N\Delta}(p), U_{N\Delta}(p), D_0(p)$, and $D(p)$ all obey the same dispersion relation. Thus Dyson's equation for the pion, Eq. (3.26), is rigorously causal. The same result also pertains to Dyson's equation for the delta particle.

3. Triple-propagator loop integral

Next, we examine the structure of the following contribution to the self-energy of a particle with either Bose or Fermi statistics:

$$L_{\alpha\beta\gamma}(\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_\alpha(\omega_\alpha)G_\beta(\omega_\beta) \times G_\gamma(\omega_\alpha+\omega_\beta+\omega)d\omega_\alpha\, d\omega_\beta. \quad (3.28)$$

(See Fig. 6.) The contribution (3.28) arises for any particle that has a dominant three-particle decay mode. Since each Green's function obeys the dispersion relations (2.5), direct application of the PB theorem for triple integrals, Eqs. (3.8) and (3.9), yields an expression for $\mathrm{Re}[L_{\alpha\beta\gamma}(\omega)]$ in terms of only the imaginary parts of the individual Green's functions, namely,

$$\mathrm{Re}[L_{\alpha\beta\gamma}(\omega)] = \frac{\mathscr{P}}{\pi}\int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta \int_{-\infty}^{\infty} d\omega_\gamma \frac{\mathrm{Im}[G_\alpha(\omega_\alpha)]\mathrm{Im}[G_\beta(\omega_\beta)]\mathrm{Im}[G_\gamma(\omega_\gamma)]}{(\omega+\omega_\alpha+\omega_\beta-\omega_\gamma)} \times [\eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta)\eta_\gamma(\omega_\gamma) - \eta_\alpha(\omega_\alpha) - \eta_\beta(\omega_\beta) + \eta_\gamma(\omega_\gamma)]. \quad (3.29a)$$

Also, the imaginary part of $L_{\alpha\beta\gamma}(\omega)$ can be written as

$$\mathrm{Im}[L_{\alpha\beta\gamma}(\omega)] = \int_{-\infty}^{\infty} d\omega_\alpha \int_{-\infty}^{\infty} d\omega_\beta\, \mathrm{Im}[G_\alpha(\omega_\alpha)]\mathrm{Im}[G_\beta(\omega_\beta)]\mathrm{Im}[G_\gamma(\omega_\alpha+\omega_\beta+\omega)] \times [\eta_\beta(\omega_\beta)\eta_\gamma(\omega_\alpha+\omega_\beta+\omega) + \eta_\gamma(\omega_\alpha+\omega_\beta+\omega)\eta_\alpha(\omega_\alpha) - \eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta) - 1]. \quad (3.29b)$$

Then from Eqs. (3.29), we see that $L_{\alpha\beta\gamma}(\omega)$ obeys the dispersion relation

$$\mathrm{Re}[L_{\alpha\beta\gamma}(\omega)] = \frac{\mathscr{P}}{\pi}\int_{-\infty}^{\infty} \eta_{\alpha\beta\gamma}(\omega')\frac{\mathrm{Im}[L_{\alpha\beta\gamma}(\omega')]d\omega'}{(\omega'-\omega)}, \tag{3.30}$$

where

$$\eta_{\alpha\beta\gamma}(\omega_\gamma-\omega_\alpha-\omega_\beta) = \frac{\eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta)\eta_\gamma(\omega_\gamma)-\eta_\alpha(\omega_\alpha)-\eta_\beta(\omega_\beta)+\eta_\gamma(\omega_\gamma)}{1+\eta_\alpha(\omega_\alpha)\eta_\beta(\omega_\beta)-\eta_\beta(\omega_\beta)\eta_\gamma(\omega_\gamma)-\eta_\alpha(\omega_\alpha)\eta_\gamma(\omega_\gamma)} \tag{3.31}$$

and, from Eq. (2.6),

$$\eta_{\alpha\beta\gamma}(\omega) = \begin{cases}\tanh\{[\omega-(\mu_\gamma-\mu_\alpha-\mu_\beta)]/2k_BT\}, & \text{for three bosons or two fermions and one boson,}\\ \coth\{[\omega-(\mu_\gamma-\mu_\alpha-\mu_\beta)]/2k_BT\}, & \text{for three fermions or two bosons and one fermion.}\end{cases} \tag{3.32}$$

Thus the contribution to the self-energy has the same statistics as the loop as a whole (or as the external propagator shown in Fig. 6). Note, also, that the net chemical potential is $\mu_\gamma-\mu_\alpha-\mu_\beta$, analogous to Eqs. (2.18) and (2.19) for the two-propagator loop integral. We can also treat two of the particles in Fig. 6 as a loop in parallel with the third since we have clearly demonstrated that a loop obeys a dispersion relation similar in form to its constituents. Then the results of Sec. II are directly applicable and we obtain an alternative derivation of Eqs. (3.29)–(3.31). Of course, this is just another way of stating that the PB theorem for triple integrals can be derived from the ordinary double-integral result.

4. Simple example of separable products

Finally, we examine triple integrals in which $f_3(x,y,z)$ is a separable function of sin and cos functions, e.g.,

$$f_3(x,y,z) = \sin x \sin y \sin z.$$

For such a function, one uses the well-known Hilbert transforms[10,11]

$$\cos u = \frac{1}{\pi}\int_{-\infty}^{\infty} dx\,\frac{\mathscr{P}}{(x-u)}\sin x,$$
$$\sin u = -\frac{1}{\pi}\int_{-\infty}^{\infty} dx\,\frac{\mathscr{P}}{(x-u)}\cos x. \tag{3.33}$$

In Table I we present the results obtained from Eqs. (3.9) and (3.33) for various separable functions. In each of the cases tabulated, $D_3(u)$ is identically equal to $A_3(u)$ defined in Eq. (3.1), which from Eq. (3.8) means that $B_n(u)$ vanishes; however, this is not a general feature (as we shall see from the examples in Table II).

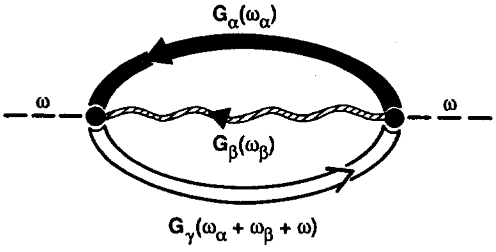


FIG. 6. A Feynman diagram for a three-propagator response function. The dashed line represents the external propagator into which the loop is inserted.

IV. GENERALIZATION OF THE PB THEOREM, FOR A DOUBLE INTEGRAL, TO HIGHER-ORDER POLES

In this section we will generalize the PB theorem[1,2] to higher-order poles. We first give the *definition* of a single higher-order principal value in a one-dimensional integral:[20–22]

$$\int_{-\infty}^{\infty} dx\,\frac{\mathscr{P}}{(x-u)^n}f(x)$$
$$= \frac{(-)^{n-1}}{(n-1)!}\lim_{\epsilon\to 0+}\int_{-\infty}^{\infty} dx \times\left[\frac{d^{(n-1)}}{dx^{(n-1)}}\frac{(x-u)}{((x-u)^2+\epsilon^2)}\right]f(x)$$
$$= \frac{1}{(n-1)!}\int_{-\infty}^{\infty} dx\,\frac{\mathscr{P}}{(x-u)}\frac{d^{(n-1)}}{dx^{(n-1)}}f(x). \tag{4.1}$$

Note that in Eq. (4.1) we have reduced the expression for a higher-order pole to that of a single pole, which is a trick we will continue to use throughout this section. Also, we shall assume that $f(x)$, or $f(x,y)$, is always finite as $x,y\to\pm\infty$, so that boundary terms vanish, e.g., when integrating by parts. The results of this section will clearly be valid for limits other than $\pm\infty$ provided that the corresponding boundary terms are zero. If the boundary terms do not vanish, then the contributions from such terms must be added onto the appropriate equations.[20] However, for most applications of physical interest, one does not have to be concerned about this problem.

Next, we will prove several important relations. Consider the integral

TABLE I. Evaluation of $D_3(u)$ for various products of sin and cos functions.

$f_3(x,y,z)$	$D_3(u)/\pi^3$
$\sin x\sin y\sin z$	$\frac{1}{2}\cos u\cos 2u$
$\sin x\sin y\cos z$ $\sin x\cos y\sin z$	$-\frac{1}{2}\sin u\cos 2u$
$\sin x\cos y\cos z$	$\frac{1}{2}\cos u(3-2\cos^2 u)$
$\cos x\sin y\sin z$	$-\frac{1}{2}\sin u(3-2\sin^2 u)$
$\cos x\sin y\cos z$ $\cos x\cos y\sin z$	$-\frac{1}{2}\cos u\cos 2u$
$\cos x\cos y\cos z$	$\frac{1}{2}\sin u\cos 2u$

$$I'(x)=\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}f(y)$$
$$=\lim_{\epsilon\to 0+}\left(\int_{-\infty}^{x-\epsilon}+\int_{x+\epsilon}^{\infty}\right)\frac{dy}{(y-x)}f(y), \quad (4.2)$$
which we differentiate with respect to x to obtain
$$\frac{dI'(x)}{dx}=\lim_{\epsilon\to 0+}\left\{-\frac{1}{\epsilon}[f(x-\epsilon)+f(x+\epsilon)]+\left(\int_{-\infty}^{x-\epsilon}+\int_{x+\epsilon}^{\infty}\right)\frac{dy}{(y-x)^2}f(y)\right\}. \quad (4.3)$$
We then integrate Eq. (4.3) by parts, giving
$$\frac{dI'(x)}{dx}=\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}\frac{df(y)}{dy}$$
or, in general, from Eq. (4.1),
$$\frac{d^{(n-1)}I'(x)}{dx^{(n-1)}}=(n-1)!\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)^n}f(y)$$
$$=\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}\frac{d^{(n-1)}f(y)}{dy^{(n-1)}}. \quad (4.4)$$
Suppose that f in Eq. (4.2) depends on x as well as y, i.e.,
$$I''(x)=\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}f(x,y). \quad (4.5)$$
Then by the Leibnitz theorem[23] we find that
$$\frac{d^{(n-1)}I''(x)}{dx^{(n-1)}}=\sum_{r=0}^{n-1}C_r^{n-1}\int_{-\infty}^{\infty}dy(r!)\frac{\mathscr{P}}{(y-x)^{r+1}}\frac{\partial^{(n-1-r)}f(x,y)}{\partial x^{(n-1-r)}}$$
$$=\sum_{r=0}^{n-1}C_r^{n-1}\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}\times\left[\frac{\partial^{(r)}}{\partial y^{(r)}}\frac{\partial^{(n-1-r)}}{\partial x^{(n-1-r)}}f(x,y)\right], \quad (4.6)$$
where
$$C_r^N=N!/r!(N-r)!$$
is the usual binomial coefficient. Also, let
$$I(x)=\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)^m}f(x,y) \quad (4.7)$$
and, from Eqs. (4.1), (4.5), and (4.6) we obtain
$$\frac{d^{(n-1)}I(x)}{dx^{(n-1)}}=\frac{1}{(m-1)!}\sum_{r=0}^{n-1}C_r^{n-1}\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}\times\left[\frac{\partial^{(n-1-r)}}{\partial x^{(n-1-r)}}\frac{\partial^{(m-1+r)}}{\partial y^{(m-1+r)}}f(x,y)\right]. \quad (4.8)$$
Now we consider the double integral
$$A(u)=\int_{-\infty}^{\infty}dx\,\frac{\mathscr{P}}{(x-u)^n}\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)^m}f(x,y)$$
$$=\int_{-\infty}^{\infty}dx\,\frac{\mathscr{P}}{(x-u)^n}I(x), \quad (4.9)$$
which, from Eqs. (4.1) and (4.8), becomes
$$A(u)=\int_{-\infty}^{\infty}dx\,\frac{\mathscr{P}}{(x-u)}\int_{-\infty}^{\infty}dy\,\frac{\mathscr{P}}{(y-x)}F_{nm}(x,y), \quad (4.10)$$
where
$$F_{nm}(x,y)=\frac{1}{(n-1)!(m-1)!}\sum_{r=0}^{n-1}C_r^{n-1}\times\frac{\partial^{(n-1-r)}}{\partial x^{(n-1-r)}}\frac{\partial^{(m-1+r)}}{\partial y^{(m-1+r)}}f(x,y). \quad (4.11)$$
Equation (4.10) involves only first-order poles, so that we can apply the ordinary PB theorem (1.1) to obtain
$$A(u)=B(u)-\pi^2F_{nm}(u,u), \quad (4.12)$$
with the function $B(u)$ given by
$$B(u)=\int_{-\infty}^{\infty}dy\int_{-\infty}^{\infty}dx\,\frac{\mathscr{P}}{(x-u)}\frac{\mathscr{P}}{(y-x)}F_{nm}(x,y). \quad (4.13)$$
Comparing Eqs. (1.1), (4.9), and (4.13), we see that the final step in the generalization of the theorem is to prove that
$$B(u)=\int_{-\infty}^{\infty}dy\int_{-\infty}^{\infty}dx\,\frac{\mathscr{P}}{(x-u)^n}\frac{\mathscr{P}}{(y-x)^m}f(x,y), \quad (4.14)$$
which is derived in Appendix B.

We emphasize that the proper defining equations for $A(u)$ and $B(u)$ are Eqs. (4.9) and (4.14), respectively. However, it is clearly much easier to evaluate these functions using Eqs. (4.10) and (4.13) since the integrations by parts have been done in order to convert all principal values to first order. We note, also, that[2]
$$\frac{\mathscr{P}}{(x-u)}\frac{\mathscr{P}}{(y-x)}=\frac{\mathscr{P}}{(y-u)}\left[\frac{\mathscr{P}}{(x-u)}-\frac{\mathscr{P}}{(x-y)}\right]; \quad (4.15)$$
this relation was also used in Sec. III and can be formally justified by using the limiting procedure indicated in Eq. (4.1). Equation (4.15) is in some sense a trivial identity

TABLE II. Evaluation of $A(u)$, $B(u)$, and $F_{nm}(u,u)$ defined in Eqs. (4.9), (4.14), and (4.11), respectively, for the various $f_i(x,y)$ functions given in Eq. (4.17). The first four lines are for the ordinary PB theorem (1.1).

n	m	$f(x,y)$	$A(u)/\pi^2$	$B(u)/\pi^2$	$-F_{nm}(u,u)$
1	1	f_1	$\frac{1}{2}\cos 2u$	$\frac{1}{2}$	$-\sin^2 u$
		f_2	$-\frac{1}{2}\sin 2u$	0	$-\frac{1}{2}\sin 2u$
		f_3	$-\frac{1}{2}\sin 2u$	0	$-\frac{1}{2}\sin 2u$
		f_4	$-\frac{1}{2}\cos 2u$	$\frac{1}{2}$	$-\cos^2 u$
2	1	f_1	$-\sin 2u$	0	$-\sin 2u$
		f_2	$-\cos 2u$	0	$-\cos 2u$
		f_3	$-\cos 2u$	0	$-\cos 2u$
		f_4	$\sin 2u$	0	$\sin 2u$
1	2	f_1	$-\frac{1}{2}\sin 2u$	0	$-\frac{1}{2}\sin 2u$
		f_2	$-\frac{1}{2}\cos 2u$	$\frac{1}{2}$	$-\cos^2 u$
		f_3	$-\frac{1}{2}\cos 2u$	$-\frac{1}{2}$	$\sin^2 u$
		f_4	$\frac{1}{2}\sin 2u$	0	$\frac{1}{2}\sin 2u$
2	2	f_1	$-\cos 2u$	0	$-\cos 2u$
		f_2	$\sin 2u$	0	$\sin 2u$
		f_3	$\sin 2u$	0	$\sin 2u$
		f_4	$\cos 2u$	0	$\cos 2u$

since it is valid as an algebraic expression and must, therefore, be valid if we exclude the values where the denominators vanish. The same remark applies to the general expressions (A1) and (A2) of Appendix A and Eq. (B11) of Appendix B. From Eq. (4.15), Eq. (4.13) becomes

$$B(u) = \int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)} \int_{-\infty}^{\infty} dx \times \left[\frac{\mathscr{P}}{(x-u)} - \frac{\mathscr{P}}{(x-y)} \right] F_{nm}(x,y), \tag{4.16}$$

which is considerably easier to use than Eq. (4.13) since each integral contains only one principal value.

Finally, we present some simple examples of Eq. (4.12) for different $f(x, y)$ functions. Consider the separable functions

$$f_1(x,y) = \sin x \sin y, \quad f_2(x,y) = \cos x \sin y,$$
$$f_3(x,y) = \sin x \cos y, \quad f_4(x,y) = \cos x \cos y. \tag{4.17}$$

From Eqs. (3.33) and (4.17), we evaluate $A(u)$, $B(u)$, and $F_{nm}(u,u)$ for various cases and list the results in Table II. [As mentioned, $A(u)$ can be evaluated using either of Eqs. (4.9) or (4.10); similarly, $B(u)$ can be evaluated using Eqs. (4.14), (4.13), or (4.16).] It is then easily verified that Eq. (4.12) is satisfied for each case tabulated.

ACKNOWLEDGMENTS

We would like to thank Phil Siemens and Madeleine Soyeur for several helpful discussions. Also, we are very grateful to Dick Becker and Phil Siemens, who carefully read the original versions of this manuscript and made a number of constructive suggestions.

The research was sponsored by the Division of Nuclear Physics, U.S. Department of Energy Contract No. DE-AC05-840R21400 with Martin Marietta Energy Systems, Inc.

APPENDIX A: DERIVATION OF THE MULTIPLE-INTEGRAL RECURSION RELATION

In this Appendix we first give the expansion in partial fractions of the function

$$\left[\prod_{i=1}^{n}(x-x_i)\right]^{-1},$$

which is then used to derive the recursion relation (3.14).

It is well known that

$$\frac{1}{\prod_{i=1}^{n}(x-x_i)} = \sum_{i=1}^{n} \frac{K_i^{(n)}}{(x-x_i)}, \tag{A1}$$

where

$$K_i^{(n)} = \left[\prod_{\substack{j=1\\ j\neq i}}^{n}(x_i-x_j)\right]^{-1}. \tag{A2}$$

Thus Eq. (A1) enables one to convert an integral containing products of principal values to a sum of terms, each of which contains only one principal value. Also, we recall the well-known result

$$\int_{-\infty}^{\infty} \frac{\mathscr{P}}{(x-a)} dx = 0 \quad (a \text{ real}), \tag{A3}$$

which ,when combined with Eq. (A1), gives

$$\int_{-\infty}^{\infty} \frac{\mathscr{P}}{\prod_{i=1}^{n}(x-x_i)} dx = 0 \quad (x_i \text{ all real}). \tag{A4}$$

Equations (2.3) and (3.16) are special cases of Eq. (A4).

We return to the discussion in Sec. III concerning Eqs. (3.10)–(3.14). Note that Eq. (3.12) is merely a *definition* of $D_n(u)$. In Eqs. (1.1), (3.9), and (3.15) $D_n(u)$ is given for the cases $n = 2$, 3, and 4, respectively. Our aim here is to derive a general recursion relation between $D_{n+1}(u)$ and $D_n(u)$. We begin by reducing the $n+1$ problem to nth order via the definition in Eq. (3.13) and, also,

$$\overline{A}_n(u) \equiv A_{n+1}(u) = \overline{B}_n(u) + \overline{D}_n(u), \tag{A5}$$

where $\overline{B}_n(u)$ is given by Eq. (3.11), but with f_n defined in terms of f_{n+1} via Eq. (3.13). We reexpress Eq. (3.11) as

$$\overline{B}_n(u) = \int dx_n \cdots \int dx_2 \int dx_1 \frac{\mathscr{P}}{(x_1-u)(x_2-x_1)\cdots(x_n-x_1)} f_n(x_1,x_2,\ldots,x_n) = \int dx_n \cdots \int dx_2\, H(x_2,x_3,\ldots,x_n), \tag{A6}$$

$$H(x_2,x_3,\ldots,x_n) \equiv \int \frac{\mathscr{P}}{(x_1-u)(x_2-x_1)\cdots(x_n-x_1)} dx_1 f_n(x_1,x_2,\ldots,x_n). \tag{A7}$$

As in Sec. III, a single $\mathscr{P}$ symbol in the numerator denotes the principal value of all possible terms. The $f_{n+1}(x_1,x_2,\ldots,x_n,x_{x+1})$ function in Eq. (3.13) is assumed to be an arbitrary, but well-behaved function of its arguments.

From Eqs. (A1) and (A2), Eq. (A7) can be rewritten as

$$H(x_2,x_3,\ldots,x_n) = (-)^{n+1} \int dx_1 \left[\frac{\mathscr{P}}{(x_1-u)\prod_{j=2}^{n}(u-x_j)} + \sum_{i=2}^{n} \frac{\mathscr{P}}{(x_1-x_i)(x_i-u)\prod_{\substack{j=2\\ j\neq i}}^{n}(x_i-x_j)} \right] f_n(x_1,x_2,\ldots,x_n),$$

which from Eq. (1.1) and (3.13) becomes

$$H(x_2,x_3,\dots,x_n)$$

$$= (-)^{n+1}\int dx_{n+1}\int dx_1 \frac{\mathcal{P}}{(x_{n+1}-x_1)}\left[\frac{\mathcal{P}}{(x_1-u)\prod_{j=2}^{n}(u-x_j)}+\sum_{i=2}^{n}\frac{\mathcal{P}}{(x_1-x_i)(x_i-u)}\left(\prod_{\substack{j=2\\ j\neq i}}^{n}(x_i-x_j)\right)^{-1}\right]$$

$$\times f_{n+1}(x_1,x_2,\dots,x_n,x_{n+1}) + (-)^n\pi^2\left[\frac{\mathcal{P}}{\prod_{j=2}^{n}(u-x_j)}f_{n+1}(u,x_2,x_3,\dots,x_n,u)\right. \tag{A8}$$

$$\left.+\sum_{i=2}^{n}\frac{\mathcal{P}}{(x_i-u)}\left[\prod_{\substack{j=2\\ j\neq i}}^{n}(x_i-x_j)\right]^{-1}f_{n+1}(x_i,x_2,x_3,\dots,x_n,x_i)\right].$$

If we then substitute Eq. (A8) into (A6), we note that the x_{n+1} integration can be moved all the way to the left, so that

$$\overline{B}_n(u) = (-)^{n+1}\int dx_{n+1}\int dx_n\cdots\int dx_2\int dx_1\frac{\mathcal{P}}{(x_{n+1}-x_1)}\left[\frac{\mathcal{P}}{(x_1-u)\prod_{j=2}^{n}(u-x_j)}\right.$$

$$\left.+\sum_{i=2}^{n}\frac{\mathcal{P}}{(x_1-x_i)(x_i-u)\prod_{\substack{j=2\\ j\neq i}}^{n}(x_i-x_j)}\right]f_{n+1}(x_1,x_2,\dots,x_n,x_{n+1})+\chi(u), \tag{A9}$$

with

$$\chi(u) = (-)^n\pi^2\int dx_n\cdots\int dx_2\left[\frac{\mathcal{P}}{\prod_{j=2}^{n}(u-x_j)}f_{n+1}(u,x_2,x_3,\dots,x_n,u)\right.$$

$$\left.+\sum_{i=2}^{n}\frac{\mathcal{P}}{(x_i-u)\prod_{\substack{j\neq 2\\ j\neq i}}^{n}(x_i-x_j)}f_{n+1}(x_i,x_2,x_3,\dots,x_n,x_i)\right]. \tag{A10}$$

Next, we observe that all the principal-value terms in Eq. (A9) can be recombined, using Eqs. (A1) and (A2), giving

$$\overline{B}_n(u) = \left\{\int dx_{n+1}\int dx_n\cdots\int dx_1\frac{\mathcal{P}}{(x_1-u)(x_2-x_1)\cdots(x_n-x_1)(x_{n+1}-x_1)}f_{n+1}(x_1,x_2,\cdots,x_n,x_{n+1})\right\}+\chi(u); \tag{A11}$$

however, from Eq. (3.11) we see that

$$\overline{B}_n(u) = B_{n+1}(u)+\chi(u). \tag{A12}$$

Finally, from Eqs. (3.12), (A5), and (A12) we obtain

$$D_{n+1}(u) = \overline{D}_n(u)+\chi(u), \tag{A13}$$

which is Eq. (3.14) if we remove the overbar from $\overline{D}_n(u)$. [We use the overbar here to emphasize that $\overline{D}_n(u)$ is evaluated using Eq. (3.13), so that the $n+1$ problem is reduced to nth order à la Eqs. (3.13) and (A5).]

APPENDIX B: PARTIAL FRACTION EXPANSION FOR HIGHER-ORDER POLES AND PROOF OF EQ. (4.14)

We begin by deriving another important partial fraction identity and then proving Eq. (4.14), which completes the generalization of the PB theorem to higher-order poles.

A. Partial fraction expansion of the function $(x-z)^{-n}(y-x)^{-m}$

We desire to expand the function $(x-z)^{-n}(y-x)^{-m}$ in summations involving the combinations $(y-z)^{-n_1}(x-z)^{-m_1}$ and $(y-z)^{-n_2}(y-x)^{-m_2}$. First, we construct the integral

$$J(z) = \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\frac{1}{(x-z)^n}\frac{\mathcal{P}}{(y-x)^m}f(x,y), \tag{B1}$$

where z is complex and $f(x,y)$ is a function that is completely arbitrary except for the requirements that it is differentiable according to Eq. (4.11) and that the integral be convergent.[2] We note that

$$\frac{1}{(x-z)^n} = \frac{(-)^{n-1}}{(n-1)!}\frac{\partial^{(n-1)}}{\partial x^{(n-1)}}\left(\frac{1}{x-z}\right) \tag{B2}$$

and, from Eq. (4.1), $\mathcal{P}/(x-u)^n$ obeys the same equation, namely[20]

$$\frac{\mathcal{P}}{(x-u)^n} = \frac{(-)^{n-1}}{(n-1)!}\frac{\partial^{(n-1)}}{\partial x^{(n-1)}}\left[\frac{(x-u)}{(x-u)^2+\epsilon^2}\right], \tag{B3}$$

with the limit $\epsilon\to 0+$ always understood. Thus the algebra for x^{-1} is exactly the same as that for $\mathcal{P}/x^{-1}$ and from Eqs. (4.9) and (4.10), we see that

$$J(z) = \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\frac{1}{(x-z)}\frac{\mathcal{P}}{(y-x)}F_{nm}(x,y), \tag{B4}$$

where $F_{nm}(x,y)$ is defined in Eq. (4.11). Then we use Eq. (4.15) to obtain

$$J(z) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \frac{1}{(y-z)} \times \left[\frac{1}{(x-z)} - \frac{\mathscr{P}}{(x-y)} \right] F_{nm}(x,y). \quad \text{(B5)}$$

Next, we integrate Eq. (B5) by parts using the derivatives contained in $F_{nm}(x,y)$ and the formula

$$\frac{\partial^{(k)}(x-z)^{-l}}{\partial x^{(k)}} = \frac{(l+k-1)!}{(l-1)!} \frac{(-)^k}{(x-z)^{k+l}}, \quad \text{(B6)}$$

a result that holds for either $(x-z)^{-l}$ or $\mathscr{P}/(x-u)^l$. After some algebra, we obtain

$$J(z) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \Bigg\{ k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \Bigg[H_{nm}^{(r)} \frac{1}{(y-z)^{m+r}} \times \frac{1}{(x-z)^{n-r}} - (-)^{n+r} \sum_{r'=0}^{m-1+r} G_{nm}^{(rr')} \times \frac{1}{(y-z)^{r'+1}} \frac{\mathscr{P}}{(y-x)^{n+m-r'-1}} \Bigg] \Bigg\} f(x,y), \quad \text{(B7)}$$

where C_r^{n-1} is a binomial coefficient and

$$k_{nm} = [(n-1)!(m-1)!]^{-1}, \quad \text{(B8)}$$

$$H_{nm}^{(r)} = (n-1-r)!(m-1+r)!, \quad \text{(B9)}$$

and

$$G_{nm}^{(rr')} = C_{r'}^{m-1+r}(n+m-r'-2)!r'!. \quad \text{(B10)}$$

Note that since $f(x,y)$ is arbitrary, we have also verified the *algebraic* relation

$$\frac{1}{(x-z)^n} \frac{1}{(y-x)^m} = k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \Bigg[H_{nm}^{(r)} \frac{1}{(y-z)^{m+r}} \frac{1}{(x-z)^{n-r}} - (-)^{n+r} \sum_{r'=0}^{m-1+r} G_{nm}^{(rr')} \frac{1}{(y-z)^{r'+1}} \times \frac{1}{(y-x)^{n+m-r'-1}} \Bigg]. \quad \text{(B11)}$$

We emphasize, also, that Eq. (B11) is formally valid in Eq. (B1) if we let $z \to u$ (a real number), so that $(x-u)^{-1} \to \mathscr{P}/(x-u)$ and $(y-u)^{-1} \to \mathscr{P}/(y-u)$. Again, as indicated in Eqs. (B2) and (B3), this is because the algebra associated with x^{-1} is exactly the same as that with $\mathscr{P}/x$. Also, note that in Eq. (B1) for $z \to u$, the two principal values arising from the second and third lines of Eq. (B11) occur in the innermost y integration. However, two higher-order principal values in the same integral can be justified using the limiting procedure indicated in Eq. (B3).[2] Of course, as we have demonstrated in Sec. IV, there are no special problems in manipulating the derivatives associated with higher-order poles if each of the principal values is in a different integral. *Reversing the order of integrations* is another matter and for this one needs the PB theorem!

B. Proof of Eq. (4.14)

Because the rhs of Eq. (4.13) contains two principal values in the same integral, great care must be exercised in deriving Eq. (4.14). In particular,[20,22]

$$(x-u \pm i\epsilon)^{-n} \underset{\epsilon \to 0+}{\to} \frac{\mathscr{P}}{(x-u)^n} \mp i\pi \frac{(-)^{n-1}}{(n-1)!} \frac{d^{(n-1)}}{dx^{(n-1)}} \delta(x-u) \quad \text{(B12)}$$

is valid only for the integrand of a Cauchy integral, which by definition contains no other singularities.[2] In the formalism leading to the derivation of Eq. (4.10), we carefully studied how derivatives of higher-order principal values are to be handled and we showed that two principal values are easily manageable provided that each is in a separate Cauchy integral. In what follows, we will treat the case in which the two principal values occur in only one of the integrals comprising a double integral.

Equation (4.14) will now be derived using a generalization of the method developed in Ref. 2 for the case of simple poles. First, we derive the functions

$$\Phi(z) = \int_{-\infty}^{\infty} dx \frac{1}{(x-z)^n} \int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-x)^m} f(x,y) \quad \text{(B13)}$$

and

$$\Psi(z) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \frac{1}{(x-z)^n} \frac{\mathscr{P}}{(y-x)^m} f(x,y), \quad \text{(B14)}$$

where z is complex. We now make use of a very important property of integrals containing singular functions. Since Eqs. (B13) and (B14) contain only *one* principal value, the order of integrations may be interchanged,[1,2] giving

$$\Phi(z) = \Psi(z). \quad \text{(B15)}$$

If we let

$$\Phi^{(\pm)}(u) \equiv \lim_{\epsilon \to 0} \Phi(u \pm i\epsilon), \quad \text{(B16a)}$$

$$\Psi^{(\pm)}(u) \equiv \lim_{\epsilon \to 0} \Psi(u \pm i\epsilon). \quad \text{(B16b)}$$

it is clear that

$$\Phi^{(+)}(u) + \Phi^{(-)}(u) = \Psi^{(+)}(u) + \Psi^{(-)}(u). \quad \text{(B17)}$$

Since $\Phi(z)$ is a Cauchy integral with respect to the x integration, we see from Eqs. (4.9), (B12), and (B13) that

$$\Phi^{(+)}(u) + \Phi^{(-)}(u) = 2A(u). \quad \text{(B18)}$$

However, $\Psi(z)$ is not a Cauchy integral with respect to the y integration: From Eq. (B11) we can convert it to sums of Cauchy integrals, namely,

$$\Psi(z) = k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \int_{-\infty}^{\infty} dy \Bigg[\frac{1}{(y-z)^{m+r}} H_{nm}^{(r)} \rho_{nr}(z,y) - (-)^{n+r} \sum_{r'=0}^{m-1+r} \frac{1}{(y-z)^{r'+1}} G_{nm}^{(rr')} R_{nm}^{(r')}(y) \Bigg], \quad \text{(B19)}$$

where

$$\rho_{nr}(z,y) = \int_{-\infty}^{\infty} dx \frac{1}{(x-z)^{n-r}} f(x,y), \quad \text{(B20)}$$

and

$$R_{nm}^{(r')}(y) = \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(y-x)^{n+m-r'-1}} f(x,y). \quad \text{(B21)}$$

Now $\Psi(z)$ is expressed in terms of Cauchy integrals and, from Eq. (B12), Eq. (B16) becomes

$$\Psi^{(\pm)}(u) = k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \Big\{ H_{nm}^{(r)} \Big[\Big[\int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)^{m+r}} \rho_{nr}^{(\pm)}(u,y) \pm i\pi \frac{(-)^{m+r-1}}{(m+r-1)!} \int_{-\infty}^{\infty} dy\, \rho_{nr}^{(\pm)}(u,y)$$

$$\times \frac{d^{(m+r-1)}}{dy^{(m+r-1)}} \delta(y-u) \Big] - (-)^{n+r} \sum_{r'=0}^{m-1+r} G_{nm}^{(rr')} \Big[\Big[\int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)^{r'+1}} R_{nm}^{(r')}(y)$$

$$\pm i\pi \frac{(-)^{r'}}{r'!} \int_{-\infty}^{\infty} dy\, R_{nm}^{(r')}(y) \frac{d^{(r')}}{dy^{(r')}} \delta(y-u) \Big] \Big] \Big\}, \quad \text{(B22)}$$

so that

$$\Psi^{(+)}(u) + \Psi^{(-)}(u) = k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \Big\{ H_{nm}^{(r)} \Big[\Big[\int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)^{m+r}} (\rho_{nr}^{(+)}(u,y) + \rho_{nr}^{(-)}(u,y))$$

$$+ \frac{i\pi}{(m+r-1)!} \int_{-\infty}^{\infty} dy\, \delta(y-u) \frac{d^{(m+r-1)}}{dy^{(m+r-1)}} (\rho_{nr}^{(+)}(u,y) - \rho^{(-)}(u,y)) \Big]$$

$$- 2(-)^{n+r} \sum_{r'=0}^{m-1+r} G_{nm}^{(rr')} \int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)^{r'+1}} R_{nm}^{(r')}(y) \Big\}, \quad \text{(B23)}$$

where we have integrated by parts $m + r - 1$ times in the second line of Eq. (B23).

We must now evaluate the integrals $\rho_{nr}^{(\pm)}(u,y)$. From Eqs. (B12) and (B20) we see that

$$\rho_{nr}^{(\pm)}(u,y) = \lim_{\epsilon \to 0+} \rho_{nr}(u \pm i\epsilon, y) = \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)^{n-r}} f(x,y) \pm \frac{i\pi}{(n-r-1)!} \Big[\frac{\partial^{(n-r-1)}}{\partial x^{(n-r-1)}} f(x,y) \Big]_{x=u} \quad \text{(B24)}$$

and in the second term we have integrated by parts $(n - r - 1)$ times and then integrated over the delta function. Thus we have

$$\rho^{(+)}(u,y) + \rho^{(-)}(u,y) = 2 \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)^{n-r}} f(x,y), \quad \text{(B25a)}$$

$$\rho^{(+)}(u,y) - \rho^{(-)}(u,y) = \frac{2i\pi}{(n-r-1)!} \Big[\frac{\partial^{(n-r-1)}}{\partial x^{(n-r-1)}} f(x,y) \Big]_{x=u} \quad \text{(B25b)}$$

and substituting Eqs. (B25) into Eq. (B23) we obtain

$$\Psi^{(+)}(u) + \Psi^{(-)}(u) = 2k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \Big[H_{nm}^{(r)} \int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)^{m+r}} \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)^{n-r}} f(x,y)$$

$$- (-)^{n+r} \sum_{r'=0}^{m-1+r} G_{nm}^{(rr')} \int_{-\infty}^{\infty} dy \frac{\mathscr{P}}{(y-u)^{r'+1}} \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(y-x)^{n+m-r'-1}} f(x,y) \Big] \quad \text{(B26)}$$

$$- 2\pi^2 k_{nm} \sum_{r=0}^{n-1} C_r^{n-1} \Big[\frac{\partial^{(m+r-1)}}{\partial y^{(m+r-1)}} \frac{\partial^{(n-r-1)}}{\partial x^{(n-r-1)}} f(x,y) \Big]_{x=y=u},$$

where we have also used Eqs. (B9) and (B21). The terms in the first two lines of Eq. (B26) can be then combined using Eq. (B11), so that

$$\Psi^{(+)}(u) + \Psi^{(-)}(u) = 2 \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)^n} \times \frac{\mathscr{P}}{(y-x)^m} f(x,y) - 2\pi^2 F_{nm}(u,u) \quad \text{(B27)}$$

and F_{nm} is defined in Eq. (4.11). Finally, from Eqs. (B17), (B18), and (B27), we obtain

$$A(u) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dx \frac{\mathscr{P}}{(x-u)^n} \frac{\mathscr{P}}{(y-x)^m} f(x,y) - \pi^2 F_{nm}(u,u), \quad \text{(B28)}$$

which, after comparing with Eq. (4.12), establishes Eq. (4.14).

APPENDIX C: POOR MAN'S DERIVATION OF THE PB THEOREM

In this Appendix we give a short, nonrigorous (physicist's) proof of the PB theorem.[1,2] We make some assumptions about analyticity which are more restrictive than the conditions required in the general theorem. However, the spirit of our derivation does parallel that of a more rigorous proof, e.g., the general derivation in Appendix B.[2]

We begin by noting that for a function $f(x)$ that is well behaved along the real axis[20,24] we have

$$\int_{-\infty}^{\infty} \frac{f(x)dx}{x-u \pm i\epsilon} \underset{\epsilon \to 0+}{\to} \int_{-\infty}^{\infty} f(x)dx \times \Big[\frac{\mathscr{P}}{(x-u)} \mp i\pi\delta(x-u) \Big]. \quad \text{(C1)}$$

It is well known[1] that the result (C1) does not trivially generalize to two poles. For example, if we define the functions

$$J_0(u_1,u_2;\epsilon_1,\epsilon_2)=\int_{-\infty}^{\infty}\frac{f(x)dx}{(x-u_1-i\epsilon_1)(x-u_2+i\epsilon_2)},\tag{C2}$$

$$K_0(u_1,u_2)=\int_{-\infty}^{\infty}f(x)dx\left[\frac{\mathscr{P}}{(x-u_1)}+i\pi\delta(x-u_1)\right]\times\left[\frac{\mathscr{P}}{(x-u_2)}-i\pi\delta(x-u_2)\right],\tag{C3}$$

then

$$\lim_{\epsilon_1,\epsilon_2\to 0+}J_0(u_1,u_2;\epsilon_1,\epsilon_2)\neq K_0(u_1,u_2).$$

Our aim is to find a relationship between J and K which is the natural generalization of Eq. (C1). We mention that $f(x)$ stands for $f(x,u_2)$ in order to allow for a generalization to a function of two variables later in this Appendix.

Next, we evaluate Eq. (C2) for the case in which $f(z)$ is analytic in the upper half-plane, giving

$$J_0(u_1u_2;\epsilon_1\epsilon_2)=2\pi i\frac{f(u_1+i\epsilon_1)}{u_1-u_2+i(\epsilon_1+\epsilon_2)},\tag{C4}$$

which becomes

$$\lim_{\epsilon_1,\epsilon_2\to 0+}J_0(u_1,u_2,\epsilon_1,\epsilon_2)=2\pi if(u_1)[\mathscr{P}/(u_1-u_2)]+2\pi^2f(u_1)\delta(u_1-u_2).\tag{C5}$$

Now, from the residue theorem for poles on the contour (principal values)[25] we find that

$$\int_{-\infty}^{\infty}\frac{\mathscr{P}}{(x-u_1)(x-u_2)}f(x)dx=\pi i\left[\frac{f(u_1)}{(u_1-u_2)}+\frac{f(u_2)}{(u_2-u_1)}\right].\tag{C6}$$

Equation (C6) is also valid if $u_1=u_2$ since[25,26]

$$\int_{-\infty}^{\infty}\frac{\mathscr{P}}{(x-u_1)^2}f(x)dx=\pi i\left(\frac{df(x)}{dx}\right)_{x=u_1}=\pi i\lim_{u_2\to u_1}\left[\frac{f(u_1)-f(u_2)}{u_1-u_2}\right].$$

Thus there is no loss in generality in writing

$$\int_{-\infty}^{\infty}\frac{\mathscr{P}}{(x-u_1)(x-u_2)}f(x)dx=\pi i[f(u_1)-f(u_2)]\frac{\mathscr{P}}{(u_1-u_2)}.\tag{C7}$$

Substituting Eq. (C7) into Eq. (C3) we obtain

$$K_0(u_1,u_2)=2\pi if(u_1)[\mathscr{P}/(u_1-u_2)]+\pi^2f(u_1)\delta(u_1-u_2)\tag{C8}$$

and comparing Eqs. (C5) and (C8) we see that

$$\lim_{\epsilon_1,\epsilon_2\to 0+}J_0(u_1,u_2;\epsilon_1\epsilon_2)=K_0(u_1,u_2)+\pi^2f(u_1)\delta(u_1-u_2).\tag{C9}$$

The result (C9) also holds if $f(z)$ is analytic in the lower half-plane; thus it is valid if $f(z)$ is the sum of functions that are separately analytic in the upper and lower half-planes.

Now, in Eqs. (C2), (C3), and (C8) we introduce another variable y, letting $f(x)=f(x,u_2)\to f(x,y)$, $u_1\to u$, and $u_2\to y$. We also integrate both sides of Eq. (C9) over y and find that

$$I(u)=\lim_{\epsilon_1,\epsilon_2\to 0+}\int_{-\infty}^{\infty}dy\int_{-\infty}^{\infty}\frac{f(x,y)dx}{(x-u-i\epsilon_1)(x-y+i\epsilon_2)}\tag{C10a}$$

$$=\int_{-\infty}^{\infty}dy\int_{-\infty}^{\infty}dx\frac{\mathscr{P}}{(x-u)(x-y)}f(x,y)+i\pi\int_{-\infty}^{\infty}dy\frac{\mathscr{P}}{(u-y)}f(u,y)-i\pi\int_{-\infty}^{\infty}dy\frac{\mathscr{P}}{(y-u)}f(y,y)+2\pi^2f(u,u).\tag{C10b}$$

However, in the double integral on the rhs of Eq. (C10a) we can reverse the order of integration before we take the limit,[2] with

$$I(u)=\lim_{\epsilon_1,\epsilon_2\to 0+}\int_{-\infty}^{\infty}\frac{dx}{(x-u-i\epsilon_1)}F(x,\epsilon_2),\tag{C11}$$

$$F(x;\epsilon_2)=\int_{-\infty}^{\infty}\frac{dy}{(x-y+i\epsilon_2)}f(x,y).\tag{C12}$$

Equation (C12) is of the form given in Eq. (C1), so that

$$\lim_{\epsilon_2\to 0+}F(x;\epsilon_2)=\int_{-\infty}^{\infty}dy\frac{\mathscr{P}}{(x-y)}f(x,y)-i\pi f(x,x).\tag{C13}$$

Similarly, in Eq. (C11) we have

$$I(u)=\int_{-\infty}^{\infty}dx\frac{\mathscr{P}}{(x-u)}\lim_{\epsilon_2\to 0+}F(x;\epsilon_2)+i\pi\lim_{\epsilon_2\to 0+}F(u;\epsilon_2)$$
$$=\int_{-\infty}^{\infty}dx\frac{\mathscr{P}}{(x-u)}\int_{-\infty}^{\infty}\frac{\mathscr{P}}{(x-y)}f(x,y)dy-i\pi\int_{-\infty}^{\infty}dx\frac{\mathscr{P}}{(x-u)}f(x,x)+i\pi\int_{-\infty}^{\infty}dy\frac{\mathscr{P}}{(u-y)}f(u,y)+\pi^2f(u,u).\tag{C14}$$

Then comparing Eq. (C14) with Eq. (C10b) we obtain Eq. (1.1), the PB theorem.

We have assumed that $f(x,y)$ consists of a sum of terms which with respect to x are separately analytic in the upper and lower half-planes. No special properties are assumed with respect to the dependence on y other than that the integrands are well behaved along the entire real axis. As mentioned, the general theorem is rigorously proved without any assumptions of analyticity.[2] Nevertheless, as we have repeatedly emphasized in earlier parts of this paper, the validity of various principal-value expressions is intimately connected to analyticity (and to causality).[13]

Finally, we define the functions

$$J_\pm(u_1,u_2;\epsilon_1,\epsilon_2)=\int_{-\infty}^{\infty}\frac{f(x)dx}{(x-u_1\pm i\epsilon_1)(x-u_2\pm i\epsilon_2)},\tag{C15}$$

$$K_\pm(u_1,u_2)=\int_{-\infty}^{\infty}f(x)dx\left[\frac{\mathscr{P}}{(x-u_1)}\mp i\pi\delta(x-u_1)\right]\times\left[\frac{\mathscr{P}}{(x-u_2)}\mp i\pi\delta(x-u_2)\right]\tag{C16}$$

corresponding to poles infinitesimally on the same side of the real axis. Thus from the analysis used to prove Eq. (C9) we find that

$$\lim_{\epsilon_1,\epsilon_2\to 0+} J_0(u_1,u_2;\epsilon_1,\epsilon_2) = K_\pm(u_1,u_2) + \pi^2 f(u_1)\delta(u_1-u_2), \qquad \text{(C17)}$$

a result that is independent of how one takes the limit $\epsilon_1,\epsilon_2\to 0+$. We see that (C9) and (C17) establish the comprehensive relation[22]

$$[(x-u_1-is_1\epsilon_1)(x-u_2-is_2\epsilon_2)]^{-1} \underset{\epsilon_1,\epsilon_2\to 0+}{\to} \left[\frac{\mathscr{P}}{(x-u_1)} + i\pi s_1\delta(x-u_1)\right]\left[\frac{\mathscr{P}}{(x-u_2)} + i\pi s_2\delta(x-u_2)\right] + \pi^2\delta(x-u_1)\delta(x-u_2), \qquad \text{(C18)}$$

where $s_i = \pm 1$. This limit applies to any integral in which the lhs of Eq. (C18) multiplies a function which is analytic in the upper or lower half-plane. Also, note that if $s_1 = s_2$ (with the poles on the same side of the real axis), the two terms in Eq. (C18) involving $\delta(x-u_1)\delta(x-u_2)$ cancel. The only way one can obtain a true delta function singularity upon integrating is when the poles are on opposite sides of the real axis. [See Eq. (C5).] However, in deriving Eq. (C18), it was necessary to consider the case $u_1 = u_2$; otherwise, the terms involving $\delta(x-u_1)\delta(x-u_2)$ never contribute. (Indeed, such terms were crucial for proving the PB theorem.) We mention, also, that for $u_1 = u_2$ the term $\mathscr{P}/(x-u_1)^2$ must be evaluated using Eq. (4.1).[20] Thus Eq. (C18) is a generalization of Eq. (C1) for the case in which there are two poles in the denominator of the integrand.[22]

[1]R. Balescu, *Statistical Mechanics of Charged Particles* (Interscience, New York, 1963), p. 399.

[2]N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen, 1953), pp. 56–61; this reference gives a very rigorous proof of the PB theorem.

[3]P. J. Siemens, M. Soyeur, G. D. White, L. J. Lantto, and K. T. R. Davies, Phys. Rev. C **40**, 2641 (1989).

[4]J. Linhard, Mat. Fys. Medd. **28**, (8) (1954).

[5]A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971), pp. 158–162, 171–197, 291–310.

[6]C. Itzykson and J.-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980), pp. 365–370.

[7]G. D. White, K. T. R. Davies, and P. J. Siemens, Ann. Phys. **187**, 198 (1988); this paper contains additional references regarding nuclear response functions.

[8]S. S. Schweber, H. A. Bethe, and F. de Hoffmann, *Mesons and Fields* (Row, Peterson, White Plains, NY, 1955), Vol. I, pp. 266–272.

[9]Reference 6, pp. 32–36.

[10]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part I, pp. 370–372.

[11]A. Erdélyi, Ed., Bateman Manuscript Project, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. II, pp. 243–262.

[12]Reference 1, p. 110. See, also, B. D. Fried and S. D. Conte, *The Plasma Dispersion Function* (Academic, New York, 1961).

[13]E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford U.P., New York, 1937), Theorem 95.

[14]K. T. R. Davies, G. D. White, P. J. Siemens, and M. Soyeur (to be published).

[15]Reference 5, pp. 72–79.

[16]M. Hirata, J. H. Koch, F. Lenz, and E. J. Moniz, Ann. Phys. **120**, 205 (1979); F. Lenz, M. Thies, and Y. Horikawa, *ibid.* **140**, 266 (1982); J. H. Koch, E. J. Moniz, and N. Ohtsuka, *ibid.* **154**, 99 (1984); B. Karaoglu and E. J. Moniz, Phys. Rev. C **33**, 974 (1986); L. Heller, S. Kumano, J. C. Martinez, and E. J. Moniz, *ibid.* **35**, 35 (1987).

[17]E. Oset, H. Toki, and W. Weise, Phys. Rep. **83**, (4), 281 (1982); see this paper for additional references on the delta-hole model.

[18]M. Baranger, in *Nuclear Structure and Nuclear Reactions, Proceedings of the International School of Physics "Enrico Fermi," Course XL, Varenna, 1967,* edited by M. Jean and R. A. Ricci (Academic, New York, 1969).

[19]J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975), 2nd ed., pp. 284–312.

[20]K. T. R. Davies and R. W. Davies, Can. J. Phys. **67**, 759 (1989).

[21]I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964); M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions,* (Cambridge U.P., Cambridge, 1962), pp. 34–39; P. J. David and P. Rabinowitz, *Methods of Numerical Integration* (Academic, New York, 1984), 2nd ed., pp. 11–14, 188–190.

[22]R. G. Newton and R. Shtokhamer, Phys. Rev. A **14**, 642 (1976).

[23]L. Brand, *Advanced Calculus* (Wiley, New York, 1955), pp. 109–111.

[24]E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1962), pp. 501–502.

[25]Reference 10, p. 368.

[26]Reference 10, p. 411.

Classical Liouville completely integrable systems associated with the solutions of Boussinesq–Burgers' hierarchy

Zhuquan Gu
Shijiazhuang Railway Institute, Shijiazhuang, People's Republic of China

(Received 1 June 1989; accepted for publication 13 December 1989)

Two new finite-dimensional completely integrable systems in the Liouville sense are obtained. The solutions of Boussinesq–Burgers' hierarchy are generated by using involutive solutions of the commutable flows in the completely integrable systems.

I. INTRODUCTION

The link between finite-dimensional completely integrable systems in the Liouville sense and infinite-dimensional soliton systems has been an important topic of concern in recent years.[1] However, known classical Liouville completely integrable systems are very few;[2] the key point is whether an N-involutive system of the Hamiltonian functions in R^{2N} can be obtained. In the present paper, according to Cao's method of finding the finite-dimensional completely integrable systems,[3] we study the hierarchy of the Boussinesq–Burgers' equation. Two new finite-dimensional completely integrable systems in the Liouville sense are obtained by means of the nonlinearization of the Lax pairs of the hierarchy of the Boussinesq–Burgers' equation; therefore the solutions of the Boussinesq–Burgers' hierarchy are generated by using the involutive solutions of the commutable flows in the completely integrable systems. In the Boussinesq–Burgers' system, a lot of work has been done, e.g., Hamiltonian structure, the Painlevé property, the Hirota bilinear form, pole expansions and a related many-body problem, Bäcklund transformations, some particular solutions, etc.[4–6] We study the following problems.

(1) How do we get the Lax pair of the high-order Boussinesq–Burgers' equation?

(2) What is the link between the Boussinesq–Burgers' hierarchy and finite-dimensional completely integrable systems in the Liouville sense?

(3) How do we use the Lax pairs of the Boussinesq–Burgers' hierarchy to generate the solutions of the high-order Boussinesq–Burgers' equations?

In this paper, we get the hierarchy of the Boussinesq–Burgers' equation and their Lax pairs by using the spectral problem and the commutator of the differential operators. We give two constraints between the potential and eigenfunctions for the spectral problem; then the Lax pairs of the high-order Boussinesq–Burgers' equations are nonlinearized. It is proved that the Lax pairs when nonlinearized become commutable flows of the finite-dimensional completely integrable systems in the Liouville sense, so that each of the high-order Boussinesq–Burgers' equations becomes an exactly involutive condition of the commutable flows; in particular, as dynamic systems, the nonlinear Lax equation systems are all completely integrable systems in the Liouville sense. Finally, the solutions of the Boussinesq–Burgers' hierarchy are obtained by means of the involutive solutions of commutable flows.

II. EVOLUTION EQUATION HIERARCHY AND LAX REPRESENTATIONS

We consider the following spectral problem:

$$L\begin{pmatrix}p\\q\end{pmatrix}=\begin{pmatrix}\partial+\frac{1}{4}u & -\frac{1}{4}(w+u_x)\\ -1 & -\partial+\frac{1}{4}u\end{pmatrix}\begin{pmatrix}p\\q\end{pmatrix}=\lambda\begin{pmatrix}p\\q\end{pmatrix}, \quad (2.1)$$

where u and w are called potentials, λ is the eigenparameter, and $\partial=\partial/\partial x$.

Propostion 1: If the operators K and J take the forms[5]

$$K=\begin{pmatrix}2\partial & \partial u\\ u\partial & 2\partial^3+2w\partial+w_x\end{pmatrix},\quad J=4\begin{pmatrix}0 & \partial\\ \partial & 0\end{pmatrix}, \quad (2.2)$$

and $(p_j,q_j)^T$ and λ_j satisfy the linear equation system (2.1), then

$$K\begin{pmatrix}2(p_jq_j+q_jq_{jx})\\ -q_j^2\end{pmatrix}=\lambda_jJ\begin{pmatrix}2(p_jq_j+q_jq_{jx})\\ -q_j^2\end{pmatrix}. \quad (2.3)$$

Proof: From (2.1) and the definitions of K and J, through direct computation, (2.3) holds.

Remark: If u and w are real and λ_j is an arbitrary real given, then the linear equation system (2.1) has the nontrivial real solution $(p_j,q_j)^T$.

Now, let

$$\partial\partial^{-1}=\partial^{-1}\partial=1. \quad (2.4)$$

We define Lenart's sequence G_m, $m=0,1,2,\ldots$, by means of the recursion relations

$$G_{-1}=(b,4)^T,\quad G_m=(b_m,c_m)^T,\qquad JG_m=KG_{m-1},\quad m=0,1,2,\ldots, \quad (2.5)$$

where b is an arbitrary constant.

From (2.5), b_m and c_m are polynomials of u, w, u', w', u'', w'',... . If the constant term of G_m $(m=0,1,2,\ldots)$ takes zero, then G_m $(m=0,1,2,\ldots)$ is determined uniquely. In this case, we call $X_m=JG_m$ $(m=0,1,2\ldots)$ an mth-order Boussinesq–Burgers' vector field; the evolution equation

$$v_{tm}=(u,w)^T_{tm}=JG_m,\quad m=0,1,2,\ldots, \quad (2.6)$$

is called an mth-order Boussinesq–Burgers' equation.

Propostion 2: In the case of $w=u_x$, the evolution equation (2.6) becomes the Burgers'-type equation

$$u_{tm}=4\Phi^m u_x,\quad m=0,1,2,\ldots, \quad (2.7)$$

where $\Phi=\frac{1}{2}(\partial+\frac{1}{2}u+\frac{1}{2}u_x\partial^{-1})$ (Ref. 7).

Proof: If $b_m=c_{mx}$, $m=0,1,2,\ldots$, then the propostion is proved. Since $b_0=w=u_x=c_{0x}$, in the case of $m=j$, we have $b_j=c_{jx}$. Then, from $X_{j+1}=KJ^{-1}X_j$,

 0022-2488/90/061374-07$03.00

$$b_{j+1}=\tfrac{1}{2}(\tfrac{1}{2}(uc_j)_x+b_{jx})=c_{j+1\,x},$$

so that $b_m=c_{mx}$, $m=0,1,2,\ldots$.

The operators L_* and N_j are defined as

$$L_*(r)=\frac{d}{d\epsilon}\bigg|_{\epsilon=0}L(v+\epsilon r),\quad \forall r, \tag{2.8}$$

where $v=(u,w)^T$, $r=(r_1,r_2)^T$, and L is defined by (2.1), and

$$N_j=\begin{pmatrix}\frac{1}{4}uc_{j-1}-\frac{1}{2}c_{j-1\,x} & -\frac{1}{2}b_{j-1\,x}-\frac{1}{2}c_{j-1\,xx}-(w+u_x)\frac{1}{4}c_{j-1}\\ c_{j-1} & -\frac{1}{4}uc_{j-1}+\frac{1}{2}c_{j-1\,x}\end{pmatrix}+\begin{pmatrix}-c_{j-1} & 0\\ 0 & c_{j-1}\end{pmatrix}L.$$

The commutator of operators A and B is defined as follows:

$$[A,B]=AB-BA.$$

Hence we have

(i) L: $R^2\to R^2$ is one to one;

(ii) $[N_j,L]=-L_*(KG_{j-1})+L_*(JG_{j-1})L$.

Theorem 1: Set

$$V_m=-\sum_{j=0}^{m}N_jL^{m-j}.$$

Then $L_*(JG_m)=[V_m,L]$, and the evolution equation $v_{tm}=JG_m$ and the Lax form $L_{tm}=[V_m,L]$ are equivalent, for $m=0,1,2,\ldots$.

Proof: Since $JG_{-1}=0$,

$$[N_jL^{m-j},L]=[N_j,L]L^{m-j},$$

so that

$$\begin{aligned}[V_m,L]&=-\sum_{j=0}^{m}[N_j,L]L^{m-j}\\ &=L_*(KG_{m-1})-L_*(JG_{-1})L^{m+1}\\ &=L_*(JG_m),\end{aligned}$$

$$L_{tm}=L_*(v_{tm}).$$

Hence the theorem is proved.

Corollary: If α_j $(j=0,1,2,\ldots)$ is constant independent from x, then (i) The Boussinesq–Burgers'-type equation

$$v_{tm}=\sum_{j=0}^{m}\alpha_jX_j$$

can be written in the Lax form

$$L_{tm}=\left[\sum_{j=0}^{m}\alpha_jV_j,L\right];$$

and (ii) the Boussinesq–Burgers'-type equation

$$v_{tm}=\sum_{j=0}^{m}\alpha_jX_j$$

is a compatible condition for the following Lax pair:

$$L\psi=\lambda\psi,\quad \psi_{tm}=\left(\sum_{j=0}^{m}\alpha_jV_j\right)\psi\quad [\psi=(p,q)^T],$$

in the case of $\lambda_{tm}=0$, $\psi_{xtm}=\psi_{tmx}$. Particularly, the mth-order Boussinesq–Burgers' equation $v_{tm}=X_m=JG_m$ is a compatible condition for the following Lax pair:

$$L\psi=\lambda\psi, \tag{2.9a}$$

$$\psi_{tm}=V_m\psi, \tag{2.9b}$$

in the case of $\lambda_{tm}=0$, $\psi_{xtm}=\psi_{tmx}$.

From (2.9), the Boussinesq–Burgers' equation

$$\begin{pmatrix}u\\ w\end{pmatrix}_{t1}=2\begin{pmatrix}w_x+uu_x\\ u_{xxx}+(uw)_x\end{pmatrix} \tag{2.10}$$

has the following Lax pair:

$$L\psi=\lambda\psi\quad [\psi=(p,q)^T],$$

$$\psi_{t1}=V_1\psi=\begin{pmatrix}+4 & 0\\ 0 & -4\end{pmatrix}L^2\psi-\begin{pmatrix}0 & -w-u_x\\ 4 & 0\end{pmatrix}L\psi$$

$$-\begin{pmatrix}\frac{1}{4}u^2-\frac{1}{2}u_x & -\frac{1}{2}w_x-\frac{1}{2}u_{xx}-\frac{1}{4}uw-\frac{1}{4}uu_x\\ u & \frac{1}{2}u_x-\frac{1}{4}u^2\end{pmatrix}\psi,$$

i.e.,

$$\begin{aligned}&q_{xx}+(-\lambda^2+(\lambda/2)u+\tfrac{1}{4}w-\tfrac{1}{16}u^2)q=0,\\ &q_{t1}+\tfrac{1}{2}u_xq-(u+4\lambda)q_x=0.\end{aligned} \tag{2.11}$$

Let $\lambda=ik$ and $t1=-\frac{1}{2}t$. Then (2.10) and (2.11) become exactly (2) and (3) in Ref. 5.

If $F(v)$ is a functional, i.e., a numerically valued function, in general, defined on the underlying linear space, then F is called differentiable if the directional derivative

$$\lim_{\epsilon\to0}[F(v+\epsilon r)-F(v)]/\epsilon$$

exists for all v and r and is a linear functional of r. Here $v=(v_1(x),v_2(x))^T$ and $r=(r_1(x),r_2(x))^T$ are vector functions defined over the underlying interval Ω, where Ω is $(-\infty,+\infty)$ for decaying at infinity and is a double periodic interval for the periodic condition. We assume that our linear space is equipped with a $L_2(\Omega)$ scalar product $(\cdot,\cdot)_1$:

$$(v,r)_1=\int_\Omega(v_2r_2+v_1r_1)dx=\int_\Omega v\cdot r\,dx.$$

Since linear functionals can be expressed as scalar products, we can write

$$\frac{d}{d\epsilon}F(v+\epsilon r)\bigg|_{\epsilon=0}=(G_F(v),r)_1=\int_\Omega G_F(v)\cdot r\,dx,$$

where $G_F(v)$ is called the gradient of F at v with respect to the specified scalar product.

If $(p_j,q_j)^T\in L_2(\Omega)$ is a nontrivial solution of the spectral equation (2.1), and the eigenparameter corresponding to the vector $(p_j,q_j)^T$ is λ_j, then we call $(p_j,q_j)^T$ an eigenvector of (2.1) and λ_j an eigenvalue of (2.1).

Proposition 3: Suppose $(p_j,q_j)^T$ is any eigenvector of (2.1), and λ_j is an eigenvalue corresponding to the vector $(p_j,q_j)^T$. Then

(i) $$G_{\lambda_j}(v)=\frac{1}{8}\left(\int_\Omega p_jq_j\,dx\right)^{-1}\begin{pmatrix}2(p_jq_j+q_jq_{jx})\\ -q_j^2\end{pmatrix},$$

where $v=(u,w)^T$;

(ii) $KG_{\lambda_j}(v) = \lambda_j JG_{\lambda_j}(v)$,

where K and J are Hamiltonian operators;[7,8] and (iii) each of the eigenvalues for the spectral equation (2.1) is all conservation integrating of the Boussinesq–Burgers' hierarchy $\{v_m = JG_m\}$ (Refs. 9, 10).

Proof: (2.1), we have

$$\begin{pmatrix} p_j \\ q_j \end{pmatrix}_x = M_j \begin{pmatrix} p_j \\ q_j \end{pmatrix} = \begin{pmatrix} \lambda_j - \frac{1}{4}u & \frac{1}{4}(w+u_x) \\ -1 & \frac{1}{4}u - \lambda_j \end{pmatrix} \begin{pmatrix} p_j \\ q_j \end{pmatrix}.$$

Let a prime denote a derivative with respect to ϵ. Then

$$(\dot p_j q_j - p_j \dot q_j)_x = \operatorname{tr} M_j \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \dot p_j \\ \dot q_j \end{pmatrix} + (p_j, q_j) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \dot M_j \begin{pmatrix} p_j \\ q_j \end{pmatrix}.$$

Since $\operatorname{tr} M_j = 0$ and $(p_j, q_j)^T \in L_2(\Omega)$,

$$\int_\Omega (p_j, q_j) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \dot M_j \begin{pmatrix} p_j \\ q_j \end{pmatrix} dx = 0;$$

hence (i) is obtained. By (2.2) and (2.3) and the definition, (ii) is proved. According to (ii), we have

$$\begin{aligned} \lambda_{jtm} &= (G_{\lambda_j}(v), v_{tm})_1 = (G_{\lambda_j}(v), KG_{m-1})_1 \\ &= -(KG_{\lambda_j}(v), G_{m-1})_1 = -(\lambda_j JG_{\lambda_j}(v), G_{m-1})_1 \\ &= (G_{\lambda_j}(v), JG_{m-1})_1 \lambda_j = \lambda_j (G_{\lambda_j}(v), KG_{m-2})_1 \\ &= \cdots, \\ \lambda_{jtm} &= \lambda_j^{m+1} (G_{\lambda_j}(v), JG_{-1})_1 = 0, \\ & m = 0,1,2,\ldots; \end{aligned}$$

thus (iii) holds.

III. TWO CLASSICAL LIOUVILLE COMPLETELY INTEGRABLE SYSTEMS[1,2,11,12]

In order to study the complete integrability of the Lax equation system (2.9) in R^{2N}, let $\langle \cdot,\cdot \rangle$ denote the standard inner product in R^N, and set $q = (q_1, q_2, \ldots, q_N)^T$, $p = (p_1, p_2, \ldots, p_N)^T$, and $A = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$, $\lambda_1 < \lambda_2 < \cdots < \lambda_N$.

The Poisson bracket for the Hamiltonian functions F and Q in the symplectic space $(R^{2N}, dp \wedge dq)$ is defined as

$$(F,Q) = \sum_{j=1}^N \frac{\partial F}{\partial q_j}\frac{\partial Q}{\partial p_j} - \frac{\partial F}{\partial p_j}\frac{\partial Q}{\partial q_j} = \langle F_q, Q_p \rangle - \langle F_p, Q_q \rangle . \tag{3.1}$$

We call the Hamiltonian functions F and Q involutive in R^{2N} if $(F,Q) = 0$. Whether the Hamiltonian system is completely integrable in R^{2N} depends on whether the N-involutive system exists or not in R^{2N} (Refs. 2, 13).

Let

$$T_i = \sum_j^{i}{}' (\lambda_i - \lambda_j)^{-1} B_{ij}^2, \quad B_{ij} = p_i q_j - p_i q_j, \qquad \sum_j^{i}{}' \equiv \sum_{\substack{j=1 \\ j \neq 1}}^{N}, \quad i = 1,2,\ldots,N. \tag{3.2}$$

Lemma:

$$\begin{aligned} &(q_i, q_j) = (p_i, p_j) = 0; \\ &(q_i, p_j) = \delta_{ij}, \quad \delta_{ij} = \begin{cases} 0, & i \neq j, \\ 1, & i = j; \end{cases} \\ &(\langle p,p\rangle, B_{ij}) = (\langle q,q\rangle, B_{ij}) = (\langle p,q\rangle, B_{ij}) = 0; \\ &(T_i, T_j) = 0; \quad (p_k, B_{ij}) = p_j \delta_{ki} - p_i \delta_{kj}; \\ &(q_k, B_{ij}) = q_j \delta_{ki} - q_i \delta_{kj}; \\ &(p_k^2, T_i) = 4(\lambda_k - \lambda_i)^{-1} p_k p_i B_{ik}; \\ &(q_k^2, T_i) = 4(\lambda_k - \lambda_i)^{-1} q_k q_i B_{ik}; \\ &(p_k q_k, T_i) = 2(\lambda_k - \lambda_i)^{-1} (p_i q_k + p_k q_i) B_{ik}; \\ &\qquad i,j,k = 1,2,\ldots,N. \end{aligned}$$

Proof: (See Ref. 14.) In fact, from (3.1), we have

$$(p_i, p_j) = 0, \quad (q_i, q_j) = 0, \quad (q_i, p_j) = \delta_{ij} .$$

From (3.1) and (3.2),

$$(p_k, B_{ij}) = p_j \delta_{ki} - p_i \delta_{kj}, \quad (q_k, B_{ij}) = q_j \delta_{ki} - q_i \delta_{kj} ,$$

$$(\langle p,p\rangle, B_{ij}) = \sum_{k=1}^N (p_k^2, B_{ij}) = 2 \sum_{k=1}^N p_k (p_k, B_{ij}) = 0 .$$

Because $i = j$, $(T_i, T_j) = 0$. If $i \neq j$, then, through computing, we have

$$\begin{aligned} (T_i, T_j) &= 4 \sum_k^{i}{}' \sum_n^{j}{}' (\lambda_i - \lambda_k)^{-1} (\lambda_j - \lambda_n)^{-1} (B_{ij} B_{in} B_{kj} \delta_{kn} \\ &\quad + B_{ij} B_{in} B_{kn} \delta_{kj} + B_{ij} B_{kn} B_{kj} \delta_{in}) \\ &= 4 \sum_n^{j}{}' (\lambda_i - \lambda_n)^{-1} (\lambda_j - \lambda_n)^{-1} B_{in} B_{nj} \\ &\quad + 4 \sum_n^{j}{}' (\lambda_i - \lambda_j)^{-1} (\lambda_j - \lambda_n)^{-1} B_{in} B_{jn} \\ &\quad + 4 \sum_k^{i}{}' (\lambda_i - \lambda_k)^{-1} (\lambda_j - \lambda_i)^{-1} B_{kj} B_{ki} . \end{aligned}$$

Since

$$(\lambda_i - \lambda_n)^{-1} (\lambda_j - \lambda_n)^{-1} = (\lambda_i - \lambda_j)^{-1} ((\lambda_j - \lambda_n)^{-1} - (\lambda_i - \lambda_n)^{-1}),$$

Thus

$$\begin{aligned} (T_i, T_j) &= 4 B_{ij} (\lambda_i - \lambda_j)^{-1} \Big(\sum_n^{j}{}' (\lambda_j - \lambda_n)^{-1} (B_{nj} + B_{jn}) B_{in} \\ &\quad - \sum_n^{i}{}' (\lambda_i - \lambda_n)^{-1} (B_{in} + B_{ni}) B_{nj} \Big) = 0, \\ (p_k^2, T_i) &= \sum_j^{i}{}' (\lambda_i - \lambda_j)^{-1} (p_k^2, B_{ij}^2) \\ &= 4 \sum_j^{i}{}' (\lambda_i - \lambda_j)^{-1} p_k B_{ij} (p_j \delta_{ki} - p_i \delta_{kj}) \\ &= 4(\lambda_k - \lambda_i) p_k p_i B_{ik}, \end{aligned}$$

and another similarly is proved.

Theorem 2: Set

$$E_j = 2p_j^2 - 4\lambda_j p_j q_j - 4\langle q,q \rangle q_j^2 + 2T_j , \qquad j = 1,2,\ldots,N. \tag{3.3}$$

Suppose the Hamiltonian functions in R^{2N} are defined as follows:

$$F_0 = -4\langle Aq,q\rangle - 2\langle p,p\rangle - 4\langle q,q\rangle\langle p,q\rangle,$$
$$F_m = -4\langle A^{m+1}p,q\rangle - 2\langle A^m p,p\rangle - 4\langle A^m q,q\rangle\langle p,q\rangle + 2\sum_{i=1}^{m}(\langle A^{i-1}p,p\rangle\langle A^{m-i}q,q\rangle - \langle A^{i-1}p,q\rangle\langle A^{m-i}p,q\rangle), \quad (3.4)$$
$$m = 1,2,\ldots .$$

Then

(i) $\{E_j,\ j=1,2,\ldots,N\}$ is an N-involutive system;

(ii) $F_m = \sum_{j=1}^{N}\lambda_j^m E_j,\quad m=0,1,2,\ldots;$

(iii) $(F_m,F_n)=0,\quad m,n=0,1,2,\ldots,$

i.e., $(R^{2N}, dp\wedge dq, F_m)$ is a completely integrable system in the Liouville sense. In particular, the system $(R^{2N}, dp\wedge dq, \frac{1}{4}F_0)$ is a Liouville completely integrable system.

Proof: By the lemma, the following terms are zero:

$$(-2p_k^2+2T_k, -2p_j^2+2T_j),$$
$$(-4\langle q,q\rangle q_k^2, -4\langle q,q\rangle q_j^2),$$
$$(-2p_k^2, -4\lambda_j p_j q_j) + (-4\lambda_k p_k q_k, -2p_j^2),$$
$$(-4\lambda_k p_k q_k, -4\lambda_j p_j q_j),$$
$$(-2p_k^2, -4\langle p,q\rangle q_j^2) + (-4\langle p,q\rangle q_k^2, -2p_j^2) + (-4\lambda_k p_k q_k, 2T_j) + (2T_k, -4\lambda_j p_j q_j),$$
$$(-4\langle p,q\rangle q_k^2, 2T_j) + (2T_k, -4\langle p,q\rangle q_j^2).$$

Since (E_k,E_j) is the sum of above terms, we have that $(E_k,E_j)=0,\quad k,j=1,2,\ldots,N.$

Let $Q_z(y,s) = \langle (z-A)^{-1}y,s\rangle$, where$Q_z(y,s)$ is a double-linear function in R^{2N}. Now expand $Q_z(y,s)$ into Laurant form and partial fraction form as follows:

$$Q_z(y,s) = \sum_{m=0}^{\infty} z^{-m-1}\langle A^m y,s\rangle = \sum_{i=1}^{N}(z-\lambda_i)^{-1}y_i s_i.$$

Since the generating function of T_k is[11]

$$\sum_{k=1}^{N}(z-\lambda_k)^{-1}T_k = Q_z(q,q)Q_z(p,p) - Q_z(p,q)Q_z(q,p),$$

the generating function of E_k is

$$\sum_{k=1}^{N}(z-\lambda_k)^{-1}E_k = -2Q_z(p,q) - 4Q_z(Aq,p) - 4\langle p,q\rangle Q_z(q,q) + 2\sum_{k=1}^{N}(z-\lambda_k)^{-1}T_k .$$

Substituting the Laurant expansion of Q_z into the generating function of E_k, and expanding $(z-\lambda_k)^{-1}$ as a power series of z^{-1}, we have

$$\sum_{k=1}^{N}(z-\lambda_k)^{-1}E_k = \sum_{m=0}^{\infty} z^{-m-1}\sum_{k=1}^{N}\lambda_k^m E_k;$$

thus

$$F_m = \sum_{k=1}^{N}\lambda_k^m E_k .$$

The involutivity of $\{E_k\}$ implies the involutivity of $\{F_m\}$.

Theorem 3: Set

$$H = \tfrac{1}{2}(\langle q,q\rangle - 1),$$
$$e_j = 4\langle q,p\rangle q_j^2 - 4\langle q,q\rangle q_j p_j - 2T_j,\quad j=1,2,\ldots,N. \quad (3.5)$$

Suppose the Hamiltonian functions h_m, $m=1,2,\ldots$, are defined as

$$h_m = -4\langle q,q\rangle\langle A^m q,p\rangle + 4\langle q,p\rangle\langle A^m q,q\rangle - 2\sum_{i=1}^{m}(\langle A^{i-1}p,p\rangle\langle A^{m-i}q,q\rangle - \langle A^{i-1}q,p\rangle\langle A^{m-i}q,p\rangle),\quad m=1,2,\ldots; \quad (3.6)$$

then

(i) $\sum_{j=1}^{N} e_j = 0$ and $(H,e_j)=0,\ j=1,2,\ldots,N,$

where $\{H,e_j,\ j=1,2,\ldots,N-1\}$ is an N-involutive system; and

(ii) $h_m = \sum_{j=1}^{N}\lambda_j^m e_j$ and $(h_m,h_n)=0,\quad m,n=1,2,\ldots,$

where the system $(R^{2N}, dp\wedge dq, h_m)$ is a completely integrable system in the Liouville sense.

Proof: By the lemma, $(H,e_j)=0,\quad j=1,2,\ldots,N.$ Since the terms

$$(4\langle q,p\rangle q_k^2, \langle q,p\rangle q_j^2),$$
$$(-4\langle q,q\rangle q_k p_k, 4\langle q,p\rangle q_j^2) + (-4\langle q,q\rangle q_k p_k, -4\langle q,q\rangle q_j p_j) + (4\langle q,p\rangle q_k^2, -4\langle q,q\rangle q_j p_j),$$
$$(-2T_k, 4\langle q,p\rangle q_j^2) + (4\langle q,p\rangle q_k^2, -2T_j),$$
$$(-2T_k, -2T_j),$$
$$(-2T_k, -4\langle q,q\rangle q_j p_j) + (-4\langle q,q\rangle q_k p_k, -2T_j)$$

are zero, and (e_k,e_j) is the sum of the above terms, $(e_k,e_j)=0,\ k,j=1,2,\ldots,N.$ Since

$$\sum_{j=1}^{N}T_j = 0,$$

we obtain

$$\sum_{j=1}^{N}e_j = 0.$$

In a way similar to (ii) and (iii) of Theorem 2, (ii) can be proved.

Now, according to Moser's method[11] of constrained Hamiltonian systems, we consider the following constraint condition:

$$H = \tfrac{1}{2}(\langle q,q\rangle - 1) = 0,\quad G = \langle q,p\rangle = 0 . \quad (3.7)$$

In Sec. IV we shall prove that the dynamic systems corresponding to the Hamiltonian systems (3.6) become the dynamic systems corresponding to the following Hamiltonian systems (3.8) in the case of the constraint condition (3.7):

$$H_m = -4\langle A^m q,p\rangle + 4\langle q,p\rangle\langle A^m q,q\rangle - 2\sum_{i=1}^{m}(\langle A^{i-1}p,p\rangle\langle A^{m-i}q,q\rangle - \langle A^{i-1}q,p\rangle\langle A^{m-i}q,p\rangle), \quad m=1,2,\ldots. \tag{3.8}$$

Theorem 4: Set

$$g_j = -4q_jp_j + 4\langle q,p\rangle q_j^2 - 2T_j, \tag{3.9}$$

$$a_j = (e_j,\langle q,p\rangle)|_{H=0,\ G=0}, \tag{3.10}$$

$$f_m = (h_m,\langle q,p\rangle)|_{H=0,\ G=0}, \tag{3.11}$$

$j=1,2,\ldots,N, \quad m=1,2,\ldots.$

Then

(i) $(g_k,g_j)=0, \quad k,j=1,2,\ldots,N;$

(ii) $H_m = \sum_{j=1}^{N}\lambda_j^m g_j, \quad (H_m,H_n)=0, \quad m,n=1,2,\ldots;$

(iii) $g_j = e_j - a_jH, \quad H_m = h_m - f_mH, \quad j=1,2,\ldots,N; \quad m=1,2,\ldots.$

Therefore the $\{H_m\}$ are completely integrable systems in the Liouville sense.

Proof: In a manner similar to the proofs of Theorems 2 and 3, by means of the lemma through direct computation, the theorem is obtained.

IV. NONLINEARIZATION OF LAX PAIRS AND SOLUTIONS OF BOUSSINESQ–BURGERS' HIERARCHY

In the symplectic space $(R^{2N}, dp\wedge dq)$, the canonical equation of the Hamiltonian function F is defined as follows:

$$\begin{pmatrix}p\\q\end{pmatrix}_{tn} = \begin{pmatrix}0 & -I\\ I & 0\end{pmatrix}\begin{pmatrix}\dfrac{\partial F}{\partial p}\\ \dfrac{\partial F}{\partial q}\end{pmatrix}, \tag{4.1}$$

where I is the $N\times N$ unit matrix.

Let g_F^{tn} denote the solution operator of the initial value problem for the dynamic system (4.1). Then the solution of (4.1) can be expressed as

$$\begin{pmatrix}p(tn)\\q(tn)\end{pmatrix} = g_F^{tn}\begin{pmatrix}p(0)\\q(0)\end{pmatrix},$$

where $(p(0),q(0))^T$ is an arbitrary initial value. The operator hierarchy $\{g_F^{tn}\}$ is called the Hamiltonian phase flow of the dynamic system (4.1), or F flows. Therefore we have

$$(F,Q) = \frac{d}{dtn}\bigg|_{tn=0} F\left\{g_Q^{tn}\begin{pmatrix}p(tn)\\q(tn)\end{pmatrix}\right\}.$$

According to Proposition 1 and the remark, suppose real λ_j and the real vector $(p_j,q_j)^T$ satisfy the linear equation system (2.1), and $j=1,2,\ldots,N$, $\lambda_1<\lambda_2<\cdots<\lambda_N$, $P=(p_1,p_2,\ldots,p_N)^T$, $q=(q_1,q_2,\ldots,q_N)^T$, and $A=\operatorname{diag}(\lambda_1,\lambda_2,\ldots,\lambda_N)$. Then (2.1) can be written in the following form:

$$\begin{pmatrix}p\\q\end{pmatrix}_x = \begin{pmatrix}A-\frac14 uI & -\frac14(w+u_x)I\\ -I & -A+\frac14 uI\end{pmatrix}\begin{pmatrix}p\\q\end{pmatrix}. \tag{4.2}$$

Now, we consider the following two constraints:

$$G_0 = \begin{pmatrix}w\\u\end{pmatrix} = -4\begin{pmatrix}-2(\langle q,p\rangle + \langle q,q_x\rangle)\\ \langle q,q\rangle\end{pmatrix} \tag{4.3}$$

and

$$G_{-1} = \begin{pmatrix}0\\4\end{pmatrix} = 4\begin{pmatrix}-2(\langle q,p\rangle + \langle q,q_x\rangle)\\ \langle q,q\rangle\end{pmatrix}. \tag{4.4}$$

We call (4.3) the Bargmann constraint and (4.4) the Neumann constraint for the spectral problem (2.1).

By (4.2), (4.3) can be written

$$\begin{pmatrix}w\\u\end{pmatrix} = -4\begin{pmatrix}+2(\langle Aq,q\rangle + \langle q,q\rangle^2)\\ \langle q,q\rangle\end{pmatrix} \tag{4.5}$$

and (4.4) can be written

$$\langle q,q\rangle = 1, \quad \langle q,p\rangle = 0. \tag{3.7'}$$

In the case of the constraint condition (4.5), by making use of the equation $KG_{m-1}=JG_m$, i.e., $G_k=(J^{-1}K)^kG_0$, $k=0,1,2,\ldots$, and using (2.3) and (2.4), the Lax pair (2.9) of the mth order Boussinesq–Burgers' equation $v_{tm}=JG_m$ $(=KG_{m-1})$ is nonlinearized as follows:

$$(F_0):\quad \begin{cases}p_x = Ap + \langle q,q\rangle P + 2\langle q,p\rangle q,\\ q_x = -p - Aq - \langle q,q\rangle q;\end{cases}$$

$$(F_m):\quad \begin{cases}p_{tm} = 4A^{m+1}p + 8\langle q,p\rangle A^m q + 4\langle A^m q,q\rangle p + 4\sum_{i=1}^{m}(\langle A^{i-1}q,p\rangle A^{m-i}p - \langle A^{i-1}p,p\rangle A^{m-i}q),\\ q_{tm} = -4A^{m+1}q - 4A^m p - 4\langle A^m q,q\rangle q - 4\sum_{i=1}^{m}(\langle A^{i-1}q,p\rangle A^{m-i}q - \langle A^{i-1}q,q\rangle A^{m-i}p),\end{cases}$$

$m=1,2,3,\ldots.$

In the case of the constraint condition (3.7), by using (2.3), (2.4), $G_0=(w,u)^T$, and $G_k=(J^{-1}K)^kG_0$, $k=0,1,2,\ldots$, we have

$$G_0 = \begin{pmatrix}w\\u\end{pmatrix} = 4\begin{pmatrix}\langle p,p\rangle + 2\langle A^2q,q\rangle + 2\langle Aq,p\rangle - 2\langle Aq,q\rangle^2\\ \langle Aq,q\rangle\end{pmatrix}. \tag{4.6}$$

Therefore the Lax pair (2.9) of the mth-order Boussinesq–Burgers' equation $v_{tm}=JG_m$ $(=KG_{m-1})$ is nonlinearized as follows:

$$(H_1):\quad \begin{cases}p_x = Ap - \langle Aq,q\rangle p + \langle p,p\rangle q - 2\langle q,p\rangle Aq - \langle q,p\rangle p,\\ q_x = -Aq + \langle Aq,q\rangle q - \langle q,q\rangle p + \langle q,p\rangle q,\end{cases}$$

$$H=\tfrac{1}{2}(\langle q,q\rangle-1)=0,\quad G=\langle q,p\rangle=0;$$

$$(H_{m+1}):\ \begin{cases} p_{tm}=4\langle q,q\rangle A^{m+1}p-4\langle A^{m+1}q,q\rangle p-4\sum_{j=0}^{m}(\langle A^jq,p\rangle A^{m-j}p-\langle A^jp,p\rangle A^{m-j}q),\\ q_{tm}=4\langle A^{m+1}q,q\rangle q-4\langle q,q\rangle A^{m+1}q+4\sum_{j=0}^{m}(\langle A^jq,p\rangle A^{m-j}q-\langle A^jq,q\rangle A^{m-j}p),\end{cases}$$

$$H=\tfrac{1}{2}(\langle q,q\rangle-1)=0,\quad G=\langle q,p\rangle=0,\quad m=1,2,3,\ldots.$$

According to the previous discussion, we obtain a new propostion as follows.

Proposition 4: Let $4to=x$. Then, as dynamic systems, the Lax pairs (F_m) nonlinearized in the case of constraint (4.5) become exactly the following Hamiltonian canonical equation systems:

$$(F_m):\ \begin{pmatrix}p_{tm}\\ q_{tm}\end{pmatrix}=\begin{pmatrix}0 & -I\\ I & 0\end{pmatrix}\begin{pmatrix}\dfrac{\partial F_m}{\partial p}\\ \dfrac{\partial F_m}{\partial q}\end{pmatrix},\quad m=0,1,2,3,\ldots, \tag{4.7}$$

where the F_m, $m=0,1,2,\ldots$, are defined by (3.4). The Lax pairs (H_m) nonlinearized by the constraint (3.7) become exactly the following Hamiltonian canonical equation systems:

$$(H_{m+1}):\ \begin{pmatrix}p_{tm}\\ q_{tm}\end{pmatrix}=\begin{pmatrix}0 & -I\\ \mathrm{I} & 0\end{pmatrix}\begin{pmatrix}\dfrac{\partial H_{m+1}}{\partial p}\\ \dfrac{\partial H_{m+1}}{\partial q}\end{pmatrix},\quad m=0,1,2,\ldots, \tag{4.8}$$

where the H_{m+1}, $m=0,1,2,..$, are defined by (3.8).

According to Theorems 2 and 4, the F_m, $m=0,1,2,\ldots$, and the H_m, $m=1,2,\ldots$, are all completely integrable systems, so that two arbitrary Hamiltonian canonical equations (F_m) and (F_n) are compatible $(m,n=0,1,2,\ldots)$, and two arbitrary Hamiltonian canonical equations (H_m) and (H_n) are compatible $(m,n=1,2,\ldots)$; therefore the Hamiltonian phase flows $g_{F_m}^{tm}$ and $g_{F_n}^{tn}$ are commutable, and the Hamiltonian phase flows $g_{H_{m+1}}^{tm}$ and $g_{H_{n+1}}^{tn}$ are commutable. Now, we arbitrarily choose an initial value $(p(0,0),q(0,0))^T$. Let

$$\begin{pmatrix}p(tm,tn)\\ q(tm,tn)\end{pmatrix}=g_{F_m}^{tm}g_{F_n}^{tn}\begin{pmatrix}p(0,0)\\ q(0,0)\end{pmatrix}, \tag{4.9}$$

$$\begin{pmatrix}P(tm,tn)\\ Q(tm,tn)\end{pmatrix}=g_{H_{m+1}}^{tm}g_{H_{n+1}}^{tn}\begin{pmatrix}p(0,0)\\ q(0,0)\end{pmatrix}. \tag{4.10}$$

Since the F_m flow and F_n flow $(m,n=0,1,2,\ldots)$ are commutable, (4.9) is called an involutive solution of the Hamiltonian canonical equations (F_m) and (F_n) and (4.10) is called an involutive solution of the Hamiltonian canonical equations (H_{m+1}) and (H_{n+1}).

By means of Theorem 1 and its corollary and (4.7)–(4.10), we have the following two theorems.

Theorem 5: Suppose $(p(x,tm),q(x,tm))^T$ is an involutive solution of the Hamiltonian canonical equation system (F_0) and (F_m). Then

$$\begin{pmatrix}w\\ u\end{pmatrix}=-4\begin{pmatrix}2(\langle Aq(x,tm),q(x,tm)\rangle+\langle q(x,tm),q(x,tm)\rangle^2)\\ \langle q(x,tm),q(x,tm)\rangle\end{pmatrix} \tag{4.5'}$$

becomes the solution of the mth-order $(m=1,2,\ldots)$ Boussinesq–Burgers' equation (2.6) $(v_{tm}=JG_m)$. Especially if $(p(x,t1),q(x,t1))^T$ is an involutive solution of the Hamiltonian canonical equation system (F_0) and (F_1), then

$$\begin{pmatrix}w\\ u\end{pmatrix}=-4\begin{pmatrix}2(\langle Aq(x,t1),q(x,t1)\rangle+\langle q(x,t1),q(x,t1)\rangle^2)\\ \langle q(x,t1),q(x,t1)\rangle\end{pmatrix}$$

satisfies the Boussinesq–Burgers' equation

$$\begin{pmatrix}u\\ w\end{pmatrix}_{t1}=2\begin{pmatrix}w_x+uu_x\\ u_{xxx}+(uw)_x\end{pmatrix}. \tag{2.10}$$

Theorem 6: Suppose $(P(x,tm),Q(x,tm))^T$ is an involutive solution of the constrained Hamiltonian canonical equation system (H_1) and (H_{m+1}). Then

$$w=4(\langle P(x,tm),P(x,tm)\rangle+2\langle A^2Q(x,tm),Q(x,tm)\rangle+2\langle AQ(x,tm),Q(x,tm)\rangle-2\langle AQ(x,tm),Q(x,tm)\rangle^2, \tag{4.11}$$

$$u=4\langle AQ(x,tm),Q(x,tm)\rangle$$

becomes the solution of the mth order $(m=1,2,\ldots)$ Boussinesq–Burgers' equation (2.6) $(v_{tm}=JG_m)$. Especially if $(P(x,t1),Q(x,t1))^T$ is an involutive solution of the constrained Hamiltonian canonical equation system (H_1) and (H_2), then

$$w=4(\langle P(x,t1),P(x,t1)\rangle+2\langle A^2Q(x,t1),Q(x,t1)\rangle+2\langle AQ(x,t1),P(x,t1)\rangle-2\langle AQ(x,t1),Q(x,t1)\rangle^2),$$

$$u=4\langle AQ(x,t1),Q(x,t1)\rangle$$

satisfies the Boussinesq–Burgers' equation (2.10)

The significance of the above two theorems is that solving a high nonlinear partial differential equation has been transformed into solving two ordinary differential equations and solving the soliton equations by making use of finite-dimensional completely integrable systems.

ACKNOWLEDGMENT

It is a pleasure to thank the referees for a number of helpful comments on the first draft of this paper.

[1]H. Flaschka, *Non-linear Integrable System Classical Theory and Quantum Theory* (Kyoto, 1981), pp. 221–240.
[2]V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1987).
[3]Cewen Cao, "A classical completely integrable system and the involutive representation for the solutions of the KdV equation" (preprint).
[4]B. A. Kupershmidt, Commun. Math. Phys. **99**, 51 (1985).
[5]R. L. Sachs, Physica D **30**, 1 (1988).
[6]R. Hirota, J. Phys. Soc. Jpn. **55**, 2173 (1986).
[7]W. Strammp, J. Phys. Soc. Jpn. **53**, 4129 (1984).
[8]A. S. Fokas and P. M. Santini, J. Math. Phys. **29**, 604 (1988).
[9]P. D. Lax, SIAM Rev. **18**, 351 (1976).
[10]P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).
[11]J. Moser, "Integrable Hamiltonian system and spectral theory," in *Proceedings of the 1983 Beijing Symposium on Differential Geometry and Differential Equations* (Science, Beijing, 1986), pp. 157–229.
[12]H. Airault, H. P. McKean, and J. Moser, Commun. Pure Appl. Math. **30**, 95 (1977).
[13]R. C. Churchill, "On proving the nonintegrability of a Hamiltonian system," in *Lecture Notes in Mathematics*, Vol. 925 (Springer, Berlin, 1982), pp. 103–122.
[14]Cewen Cao, Henan Sci. (Chinese) **5**, 1 (1987).

Auxiliary linear equations for a class of nonlinear partial differential equations via jet-bundle formulation

You-Quan Lee
Center of Theoretical Physics CCAST (World Laboratory), Beijing 100080, People's Republic of China and Department of Physics, Yantai University, Yantai 264005, People's Republic of China

(Received 6 September 1989; accepted for publication 17 January 1990)

It is shown that a class of two-variable nonlinear partial differential equations, such as the Liouville equation, the sine–Gordon equation, the Ernst equation, and the Ernst–Maxwell equations, can be Ricatti-type quasilinear systems through maps from $J^1(R^2, R^n)$ to $J^0(R^2, R^l)$. The auxiliary linear equations (Laxpair) for them are formulated, respectively, by using the W–E prolongation procedure. The Lax pairs for the Liouville equation and the sine–Gordon equation contain an arbitrary parameter besides the usual spectral parameter.

I. INTRODUCTION

The discovery of the inverse scattering method in 1967[1] is undoubtedly one of the most elegant contributions to mathematical physics in the 20th century. It led to the development of a powerful method for integrating important nonlinear partial differential equations. Auxiliary linear systems or the so-called Lax pair[2] of a given nonlinear partial differential equation is indispensable in this method. The W–E prolongation procedure[3] provides a useful program for deriving an auxiliary linear system from it.[4,5]

In the present paper, using the W–E prolongation procedure, we formulate the auxiliary linear system of the Liouville equation and sine–Gordon equation. It is shown that these two equations can be turned into a Ricatti-type, quasilinear system through certain maps from $J^1(R^2, R)$ to $J^0(R^2, R^l)$. As these maps permit us to introduce an arbitrary parameter, the auxiliary linear system naturally contains two parameters: One is the spectral parameter but the other is not. Finally, we briefly recall the results in the author's previous paper about the Ernst equation and the Ernst–Maxwell equations showing that they also become the Ricatti-type, quasilinear system through maps from $J^1(R^2,C^2)$ to $J^0(R^2,C^6)$ and from $J^1(R^2,C^3)$ to $J^0(R^2,C^{10})$, respectively.

II. JET-BUNDLE FORMULATION

First of all, we summarize the relevant definitions and notations from the theory of jet bundles.[6]

Let $C^\infty(M,N)$ denote the set of C^∞ maps from M to N and let $J^k(M,N)$ denote the k-jet bundle of these maps. The k-jet bundle $J^k(M,N)$ is given by the equivalent classes of maps in $C^\infty(M,N)$ having k th order contact. Especially, the 0-jet bundle $J^0(M,N)$ is identified with $M\times N$. For $k>l$ there is a natural projection Π^k_l from $J^k(M,N)$ to $J^l(M,N)$ given by $j^k_x f\mapsto j^l_x f$, where $j^k_x f$ denotes the k jet of $f\in C^\infty(M,N)$ at $x\in M$. The source projection is the map α: $J^k(M,N)\to M$ given by $j^k_x\mapsto x\ \forall f\in C^\infty(M,N)$. The cross section of α are defined as smooth maps s: $M\to J^k(M,N)$ such that $\alpha\cdot s = id_M$. An important example of such a cross section is the k-jet extension of a map $f\in C^\infty(M,N)$ defined by j^kf: $x\mapsto j^k_x f$.

If $M=R^m$, $N=R^n$, let x^a, $a=1,\dots,m$ and Z^A, $A=1,\dots,n$ be coordinates on R^m and R^n, respectively. The standard coordinates on $J^k(R^m,R^n)$ are then $x^a, Z^A, Z^A_a,\dots, Z^A_{a_1a_2\dots a_k}$.

In these coordinates, the solution manifold of a general system of k th-order partial differential equations that reads

$$F_\lambda(x^a,u^A,\dots,\ \partial_{a_1\dots a_k}u^A)=0, \tag{1}$$

can be associated with the submanifold S^k of $J^k(R^m,R^n)$. This submanifold is given by the following constraint equations:

$$F_\lambda(x^a,Z^A,Z^A_a,\dots,\ Z^A_{a_1\dots a_k})=0. \tag{2}$$

Obviously, a solution of Eq. (1) is a map $f\in C^\infty(R^m,R^n)$ such that the pull back $j^kf^*F\ (x^a,Z^A,\dots,Z^A_{a_1\dots a_k})$ vanishes. Then, S^k will be referred as k th-order differential equations.

For the aim of this paper, we study the case that S_k is a quasilinear system of equations on $J_k(R^m,R^n)$. The constraint equations of a quasilinear system on $J^k(R^m,R^n)$ are generally given by

$$F{:}=F_\lambda{}^{a_1\dots a_k}{}_A Z^A_{a_1\dots a_k}+G_\lambda=0, \tag{3}$$

where $F_\lambda{}^{a_1\dots a_k}{}_A$ and G_λ are functions independent of $Z^A_{a_1\dots a_k}$. A map $f\in C^\infty(R^m,R^n)$ is a solution of S^k iff $j_kf^*F=0$, then it can be shown in almost the same technique as in Ref. 6 that f is a solution of S_k iff

$$j^kf^*\sigma_\lambda=0, \tag{4}$$

where σ^λ is a system of m-forms on $J^{k-1}(R^m,R^n)$ associated with Eq. (1). In the system

$$\sigma^\lambda{:}=F_{\lambda A}{}^{a_1\dots a_{k-1}b}\,dZ^A_{a_1\dots a_{k-1}}\wedge\omega_b+G_\lambda\omega, \tag{5}$$

the notation ω in Eq. (5) stands for the volume form on R^m and $\omega_b=\partial_b\,\lrcorner\,\omega$.

Furthermore, let $\theta^A, \theta^A_a,\dots,\theta^A_{a_1\dots a_{k-2}}$ denote the contact forms on $J^{k-1}(R^m,R^n)$ defined by

$$\begin{aligned}\theta^A &{:}= dZ^A - Z^A_a\,dx^a,\\ &\ \vdots\\ \theta^A_{a_1\dots a_{k-2}} &{:}= dZ^A_{a_1\dots a_{k-2}} - Z^A_{a_1\dots a_{k-2}a}\,dx^a,\end{aligned} \tag{6}$$

and let Σ^{k-1} denote the exterior differential system genera-

 0022-2488/90/061381-07$03.00

ted by $\{\theta^A \wedge \omega_a, \ldots \theta^A_{a_1 \ldots a_{k-2}} \wedge \omega_a\}$ and σ_λ. It is standard to show that the solutions of Σ^{k-1} have a one-to-one correspondence with the solutions of S^k. We may therefore consider Σ^{k-1} and S^k as completely equivalent.

An important simple case is such a class of second-order partial differential equations that can be reduced to the Ricatti-type form. For S^2 given by the following constraint equations on $J^2(R^2,R^n)$:

$$F_\lambda(x^a, Z^A, Z^A_a, Z^A_{ab}) = 0, \quad \lambda = 1,\ldots,r, \quad a,b = 1,2. \tag{7}$$

We consider the case that a smooth map denoted by Q exists

$$Q: \; J^1(R^2,R^n) \to J^0(R^2,R^l), \tag{8}$$

such that Eq. (7) becomes the Ricatti-type form under this mapping, i.e.,

$$F_\mu := F^a_{\mu\alpha} Z^\alpha_a + G_{\mu\alpha\beta} Z^\alpha Z^\beta = 0,$$
$$\mu = 1,\ldots,s, \quad \alpha,\beta = 1,\ldots,1, \tag{9}$$

where $F^a_{\mu\alpha}$ and $G_{\mu\alpha\beta}$ are constants. It will be seen in the next section that the Liouville equation, the sine–Gordon equation, the Ernst equation, and the Ernst–Maxwell equations belong to this case.

According to the W–E prolongation procedure, the contact one-form on $J^1(R^2,N')$ is introduced:

$$\theta'^i := dq'^i_a - q'^i_a \, dx^a. \tag{10}$$

Here, the dimension of N' is unlimited. Consider the map ϕ: $J^0(R^2,R^l) \times J^0(R^2,N') \to J^1(R^2,N')$ such that the following figures commute:

$$\begin{array}{ccccc} J^0(R^2,R^l) & \underset{R^2}{\times} & J^0(R^2,N') & \to & J^1(R^2,N') \\ \alpha \cdot \mathrm{pr}_1 & \downarrow & & \swarrow & \alpha \\ & R^2 & & & \end{array}$$

$$\begin{array}{ccccc} J^0(R^2,R^l) & \underset{R^2}{\times} & J^0(R^2,N') & \to & J^1(R^2,N') \\ \mathrm{pr}_2 & \downarrow & & \swarrow & \beta \\ & N' & & & \end{array}$$

where pr_i is the projection of the ith factor and β is the target projection map defined by β: $J^k(M,N) \to N$. This requires that the map is given in coordinates by

$$x'^a = x^a, \quad q'^i = q^i,$$
$$q'^i_a = \phi^i_a(x^b, Z^\alpha, q^i). \tag{11}$$

The closure condition requires

$$d\phi^* \phi'^i \in \mathscr{I}(\Sigma^0, \phi^*\theta'^i). \tag{12}$$

For the Riccatti-type, quasilinear system, we only need to consider the linear part of Z^α in ϕ^i_a, because the higher-power terms contribute vanishing commutators in prolongation algebra. With ϕ^i_a admitted by condition Eq. (12) we can determine the pseudopotentials q^i from $\phi^*\theta'^i = 0$ on the solution manifold of Σ^0.

III. APPLICATIONS

In this section we discuss the Liouville equation, the sine–Gordon equation, the Ernst equation, and the Ernst–Maxwell equations, respectively.

A. Liouville equation

The standard Liouville equation reads

$$\partial_1 \partial_2 u - e^{-2u} = 0. \tag{13}$$

In terms of jet-bundle formulation, the constraint equation on $J^2(R^2,R)$ takes the form

$$Z_{12} - e^{-2u} = 0. \tag{14}$$

Through a map from $J^1(R^2,R)$ to $J^0(R^2,R^4)$ defined by

$$Z^1 = Z_1, \quad Z^2 = Z_2,$$
$$Z^3 = e^{-(1+\rho)z}, \quad Z^4 = e^{-(1-\rho)z}, \tag{15}$$

the constraint equation becomes

$$Z^2_1 - Z^3 Z^4 = 0,$$
$$Z^1_2 - Z^3 Z^4 = 0,$$
$$Z^3_1 - (1+\rho) Z^3 Z^1 = 0,$$
$$Z^3_2 - (1+\rho) Z^3 Z^2 = 0,$$
$$Z^4_1 + (1-\rho) Z^4 Z^1 = 0,$$
$$Z^4_2 + (1-\rho) Z^4 Z^2 = 0, \tag{16}$$

where ρ is a permitted arbitrary constant. The two-forms on $J^0(R^2,R^4)$ associated with Eq. (16) then read

$$\sigma_1 := dZ^1 \wedge dx^1 + Z^3 Z^4 \, dx^1 \wedge dx^2,$$
$$\sigma_2 := dZ^2 \wedge dx^2 - Z^3 Z^4 \, dx^1 \wedge dx^2,$$
$$\sigma_3 := dZ^3 \wedge dx^2 + (1+\rho) Z^3 Z^1 \, dx^1 \wedge dx^2,$$
$$\sigma_4 := dZ^3 \wedge dx^1 - (1+\rho) Z^3 Z^2 \, dx^1 \wedge dx^2,$$
$$\sigma_5 := dz^4 \wedge dx^2 + (1-\rho) Z^4 Z^1 \, dx^1 \wedge dx^2,$$
$$\sigma_6 := dZ^4 \wedge dx^1 - (1-\rho) Z^4 Z^2 \, dx^1 \wedge dx^2. \tag{17}$$

The pull back of contact one-forms on $J^1(R^2,N')$ is written as

$$\phi^*\theta'^i = dq^i - \phi^i_a(Z^1,\ldots,Z^4,q^i) dx^a. \tag{18}$$

For simplicity, here $\partial\phi^i_a/\partial x^b = 0$ is assumed, which implies that the map with translational invariance is considered.

When Eq. (12) is written out in detail by using Eqs. (17) and (18), the following set of partial differential equations for ϕ^i_a is obtained:

$$\frac{\partial\phi^i_1}{\partial Z^2} = \frac{\partial\phi^i_2}{\partial Z^1} = 0, \tag{19}$$

$$\phi^j_2 \frac{\partial\phi^i_1}{\partial q^j} - \phi^j_1 \frac{\partial\phi^i_2}{\partial q^j} + \frac{\partial\phi^i_1}{\partial Z^1} Z^3 Z^4 - (1+\rho) \frac{\partial\phi^i_1}{\partial Z^3} Z^3 Z^2$$
$$- (1-\rho) \frac{\partial\phi^i_1}{\partial Z^4} Z^4 Z^2$$
$$- \frac{\partial\phi^i_2}{\partial Z^2} Z^3 Z^4 + (1+\rho) \frac{\partial\phi^i_2}{\partial Z^3} Z^3 Z^1$$
$$+ (1-\rho) \frac{\partial\phi^i_1}{\partial Z^4} Z^4 Z^1 = 0. \tag{20}$$

Due to Eq. (19) ϕ^i_a can be expressed as

$$\phi^i_1 = X^i_1(q) Z^1 + X^i_2(q) Z^3 + X^i_3(q) Z^4,$$

$$\phi^i_2 = X^i_6(q) Z^2 + X^i_5(q) Z^3 + X^i_4(q) Z^4. \tag{21}$$

Substituting Eq. (21) into Eq. (20), we have eight partial differential equations for X^i_1 ($l = 1,...,6$). If the notations

$$X_l := X^i_l \frac{\partial}{\partial q^i},$$

$$[X_l, X_m] = X^j_l \frac{\partial}{\partial q^j} X^i_m - X^j_m \frac{\partial}{\partial q^j} X^i_l, \tag{22}$$

are adopted, the eight equations are written as the following incompleted algebra:

$$\begin{aligned}
&[X_6, X_1] = 0, \quad [X_5, X_2] = 0, \quad [X_4, X_3] = 0,\\
&[X_6, X_2] = (1+\rho)X_2, \quad [X_6, X_3) = (1-\rho)X_3,\\
&[X_5, X_1] = -(1+\rho)X_5, \quad [X_4, X_1] = -(1-\rho)X_4,\\
&[X_5, X_3] + [X_4, X_2] = -X_1 + X_6.
\end{aligned} \tag{23}$$

This is the prolongation structure of the nonlinear equation under consideration.

Considering the compatibility of these commutator relations and the requirements of Jacobbi identities, we can choose

$$\begin{aligned}
&X_3 = 0, \quad X_5 = 0,\\
&X_6 = -[(1+\rho)/(1-\rho)]X_1.
\end{aligned} \tag{24}$$

The Eq. (23) is reduced to the following:

$$\begin{aligned}
&[X_1, X_2] = -(1-\rho)X_2,\\
&[X_1, X_4] = (1-\rho)X_4,\\
&[X_2, X_4] = -[2/(1-\rho)]X_2.
\end{aligned} \tag{25}$$

An infinite-dimensional linear realization of Eq. (25) is found;

$$\begin{aligned}
&X_1 = [(1-\rho)/2](A^{(0)}_{11} - A^{(0)}_{22}),\\
&X_2 = A^{(1)}_{21},\\
&X_4 = A^{(-1)}_{12},
\end{aligned} \tag{26}$$

where

$$A^{(m)}_{ij} = \sum_{n=-\infty}^{\infty} q^{(m+n)}_i \frac{\partial}{\partial q^{(n)}_j},$$

which satisfies

$$[A^{(m)}_{ij}, A^{(n)}_{kl}] = \delta_{jk} A^{(m+n)}_{il} - \delta_{li} A^{(m+n)}_{kj}.$$

This is a Kac–Moody algebra without the center term.

Using Eqs. (21) and (26), and writing out $\phi^*\theta'^{(n)}_a = 0$ ($a = 1,2$) on the solution manifold, we have

$$\begin{aligned}
&dq^{(n)}_1 - ((1-\rho)/2q^{(n)}_1 Z^1 + q^{(n+1)}_2 Z^3)dx^1\\
&\quad + [(1+\rho)/2]q^{(n)}_1 Z^2\,dx^2 = 0,\\
&dq^{(n)}_2 + [(1-\rho)/2]q^{(n)}_2 Z^1\,dx^1 - ([(1+\rho)/2]q^{(n)}_2 Z^2\\
&\quad + q^{(n-1)}_1 Z^4)dx^2 = 0.
\end{aligned} \tag{27}$$

If definitions

$$\psi^{(\lambda)}_a = \sum_{n=-\infty}^{\infty} \lambda^n q^{(n)}_a, \quad a = 1,2, \tag{28}$$

are introduced, Eq. (27) gives an auxiliary linear system of equations for the Liouville equation. In terms of a matrix they are written as follows:

$$d\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} - \begin{pmatrix} \frac{1-\rho}{2} Z^1\,dx^1 - \frac{1+\rho}{2} Z^2\,dx^2, & \frac{1}{\lambda} Z^3\,dx^1\\ \lambda Z^4\,dx^2, & -\frac{1-\rho}{2} Z^1\,dx^1 + \frac{1+\rho}{2} Z^2\,dx^2 \end{pmatrix} \times \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} = 0, \tag{29}$$

where λ is the spectral parameter and ρ is a new parameter involved in our formulation.

B. Sine–Gordon equation

The well-known sine–Gordon equation reads

$$\partial_1 \partial_2 u = 2\sin 2u. \tag{30}$$

Consider a map $Q: J^1(R^2,R) \to J^0(R^2,R^6)$ defined by

$$\begin{aligned}
&Z^1 = Z_1, \quad Z^2 = Z_2,\\
&Z^3 = \sin(1+\rho)Z, \quad Z^4 = \sin(1-\rho)Z,\\
&Z^5 = \cos(1+\rho)Z, \quad Z^6 = \cos(1-\rho)Z.
\end{aligned}$$

The constraint equation on $J^2(R^2,R)$ becomes the following Ricatti-type quasilinear forms on $J^1(R^2,R^6)$ under this map:

$$\begin{aligned}
&Z^1_2 = Z^2_1 = 2(Z^3Z^6 + Z^4Z^5),\\
&Z^3_a = (1+\rho)Z^5 Z^a,\\
&Z^4_a = (1-\rho)Z^6 Z^a,\\
&Z^5_a = -(1+\rho)Z^3 Z^a,\\
&Z^6_a = -(1-\rho)Z^4 Z^a,
\end{aligned} \tag{31}$$

where $a = 1,2$. The associated two-forms on $J^0(R^2,R^6)$ are

$$\begin{aligned}
&\sigma_1 := dZ^2 \wedge dx^2 - 2(Z^3Z^6 + Z^4Z^5)dx^1 \wedge dx^2,\\
&\sigma_2 := dZ^1 \wedge dx^1 + 2(Z^3Z^6 + Z^4Z^5)dx^1 \wedge dx^2,\\
&\sigma_3 := dZ^3 \wedge dx^2 - (1+\rho)Z^5Z^1\,dx^1 \wedge dx^2,\\
&\sigma_4 := dZ^3 \wedge dx^1 + (1+\rho)Z^5Z^2\,dx^1 \wedge dx^2,\\
&\sigma_5 := dZ^4 \wedge dx^2 - (1-\rho)Z^6Z^1\,dx^1 \wedge dx^2,\\
&\sigma_6 := dZ^4 \wedge dx^1 + (1-\rho)Z^6Z^2\,dx^1 \wedge dx^2,\\
&\sigma_7 := dZ^5 \wedge dx^2 + (1+\rho)Z^3Z^1\,dx^1 \wedge dx^2,\\
&\sigma_8 := dZ^5 \wedge dx^1 - (1+\rho)Z^3Z^2\,dx^1 \wedge dx^2,\\
&\sigma_9 := dZ^6 \wedge dx^2 + (1-\rho)Z^4Z^1\,dx^1 \wedge dx^2,\\
&\sigma_{10} := dZ^6 \wedge dx^1 - (1-\rho)Z^4Z^2\,dx^1 \wedge dx^2.
\end{aligned} \tag{32}$$

In the same manner as in previous discussion on the Liouville equation, we introduce the pull back of contact one-forms. Then we obtain from the closure condition Eq. (12) that

$$\begin{aligned}
&\phi^i_1 = X^i_1 Z^1 + X^i_3 Z^3 + X^i_4 Z^4 + X^i_5 Z^5 + X^i_6 Z^6,\\
&\phi^i_2 = X^i_2 Z^2 + X^i_7 Z^3 + X^i_8 Z^4 + X^i_9 Z^5 + X^i_{10} Z^6,
\end{aligned} \tag{33}$$

and $X_1 = X^i_1 (\partial/\partial q^i)$ ($l = 1,2,...,10$) satisfies the following commutator relations:

$$[X_1,X_2]=0,\quad [X_3,X_7]=0,\quad [X_4,X_8]=0,$$
$$[X_5,X_9]=0,\quad [X_6,X_{10}]=0,$$
$$[X_3,X_8]+[X_4,X_7]=0,\quad [X_3,X_9]+[X_5,X_7]=0,$$
$$[X_4,X_{10}]+[X_6,X_8]=0,\quad [X_5,X_{10}]+[X_6,X_9]=0,$$
$$[X_1,X_7]=(1+\rho)X_9,\quad [X_1,X_8]=(1-\rho)X_{10},$$
$$[X_1,X_9]=-(1+\rho)X_7,\quad [X_1,X_{10}]=-(1-\rho)X_8,$$
$$[X_3,X_2]=-(1+\rho)X_5,\quad [X_4,X_2]=-(1-\rho)X_6,$$
$$[X_5,X_2]=(1+\rho)X_3,\quad [X_6,X_2]=(1+\rho)X_4,$$
$$[X_3,X_{10}]+[X_6,X_7]=2(X_1-X_2),$$
$$[X_4,X_9]+[X_5,X_8]=2(X_1-X_2). \tag{34}$$

The compatibility of Eq. (34) and requirements of Jacobi identities permit the simple choice $X_3=X_5=X_8=X_{10}=0$, or $X_4=X_6=X_7=X_9=0$. However the results given by the two choices have no difference but a sign of the parameter ρ. Thus we only need to consider the first case; under this choice Eq. (34) reduces to

$$[X_1,X_7]=(1+\rho)X_9,\quad [X_9,X_1]=(1+\rho)X_7,$$
$$[X_2,X_4]=(1-\rho)X_6,\quad [X_6,X_2]=(1-\rho)X_4,$$
$$[X_6,X_7]=2(X_1-X_2),\quad [X_4,X_9]=2(X_1-X_2),$$
$$[X_4,X_7]=0,\quad [X_6,X_9]=0,\quad [X_1,X_2]=0. \tag{35}$$

An infinite-dimensional linear realization of Eq. (35) is found to be

$$X_1=(1+\rho)T_3^{(0)},\quad X_7=2T_1^{(1)},\quad X_9=2T_2^{(1)},$$
$$X_2=-(1-\rho)T_3^{(0)},\quad X_4=2T_1^{(-1)}, \tag{36}$$
$$X_6=-2T_2^{(-1)},$$

where

$$T_1^{(m)}=\sum_{n=-\infty}^{+\infty}\frac{i}{2}\left(q_1^{(m+n)}\frac{\partial}{\partial q_2^{(n)}}+q_2^{(m+n)}\frac{\partial}{\partial q_1^{(n)}}\right),$$
$$T_2^{(m)}=\sum_{n=-\infty}^{+\infty}\frac{1}{2}\left(q_2^{(m+n)}\frac{\partial}{\partial q_1^{(n)}}-q_1^{(m+n)}\frac{\partial}{\partial q_2^{(n)}}\right), \tag{37}$$
$$T_3^{(m)}=\sum_{n=-\infty}^{+\infty}\frac{i}{2}\left(q_1^{(m+n)}\frac{\partial}{\partial q_1^{(n)}}-q_2^{(m+n)}\frac{\partial}{\partial q_2^{(n)}}\right),$$

satisfying the su(2) $\otimes$ C Kac–Moody commutator relations (without the center term)

$$[T_i^{(m)},T_j^{(n)}]=\epsilon_{ijk}T_k^{(m+n)}. \tag{38}$$

Therefore, on the solution manifold, $q_a^{(n)}$ satisfies the following equations:

$$dq_1^{(n)}=\{[i(1+\rho)/2]q_1^{(n)}Z^1+iq_2^{(n-1)}Z^4-q_2^{(n-1)}Z^6\}dx^1+\{-[i(1-\rho)/2]q_1^{(n)}Z^2+iq_2^{(n+1)}Z^3+q_2^{(n+1)}Z^5\}dx^2,$$
$$dq_2^{(n)}=\{-(i/2)(1+\rho)q_2^{(n)}Z^1+iq_1^{(n-1)}Z^4+q_1^{(n-1)}Z^6\}dx^1+\{(i/2)(1-\rho)q_2^{(n)}Z^2+iq_1^{(n+1)}Z^3-q_1^{(n+1)}Z^5\}dx^2. \tag{39}$$

In order to introduce the spectral parameter, definition (28) is adopted. Then the following auxiliary linear system (Lax pair) for the Sine-Gordon equation is obtained from Eq. (39):

$$d\Psi=M\Psi, \tag{40}$$

where

$$M:=\begin{pmatrix}\frac{i(1+\rho)}{2}Z^1dx^1-\frac{i(1-\rho)}{2}Z^2dx^2, & -\lambda(Z^6-iZ^4)dx^1+\frac{1}{\lambda}(Z^5+iZ^3)dx^2\\ \lambda(Z^6+iZ^4)dx^1-\frac{1}{\lambda}(Z^5-iZ^3)dx^2, & -\frac{i(1+\rho)}{2}Z^1dx^1+\frac{i(1-\rho)}{2}Z^2dx^2\end{pmatrix},$$
$$\psi:=\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \tag{41}$$

The above Lax pair contains the new parameter ρ, which is not the spectral parameter.

C. The Ernst equation and the Ernst–Maxwell equations

The Ernst equations of stationary axisymmetric exterior gravity field are

$$\partial_1\partial_2\mathscr{E}=-\frac{1}{2W}(\partial_1W\partial_2\mathscr{E}+\partial_2W\partial_1\mathscr{E})+\frac{1}{T}\partial_1\mathscr{E}\,\partial_2\mathscr{E},$$
$$\partial_1\partial_2W=0, \tag{42}$$

where $T=(\mathscr{E}+\mathscr{E}^*)/2$. Obviously, the related constraint equations on $J^2(R^2,C^2)$ become the following Ricatti-type quasilinear forms on $J^1(R^2,C^6)$:

$$Z_1^2=(Z^1-Z^3)Z^2-\tfrac{1}{2}(Z^5Z^2+Z^6Z^1),$$
$$Z_2^1=(Z^2-Z^4)Z^1-\tfrac{1}{2}(Z^6Z^1+Z^5Z^2),$$
$$Z_1^4=(Z^3-Z^1)Z^4-\tfrac{1}{2}(Z^5Z^4+Z^6Z^3),$$
$$Z_2^3=(Z^4-Z^2)Z^3-\tfrac{1}{2}(Z^6Z^3+Z^5Z^4),$$
$$Z_1^6=-Z^5Z^6,$$
$$Z_2^5=-Z^5Z^6, \tag{43}$$

under the maps from $J^1(R^2,C^2)$ to $J^0(R^2,C^6)$:

$$Z^1=\mathscr{E}_1/2T,\quad Z^2=\mathscr{E}_2/2T,\quad Z^3=\mathscr{E}_1^*/2T,$$
$$Z^4=\mathscr{E}_2^*/2T, \tag{44}$$
$$Z^5=W_1/W,\quad Z^6=W_2/W.$$

The related two-forms on $J^0(R^2,C^6)$ can be written down on the basis of Eq. (5). Then the prolongation structure of the Ernst equations are obtained by introducing the pull back of contact one-forms:

$$\begin{aligned}
&[X_1,Y_1]=X_1-Y_2,\quad [X_1,Y_4]=-X_1+Y_4,\\
&[X_1,Y_6]=\tfrac{1}{2}(-X_1+Y_2),\\
&[X_3,Y_2]=-X_3+Y_2,\quad [X_3,Y_4]=X_3-Y_4,\\
&[X_3,Y_6]=\tfrac{1}{2}(-X_3+Y_4),\\
&[X_5,Y_2]=\tfrac{1}{2}(-X_1+Y_2),\\
&[X_5,Y_4]=\tfrac{1}{2}(-X_3+Y_4),\\
&[X_5,Y_6]=-X_5+Y_6.
\end{aligned} \tag{45}$$

The pull back of contact one-forms are determined by

$$\begin{aligned}
\phi_1^i&=X_1^iZ^1+X_3^iZ^3+X_5^iZ^5\\
\phi_2^i&=X_2^iZ^2+X_4^iZ^4+X_6^iZ^6.
\end{aligned} \tag{46}$$

An infinite-dimensional linear realization of Eq. (45) is found,

$$\begin{aligned}
&X_1=A_{11}^{(0)}+A_{21}^{(-1)},\quad X_2=A_{11}^{(0)}+A_{21}^{(1)},\\
&X_3=A_{22}^{(0)}+A_{12}^{(-1)},\quad X_4=A_{22}^{(0)}+A_{12}^{(1)},\\
&X_5=\tfrac{1}{2}(D^{(0)}-D^{(-2)}),\quad X_6=\tfrac{1}{2}(D^{(2)}-D^{(0)}),
\end{aligned} \tag{47}$$

where the definition of A_{ij}^{m} have been given previously; $D^{(m)}$ is defined by

$$D^{(m)}=\sum_{n=-\infty}^{\infty}(m+n)\left(\sum_{i=1}^{r}q_i^{(m+n)}\frac{\partial}{\partial q_i^{(n)}}\right), \tag{48}$$

where they satisfy the following coupled Kac–Moody and Virasoro (without center terms) commutator relations:

$$\begin{aligned}
&[A_{ij}^{(m)},A_{kl}^{(n)}]=\delta_{jk}A_{il}^{(m+n)}\\
&\qquad\qquad -\delta_{li}A_{kj}^{(m+n)}\quad\text{(Kac–Moody)},\\
&[D^{(m)},A_{ij}^{(n)}]=nA_{ij}^{(m+n)},\\
&[D^{(m)},D^{(n)}]=(n-m)D^{(m+n)}\quad\text{(Virasoro)}.
\end{aligned} \tag{49}$$

The pull back of the contact one-forms vanish on the solution manifold of Eq. (43). It gives the following partial differential equations for $q_i^{(n)}$ $(i=1,2)$:

$$\begin{aligned}
dq_1^{(n)}&=\{(q_1^{(n)}+q_2^{(n-1)})Z^1+\tfrac{1}{2}(nq_1^{(n)}\\
&\quad-(n-2)q_1^{(n-2)})Z^5\}dx^1\\
&\quad+\{(q_1^{(n)}+q_2^{(n+1)})Z^2\\
&\quad+\tfrac{1}{2}[(n+2)q_1^{(n+2)}-nq_1^{(n)}]Z^6\}dx^2,\\
dq_2^{(n)}&=\{(q_2^{(n)}+q_1^{(n-1)})Z^3\\
&\quad+\tfrac{1}{2}(nq_2^{(n)}-(n-2)q_2^{(n-2)})Z^5\}dx^1\\
&\quad+\{(q_2^{(n)}+q_1^{(n+1)})Z^4\\
&\quad+\tfrac{1}{2}[(n+2)q_2^{(n+2)}-nq_2^{(n)}]Z^6\}dx^2,
\end{aligned} \tag{50}$$

where Eqs. (46)–(48) have been used. If definition Eq. (28) (here λ must be regarded as function of x) is adopted, the following auxiliary linear system of the Ernst equation is obtained from Eq. (50):

$$d\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}=\begin{pmatrix}Z^1dx^1+Z^2dx^2, & \lambda Z^1dx^1+\lambda^{-1}Z^2dx^2\\ \lambda Z^3dx^1+\lambda^{-1}Z^4dx^2, & Z^3dx^1+Z^4dx^2\end{pmatrix}\times\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}, \tag{51}$$

where

$$d\lambda=\tfrac{1}{2}(\lambda^2-1)(\lambda Z^5dx^1+\lambda^{-1}Z^6dx^2).$$

The Ernst–Maxwell equations are

$$\begin{aligned}
\partial_1\partial_2\mathscr{E}&=-\frac{1}{2W}(\partial_1W\,\partial_2\mathscr{E}+\partial_2W\,\partial_1\mathscr{E})\\
&\quad+\frac{1}{Q}(\partial_1\mathscr{E}+2\phi\,\partial_1\phi)\partial_2\mathscr{E}\\
&\quad+(\partial_2\mathscr{E}+2\bar\phi\,\partial_2\phi)\partial_1\mathscr{E},\\
\partial_1\partial_2\phi&=-\frac{1}{2W}(\partial_1W\,\partial_2\phi+\partial_2W\,\partial_1\phi)\\
&\quad+\frac{1}{Q}(\partial_1\mathscr{E}+2\bar\phi\,\partial_1\phi)\partial_2\phi\\
&\quad+(\partial_2\mathscr{E}+2\bar\phi\,\partial_2\phi)\partial_1\phi,\\
\partial_1\partial_2W&=0,
\end{aligned} \tag{52}$$

where $Q=\mathscr{E}+\bar{\mathscr{E}}+2\bar\phi\phi$ and the "overbar" stands for complex conjugation. The related constraint equations on $J^2(R^2,C^3)$ are turned into Ricatti-type forms on $J^1(R^2,C^{10})$ through the following maps from $J^1(R^2,C^3)$ to $J^0(R^2,C^{10})$:

$$\begin{aligned}
&Z_1=(\mathscr{E}_1+2\bar\phi\phi_1)/Q,\quad Z^2=(\mathscr{E}_2+2\bar\phi\phi_2)/Q,\\
&Z^3=(\bar{\mathscr{E}}_1+2\phi\bar\phi_1)/Q,\quad Z^4=(\bar{\mathscr{E}}_2+2\phi\bar\phi_2)/Q,\\
&Z^5=i(1/Q)^{1/2}\bar\phi_1,\quad Z^6=i(1/Q)^{1/2}\bar\phi_2,\\
&Z^7=i(1/Q)^{1/2}\phi_1,\quad Z^8=i(1/Q)^{1/2}\phi_2,\\
&Z^9=W_1/W,\quad Z_{10}=W_2/W.
\end{aligned} \tag{53}$$

The associated two-forms on $J^0(R^2,C^{10})$ are

$$\begin{aligned}
\sigma_1&:=dZ^1\wedge dx^1+\{(Z^2-Z^4)Z^1\\
&\quad-\tfrac{1}{2}(Z^9Z^2+Z^{10}Z^1)-Z^6Z^7\}dx^1\wedge dx^2,\\
\sigma_2&:=dZ^2\wedge dx^2+\{(Z^1-Z^3)Z^2\\
&\quad-\tfrac{1}{2}(Z^9Z^2+Z^{10}Z^1)-Z^5Z^8\}dx^2\wedge dz^1,\\
\sigma_3&:=dZ^3\wedge dx^1+\{(Z^4-Z^2)Z^3\\
&\quad-\tfrac{1}{2}(Z^9Z^4+Z^{10}Z^3)-Z^8Z^5\}dx^1\wedge dx^2,\\
\sigma_4&:=dZ^4\wedge dx^2+\{(Z^3-Z^1)Z^4\\
&\quad-\tfrac{1}{2}(Z^9Z^4+Z^{10}Z^3)-Z^7Z^6\}dx^2\wedge dx^1,\\
\sigma_5&:=dZ^5\wedge dx^1+\{\tfrac{1}{2}(Z^4-Z^2)Z^5\\
&\quad-\tfrac{1}{2}(Z^9Z^6+Z^{10}Z^5)+Z^6Z^3\}dx^1\wedge dx^2,\\
\sigma_6&:=dZ^6\wedge dx^2+\{\tfrac{1}{2}(Z^3-Z^1)Z^6\\
&\quad-\tfrac{1}{2}(Z^9Z^6+Z^{10}Z^5)+Z^5Z^4\}dx^2\wedge dx^1,\\
\sigma_7&:=dZ^7\wedge dx^1+\{\tfrac{1}{2}(Z^2-Z^4)Z^7\\
&\quad-\tfrac{1}{2}(Z^9Z^8+Z^{10}Z^7)+Z^8Z^1\}dx^1\wedge dx^2,\\
\sigma_8&:=dZ^8\wedge dx^2+\{\tfrac{1}{2}(Z^1-Z^3)Z^8\\
&\quad-\tfrac{1}{2}(Z^9Z^8+Z^{10}Z^7)+Z^7Z^2\}dx^2\wedge dx^1,
\end{aligned}$$

$$\sigma_9 := dZ^9 \wedge dx^1 + Z^9 Z^{10} dx^2 \wedge dx^1,$$
$$\sigma_{10} := dZ^{10} \wedge dx^2 + Z^9 Z^{10} dx^1 \wedge dx^2. \tag{54}$$

In the same manner as in the previous discussion, we find that the contact one-forms on the solution manifold are determined from the closure condition as follows:

$$\phi_1^i = X_1^i Z^1 + X_3^i Z^3 + X_5^i Z^5 + X_7^i Z^7 + X_9^i Z^9,$$
$$\phi_2^i = X_2^i Z^2 + X_4^i Z^4 + X_6^i Z^6 + X_8^i Z^8 + X_{10}^i Z^{10}, \tag{55}$$

where X_l^i $(l = 1,2,...,10)$ satisfy the following commutator relations:

$$\begin{aligned}
&[X_1,X_2] = X_1 - X_2, \quad [X_1,X_6] = \tfrac{1}{2}X_6,\\
&[X_1,X_{10}] = \tfrac{1}{2}(-X_1 + X_2),\\
&[X_1,X_4] = -X_1 + X_4, \quad [X_1,X_8] = X_7 - \tfrac{1}{2}X_8,\\
&[X_3,X_2] = -X_3 + X_2, \quad [X_3,X_6] = X_5 - \tfrac{1}{2}X_6,\\
&[X_3,X_{10}] = \tfrac{1}{2}(-X_3 + X_4),\\
&[X_3,X_4] = X_3 - X_4, \quad [X_3,X_8] = \tfrac{1}{2}X_8,\\
&[X_5,X_2] = -\tfrac{1}{2}X_3, \quad [X_5,X_6] = 0,\\
&[X_5,X_{10}] = \tfrac{1}{2}(-X_5 + X_6),\\
&[X_5,X_4] = \tfrac{1}{2}X_5 - X_6, \quad [X_5,X_8] = -X_3 + X_2,\\
&[X_7,X_2] = \tfrac{1}{2}X_7 - X_8, \quad [X_7,X_6] = -X_1 + X_4,\\
&[X_7,X_{10}] = \tfrac{1}{2}(-X_7 + X_8),\\
&[X_7,X_4] = -\tfrac{1}{2}X_7, \quad [X_7,X_8] = 0,\\
&[X_9,X_2] = \tfrac{1}{2}(-X_1 + X_2),\\
&[X_9,X_6] = \tfrac{1}{2}(-X_5 + X_6),\\
&[X_9,X_{10}] = -X_9 + X_{10},\\
&[X_9,X_4] = \tfrac{1}{2}(-X_3 + X_4),\\
&[X_9,X_8] = \tfrac{1}{2}(-X_7 + X_8),
\end{aligned} \tag{56}$$

where the notations given by Eq. (22) are used. This is the prolongation structure of the Ernst–Maxwell equations. Its infinite-dimensional linear realization is found to be

$$\begin{aligned}
&X_1 = A_{11}^{(0)} + A_{21}^{(-1)} + \tfrac{1}{2}A_{33}^{(0)}, \quad X_2 = A_{11}^{(0)} + A_{21}^{(1)} + \tfrac{1}{2}A_{33}^{(0)},\\
&X_3 = A_{22}^{(0)} + A_{12}^{(-1)} + \tfrac{1}{2}A_{33}^{(0)}, \quad X_4 = A_{22}^{(0)} + A_{12}^{(1)} + \tfrac{1}{2}A_{33}^{(0)},\\
&X_5 = A_{31}^{(0)} - A_{32}^{(-1)}, \quad X_6 = A_{31}^{(0)} - A_{32}^{(1)},\\
&X_7 = -A_{13}^{(0)} - A_{23}^{(-1)}, \quad X_8 = -A_{13}^{(0)} - A_{23}^{(1)},\\
&X_9 = \tfrac{1}{2}(D^{(0)} - D^{(-2)}), \quad X_{10} = \tfrac{1}{2}(D^{(2)} - D^{(0)}).
\end{aligned} \tag{57}$$

The partial differential equations for $q_j^{(n)}$ $(n = -\infty,..., +\infty; j = 1,2,3)$ can be obtained from the above linear realization. From those equations we then obtain an auxiliary linear system for the Ernst–Maxwell equations:

$$d\Psi = M\Psi$$
$$M := \begin{pmatrix} Z^1 dx^1 + Z^2 dx^2, & \lambda Z^1 dx^1 + \lambda^{-1} Z^2 dx^2, & Z^5 dx^1 + Z^6 dx^2 \\ \lambda Z^3 dx^1 + \lambda^{-1} Z^4 dx^2, & Z^3 dx^1 + Z^4 dx^2, & -\lambda Z^5 dx^1 - \lambda^{-1} Z^6 dx^2 \\ -Z^7 dx^1 - Z^8 dx^2, & -\lambda Z^7 dx^1 - \lambda^{-1} Z^8 dx^2, & \tfrac{1}{2}(Z^1 + Z^3)dx^1 + \tfrac{1}{2}(Z^2 + Z^4)dx^2 \end{pmatrix}, \tag{58}$$

where

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix},$$

the spectral parameter λ is desired to obey

$$d\lambda = \tfrac{1}{2}(\lambda^2 - 1)(\lambda Z^9 dx^1 + \lambda^{-1} Z^{10} dx^2),$$

so as to reach Eq. (58).

IV. CONCLUSION

In the above we discussed the Liouville equation, the sine–Gordon equation, the Ernst equations, and the Ernst–Maxwell equations by means of the W–E prolongation procedure in jet-bundle formulation. These equations are second-order nonlinear partial differential equations, their associated jet bundle is $J^2(R^2,N)$, where $N = R^1$ for the Liouville equation and sine–Gordon equation, $N = C^2$ for Ernst equations, and $N = C^3$ for the Ernst–Maxwell equations. It was shown that the related constraint equations on $J^2(R^2,N)$ for those equations can be turned into Ricatti-type ones on $J^1(R^2,N')$ under certain maps from $J^1(R^2,N)$ to $J^0(R^2,N')$ where $N' = R^4, R^6, C^6$, and C^{10} for the Liouville equation, the sine–Gordon equation, the Ernst equations, and the Ernst–Maxwell equations, respectively. The maps for the cases of the Liouville equation and the sine–Gordon equation naturally contain an arbitrary parameter. From the Ricatti-type constraint equations on $J^1(R^2, N')$, we can immediately write down the associated two-forms on $J^0(R^2, N')$.

Introducing contact one-forms, we obtained a set of partial differential equations from their closure condition. The set of equations can be expressed as an incompleted algebra in terms of tangent vectors in prolongation space. The incompleted algebra is called the prolongation structure of the equation under consideration. For the four equations, the number of tangent vectors that are needed to determine the pull back of contact one-forms equal to the number of constraint equations on $J^1(R^2, N')$ or that of two-forms on $J^0(R^2, N')$. The prolongation structure of the Liouville equation or Sine–Gordon equation takes some set of Kac–Moody algebra (without the center term); for the Ernst equations or Ernst–Maxwell equations, its prolongation structure takes some set of the coupled Kac–Moody and Virasoro algebra (without the center terms). These can be seen from the infinite-dimensional linear realization.

Due to that, the maps for the Liouville equation and the sine–Gordon equation permit an arbitrary parameter, the auxiliary linear system obtained naturally contain two parameters, one is the spectral parameter but the other is not. As there are generators of Virasoro algebra appearing in prolongation structure of the Ernst equation or the Ernst–Maxwell equations, the spectral parameter in the Lax pair must be functions on R^2 and obey a constraint equation.

[1]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[2]P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968); **38**, 141 (1975); M. J. Ablowitz, D. J. Kaup, A. C. Newll, and H. Segur, Phys. Rev. Lett. **31**, 125 (1973); Stud. Appl. Math. **53**, 249 (1974); V. A. Belinsky and V. E. Zakharov, Sov. Phys. JETP **75**, 1953 (1978); **77**, 3 (1979).
[3]H. D. Wahlquist and E. B. Estabook, J. Math. Phys. **16**, 1 (1975); **17**, 1293 (1976).
[4]B. K. Harrison, Phys. Rev. Lett. **41**, 1197 (1978); M. Omote, J. Math. Phys. **27**, 2853 (1986).
[5]Y. Q. Lee and M. L. Ge, Commun. Theor. Phys. **10**, 79 (1988); Y. Q. Lee and J. F. Lu, Phys. Lett. A **135**, 117 (1989).
[6]C. Rogers and W. F. Shadwick, *Backlund Transformation and Their Applications* (Academic, New York, 1982).

The quantization condition in the presence of a magnetic field and quasiclassical eigenvalues of the Kepler problem with a centrifugal potential and Dirac's monopole field

Akira Yoshioka
Department of Mathematics, Faculty of Science and Technology, Science University of Tokyo, Noda, Chiba 278, Japan
Kiyotaka Ii
Department of Mathematics, Yamagata University, Yamagata 990, Japan

(Received 8 September 1989; accepted for publication 24 January 1990)

In the presence of a magnetic field, the Maslov quantization condition is not available in the original form. An alternative quantization condition is proposed with the aid of a principal $U(1)$ bundle over a phase space and a connection whose curvature form is the charged symplectic form. By means of this quantization condition, quasiclassical eigenvalues of the Kepler problem with a centrifugal potential and Dirac's monopole field are calculated, which turn out to coincide with the eigenvalues of the quantized problem.

I. INTRODUCTION

The Maslov quantization condition, which is considered a generalization of the Bohr quantization rule, is given as follows.[1] Let M be a C^∞ manifold and T^*M be its cotangent bundle. The symplectic manifold $(T^*M, d\theta_M)$ is considered, where θ_M is the canonical one-form on T^*M. Suppose $L\subset (T^*M, d\theta_M)$ is a Lagrangian submanifold, namely, $d\theta_M|L = 0$ and $\dim L = \dim M$. Then θ_M defines an element of $H^1(L;\mathbb{R})$, so that $\int_\gamma \theta_M$, the action integral for a closed curve γ in L, depends only on the homology class $[\gamma]\in H_1(L;\mathbb{Z})$. Now L satisfies the Maslov quantization condition if and only if

$$\frac{1}{2\pi}\int_\gamma \theta_M - \frac{1}{4}<\mu_L,[\gamma]> = \text{integer}, \tag{1.1}$$

for any closed curve γ in L, where $\mu_L \in H^1(L;\mathbb{Z})$ is the Maslov class of L (for the definition of the Maslov class, see Appendix A or Ref. 2). By means of (1.1), approximate eigenvalues of operators, which are referred to as quasiclassical eigenvalues, are calculated.[1] Moreover, it is shown that in some specific models, quasiclassical eigenvalues coincide, or coincide modulo a certain constant, with eigenvalues of operators.[3–6] However, note that (1.1) is not applicable to a mechanical system under the influence of a magnetic field in general. In this paper, we improve (1.1) so it can be applied to the case where a magnetic field is considered.

In mathematical language, the quantum mechanical motion of a charged particle in a magnetic field is described within the theory of complex line bundles in the following way (for details, see Ref. 7). A magnetic field is a real closed two-form Ω defined on M. This Ω is assumed to be integral, i.e., $[\Omega]/2\pi\in H^2(M;\mathbb{Z})$. The existence theorem ensures that one has a complex line bundle $\pi_E:E\to M$ and a linear connection ∇ with the curvature $i\Omega$. This E admits a Hermitian inner product $<\ ,\ >_E$ such that

$$X\langle s_1,s_2\rangle_E = \langle \nabla_X s_1,s_2\rangle_E + \langle s_1,\nabla_X s_2\rangle_E, \tag{1.2}$$

for arbitrary cross sections s_1 and s_2 and vector field X. Moreover, if M is simply connected, E and ∇ are uniquely determined up to strong bundle isomorphism.

The Schrödinger operator is then given as

$$\hat{H}\Psi(x) = \Big(-\frac{1}{2}\sum_{j,k} g^{jk}(x)\nabla_j\nabla_k + V(x)\Big)\Psi(x), \tag{1.3}$$

where ∇_j is a covariant derivative in the direction of $\partial/\partial x^j$ (the local coordinate frame), Ψ is a cross section of E, and g is a Riemannian metric.

On the other hand, in classical mechanics one has to introduce into T^*M a modified symplectic form

$$\sigma = d\theta_M + \pi_M^*\Omega \tag{1.4}$$

(called a charged symplectic form[8]) to describe the corresponding motion, where π_M: $T^*M\to M$ is the canonical projection. The motion is given by the Hamiltonian vector field on T^*M associated with

$$H(x;p) = \tfrac{1}{2}|p|^2 + V(x),$$

with respect to σ, where $(x;p)\in T^*M$, $x\in M$, and $p\in T_x^*M$. Thus we need to consider the symplectic manifold (T^*M,σ) instead of $(T^*M,d\theta_M)$. Let L be a Lagrangian submanifold of (T^*M,σ). While $\sigma|L = 0$, $d\theta_M$ is not necessarily vanishing on L. Hence θ_M cannot define an element of $H^1(L;\mathbb{R})$ so that action integrals $\int_\gamma \theta_M$ on L are not determined by the homology class $[\gamma]\in H_1(L;\mathbb{Z})$ in general. Thus the Maslov quantization condition (1.1) is not effective in the original form. In Sec. II, we propose an alternative quantization condition by means of a certain principal $U(1)$ bundle and connection form (cf. Ref. 9).

As an application of this quantization condition, we calculate quasiclassical eigenvalues of the Kepler problem with a centrifugal potential and Dirac's monopole field, which is referred to as the MIC–Kepler problem.[10] The quantized MIC–Kepler problem is given as follows.[11]

For every $m\in\mathbb{Z}$, Dirac's monopole field is defined by a closed two-form on $\mathring{\mathbb{R}}^3 = \mathbb{R}^3 - \{0\}$ such that

$$\widetilde{\Omega}_m = -(m/2)|\tilde{x}|^{-3}(\tilde{x}_1\, d\tilde{x}_2\wedge d\tilde{x}_3 + \tilde{x}_2\, d\tilde{x}_3\wedge d\tilde{x}_1 + \tilde{x}_3\, d\tilde{x}_1\wedge d\tilde{x}_2), \tag{1.5}$$

where $\tilde{x} = (\tilde{x}_1,\tilde{x}_2,\tilde{x}_3)\in\mathring{\mathbb{R}}^3$, $|\tilde{x}|^2 = \tilde{x}_1^2 + \tilde{x}_2^2 + \tilde{x}_3^2$. A simple

 0022-2488/90/061388-07$03.00

calculation yields

$$\int_{S(2)} \tilde{\Omega}_m = 2\pi m,$$

where $S(2)$ is the unit two-sphere, that is, $\tilde{\Omega}_m$ is integral. Since $\mathring{\mathbb{R}}^3$ is simply connected, we have the unique complex line bundle E_m over $\mathring{\mathbb{R}}^3$ with the Hermitian inner product $\langle\ ,\ \rangle_m$ and the linear connection ∇^m with the curvature form $i\tilde{\Omega}_m$ by virtue of the existence theorem. The Hamiltonian of the quantized MIC–Kepler problem is given by

$$\hat{H}_m = -\frac{1}{2}\sum_{j=1}^{3} (\nabla_j^m)^2 + \frac{(m/2)^2}{2|\tilde{x}|^2} - \frac{k}{|\tilde{x}|}, \tag{1.6}$$

where ∇_j^m stands for the covariant derivative in the direction of $\partial/\partial\tilde{x}_j$, $j = 1,2,3$, and k is a positive constant. The domain of $\hat{H}_m$ is contained in the space of all L^2 sections of E_m. The eigenvalue problem is exactly solved.[11] Consider the nonnegative integer n subject to the condition

$$|m| \leqslant n,\ n - m \text{ is even.} \tag{1.7}$$

The eigenvalues of $\hat{H}_m$ and their multiplicities are

$$e_n = -2k^2/(n+2)^2, \tag{1.8}$$

$$(n-m+2)(n+m+2)/4, \tag{1.9}$$

respectively, where n satisfies (1.7).

On the other hand, the corresponding classical mechanical system is of the following form. The symplectic manifold is $(T^*\mathring{\mathbb{R}}^3, \sigma_m)$, where

$$\sigma_m = \sum_{j=1}^{3} d\tilde{p}_j \wedge d\tilde{x}_j + \pi^*\tilde{\Omega}_m, \tag{1.10}$$

$(\tilde{x};\tilde{p}) \in T^*\mathring{\mathbb{R}}^3 = \mathring{\mathbb{R}}^3 \times \mathbb{R}^3$ and π: $T^*\mathring{\mathbb{R}}^3 \to \mathring{\mathbb{R}}^3$ is the canonical projection. The classical Hamiltonian of the MIC–Kepler problem is given by

$$H_m(\tilde{x};\tilde{p}) = \frac{1}{2}|\tilde{p}|^2 + \frac{(m/2)^2}{2|\tilde{x}|^2} - \frac{k}{|\tilde{x}|}. \tag{1.11}$$

Iwai and Uwano[10] showed $(T^*\mathring{\mathbb{R}}^3, \sigma_m, H_m)$ is obtained by the $U(1)$ reduction of the conformal Kepler problem. In Sec. III, in the above framework we relate the principal $U(1)$ bundle and the connection form, which are used in defining the quantization condition, to the cotangent bundle $T^*\mathring{\mathbb{R}}^4$ and its canonical one-form.

In Sec. IV, we investigate the $U(1)$ reduction of the Maslov class. With the aid of the results in Sec. III and IV, we calculate the quasiclassical eigenvalues and their multiplicities of $\hat{H}_m$ in Sec. V, to see these coincide with (1.8) and (1.9), respectively.

II. A QUANTIZATION CONDITION IN THE PRESENCE OF A MAGNETIC FIELD

Consider the principal $U(1)$ bundle ν_p:$P \to M$ associated with the complex line bundle π_E:$E \to M$ with the Hermitian metric $\langle\ ,\ \rangle_E$. Let ∇ be a linear connection in E with the property (1.2), whose curvature form is $i\Omega$. Here P is endowed with the connection form β induced by the linear connection ∇. The curvature form of β is Ω. We define the pullback bundle $B = \pi_M^{-1}P$ over T^*M with the commutative diagram

$$\begin{array}{ccc} B & \xrightarrow{\ p\ } & P \\ {\scriptstyle \nu_B}\downarrow & & \downarrow{\scriptstyle \nu_P} \\ T^*M & \xrightarrow{\ \pi_M\ } & M \end{array} \tag{2.1}$$

Set a one-form on B:

$$\alpha = \nu_B^*\theta_M + p^*\beta. \tag{2.2}$$

Then α turns out to be a connection form on B with the curvature form σ defined by (1.4). Suppose L is a Lagrangian submanifold of (T^*M,σ). It holds that $d\alpha|\nu_B^{-1}(L) = 0$ since $\sigma|L = 0$. Thus α defines an element of $H^1(\nu_B^{-1}(L);\mathbb{R})$ so that $\int_\gamma \alpha$, for an arbitrary closed curve γ in $\nu_B^{-1}(L)$, depends on $[\gamma] \in H_1(\nu_B^{-1}(L);\mathbb{Z})$. Now, L satisfies the quantization condition if and only if

$$\frac{1}{2\pi}\int_\gamma \alpha - \frac{1}{4}\langle \nu_B^*\mu_L,[\gamma]\rangle = \text{integer}, \tag{2.3}$$

for any closed curve γ in $\nu_B^{-1}(L)$. If $\Omega = 0$ and $E = M \times \mathbb{C}$, then we may take $B = T^*M \times U(1)$ and $\alpha = \theta_M + dz/iz$, where $z \in \mathbb{C}$, $|z| = 1$. In this case, (2.3) is equivalent to (1.1). Thus the quantization condition (2.3) can be regarded as a generalization of the Maslov one. Using (2.3) instead of (1.1), we can define a quasiclassical state (QCS, for short), a quasiclassical eigenstate (QCE), and a quasiclassical eigenvalue in parallel with those in Ref. 4. Namely, a compact connected Lagrangian submanifold $L \subset (T^*M,\sigma)$ is a QCS if L satisfies (2.3). Let H be a smooth function on T^*M. A QCS L is called a QCE of H when $H|L = c$ (const). This c is called a quasiclassical eigenvalue of H.

III. (B, α) AND THE $U(1)$ REDUCTION OF A PHASE SPACE

In this section, we review the construction of the complex line bundle E_m and the connection ∇^m for the quantized MIC–Kepler problem first.[11,7] Next, we recall briefly the $U(1)$ reduction of a phase space, through which the classical MIC–Kepler problem is obtained from the conformal Kepler problem in $T^*\mathring{\mathbb{R}}^4$.[10] Within this framework, we show the principal $U(1)$ bundle and the connection form given in (2.1) and (2.2) can be related to $T^*\mathring{\mathbb{R}}^4$ and its canonical one-form, respectively.

We identify $x = (x_1,x_2,x_3,x_4) \in \mathbb{R}^4$ with $z = (z_1,z_2) \in \mathbb{C}^2$ through $z_1 = x_1 + ix_2$, $z_2 = x_3 + ix_4$. Consider a $U(1)$ action $R(e^{it})$ on $\mathring{\mathbb{R}}^4 = \mathbb{R}^4 - \{0\}$ such that

$$R(e^{it})z = (e^{it}z_1, e^{it}z_2). \tag{3.1}$$

Then we have a principal $U(1)$ bundle over $\mathring{\mathbb{R}}^3$, ν:$\mathring{\mathbb{R}}^4 \to \mathring{\mathbb{R}}^3$, where ν is given by $\tilde{x} = \nu(x) \in \mathring{\mathbb{R}}^3$, $x \in \mathring{\mathbb{R}}^4$,

$$\tilde{x}_1 = 2\,\mathrm{Re}\, z_1\bar{z}_2, \quad \tilde{x}_2 = 2\,\mathrm{Im}\, z_1\bar{z}_2, \quad \tilde{x}_3 = |z_1|^2 - |z_2|^2. \tag{3.2}$$

The fundamental vector field $\gamma(z) = (d/dt)|_{t=0} R(e^{it})z$ is written as $\gamma(z) = iz$. Let $\mathring{\mathbb{R}}^4$ be equipped with the Riemannian metric $g_0 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$. The one-form β_1 such that

$$\beta_1(z) = g_0(\gamma(z),*)/|\gamma(z)|^2 \tag{3.3}$$

defines a connection on $\mathring{\mathbb{R}}^4$. A calculation shows the curvature form $\tilde{\Omega}_1$ of β_1, i.e., $d\beta_1 = \nu^*\tilde{\Omega}_1$, is given by

$$\tilde{\Omega}_1(\tilde{x}) = -(1/2|\tilde{x}|^3)(\tilde{x}_1\, d\tilde{x}_2\wedge d\tilde{x}_3 + \tilde{x}_2\, d\tilde{x}_3\wedge d\tilde{x}_1 + \tilde{x}_3\, d\tilde{x}_1\wedge d\tilde{x}_2). \quad (3.4)$$

For each $m\in \mathbf{Z}$, let ρ_m be a $U(1)$ action on $\mathbf{C}$ defined by $\rho_m(e^{it})w = e^{imt}w$, $w\in\mathbf{C}$. Then the complex line bundle E_m and the linear connection ∇^m are obtained as follows: E_m is the associated bundle $E_m = \mathring{\mathbf{R}}^4\times_{\rho_m}\mathbf{C}$. The connection form β_1 naturally induces the linear connection ∇^m in E_m. The curvature form of ∇^m equals $im\tilde{\Omega}_1$.

As stated in the Introduction, the symplectic manifold $(T^*\mathring{\mathbf{R}}^3,\sigma_m)$, $\sigma_m = d\tilde{\theta} + \pi^* m\tilde{\Omega}_1$, is considered, where $\tilde{\theta}$ is the canonical one-form on $T^*\mathring{\mathbf{R}}^3$ and $\pi: T^*\mathring{\mathbf{R}}^3\to\mathring{\mathbf{R}}^3$ is the canonical projection. We denote by θ the canonical one-form on $T^*\mathring{\mathbf{R}}^4$. Let H_c be a Hamiltonian function of the conformal Kepler problem:

$$H_c(x;p) = \frac{1}{8|x|^2}\sum_{j=1}^{4} p_j^2 - \frac{k}{|x|^2}, \quad (3.5)$$

where $(x;p)\in T^*\mathring{\mathbf{R}}^4 = \mathring{\mathbf{R}}^4\times\mathbf{R}^4$. In what follows, we review the $U(1)$ reduction, through which $(T^*\mathring{\mathbf{R}}^3, \sigma_m, H_m)$ is obtained from $(T^*\mathring{\mathbf{R}}^4, d\theta, H_c)$ (for details, see Ref. 10). The $U(1)$ action R is lifted to be a symplectic action on $T^*\mathring{\mathbf{R}}^4$, which is denoted by $\hat{R}$. The moment map $\psi: T^*\mathring{\mathbf{R}}^4\to\mathbf{R}$ of $\hat{R}$ is defined by $\psi = \theta(\hat{\gamma})$, where $\hat{\gamma}$ is the induced vector field of $\hat{R}$ such that

$$\hat{\gamma}(x;p) = \frac{d}{dt}\bigg|_{t=0} \hat{R}(e^{it})(x;p).$$

We denote by H_ψ the Hamiltonian vector field of ψ. We note $\hat{\gamma} = H_\psi$. In an explicit manner, we have

$$\psi(x;p) = \langle p,\gamma(x)\rangle = x_1p_2 - x_2p_1 + x_3p_4 - x_4p_3. \quad (3.6)$$

We set a submanifold $F = \psi^{-1}(0)$ and the inclusion map ι_F: $F\to T^*\mathring{\mathbf{R}}^4$. Since ψ is $U(1)$ invariant, $U(1)$ acts freely on F. Hence we get a principal $U(1)$ bundle over $T^*\mathring{\mathbf{R}}^3$ with the $U(1)$ action $\hat{R}$, $\nu_F: F\to T^*\mathring{\mathbf{R}}^3$. It is easily checked that

$$\nu_F^*\tilde{\theta} = \iota_F^*\theta. \quad (3.7)$$

We define a diffeomorphism τ_m: $T^*\mathring{\mathbf{R}}^4\to T^*\mathring{\mathbf{R}}^4$ by $\tau_m(x;p) = (x;p - m\beta_1(x))$, where $\beta_1(x)$ is regarded as an element of $T_x^*\mathring{\mathbf{R}}^4$. The diffeomorphism τ_m commutes with $\hat{R}$. Let $\pi_{P_1}: T^*\mathring{\mathbf{R}}^4\to\mathring{\mathbf{R}}^4$ be the canonical projection. It follows that

$$\tau_m^*\theta = \theta - m\pi_{P_1}^*\beta_1. \quad (3.8)$$

Consider a submanifold $\psi^{-1}(m)$. Then τ_m maps $\psi^{-1}(m)$ diffeomorphically to F. We set $\nu_m = \nu_F\circ\tau_m$.

Theorem 3.1 (Iwai and Uwano[10]): $(T^*\mathring{\mathbf{R}}^4, d\theta, H_c)$ is reduced to $(T^*\mathring{\mathbf{R}}^3,\sigma_m, H_m)$. Namely, (i) ν_m: $\psi^{-1}(m)\to T^*\mathring{\mathbf{R}}^3$, is a principal $U(1)$ bundle with the $U(1)$ action $\hat{R}$; (ii) $\nu_m^*\sigma_m = \iota_m^*\, d\theta$; and (iii) $\nu_m^* H_m = \iota_m^* H_c$, where $\iota_m: \psi^{-1}(m)\to T^*\mathring{\mathbf{R}}^4$, is an inclusion map.

In addition to H_c, we set $U(1)$ invariant functions on $T^*\mathring{\mathbf{R}}^4$:

$$J(x;p) = \tfrac{1}{2}(x_1p_4 - x_4p_1 + x_3p_2 - x_2p_3), \quad (3.9)$$

$$D(x;p) = \tfrac{1}{4}(p_1p_3 + p_2p_4) - 2(x_1x_3 + x_2x_4)H_c(x;p). \quad (3.10)$$

Here H_c, J, and D are in involution, i.e., commutative with respect to the Poisson bracket defined by $d\theta$. On the other hand, we consider functions on $T^*\mathring{\mathbf{R}}^3$:

$$J_m(\tilde{x};\tilde{p}) = \tilde{x}_2\tilde{p}_3 - \tilde{x}_3\tilde{p}_2 + (m/2)\tilde{x}_1/|\tilde{x}|, \quad (3.11)$$

$$D_m(\tilde{x};\tilde{p}) = -\tilde{x}_1|\tilde{p}|^2 + \tilde{p}_1\langle\tilde{x},\tilde{p}\rangle + (m/2)(\tilde{x}_2\tilde{p}_3 - \tilde{x}_3\tilde{p}_2)/|\tilde{x}| + k\tilde{x}_1/|\tilde{x}|. \quad (3.12)$$

Proposition 3.2 (Iwai and Uwano[10]): It holds that

$$\nu_m^* J_m = J, \quad \nu_m^* D_m = D.$$

Moreover, H_m, J_m, and D_m are commutative with respect to the Poisson bracket defined by σ_m.

Now, let $\nu_{P_m}: P_m\to\mathring{\mathbf{R}}^3$ be the principal $U(1)$ bundle associated with the Hermitian line bundle $(E_m,\langle\,;\,\rangle_m)$. Here we have $P_1 = \mathring{\mathbf{R}}^4$, $\nu_{P_1} = \nu$. We denote by R_m the $U(1)$ action of the principal bundle P_m and by γ_m the fundamental vector field on it. The bundle P_m is equipped with the connection form β_m induced by ∇^m. For P_m and β_m, we denote by $\nu_{B_m}: B_m\to T^*\mathring{\mathbf{R}}^3$, the principal $U(1)$ bundle over $T^*\mathring{\mathbf{R}}^3$ defined by (2.1) and by α_m the connection form defined by (2.2), respectively. In what follows, we see B_m and α_m are related to $T^*\mathring{\mathbf{R}}^4$ and θ, respectively. Let $\{\varphi_\alpha^{(m)}, U_\alpha\}_{\alpha\in\Lambda}$ be the local trivialization of P_m; $\{U_\alpha\}_{\alpha\in\Lambda}$ is an open covering of $\mathring{\mathbf{R}}^3$ and

$$\varphi_\alpha^{(m)}: U_\alpha\times U(1)\tilde{\to}\nu_{P_m}^{-1}(U_\alpha)$$

is a diffeomorphism with the $U(1)$ equivariancy. Define a bundle homomorphism $\Phi^{(m)}: P_1 - P_m$ by

$$\varphi_\alpha^{(m)-1}\circ\Phi^{(m)}\circ\varphi_\alpha^{(1)}(x,w) = (x,w^m),$$

$(x,w)\in U_\alpha\times U(1)$, for every $\alpha\in\Lambda$. Then it holds that $\Phi^{(m)}\circ R_1(e^{it}) = R_m(e^{it})^m\circ\Phi^{(m)}$, so that $d\Phi^{(m)}\ \gamma_1 = m\gamma_m$. It follows that

$$\Phi^{(m)*}\beta_m = m\beta_1. \quad (3.13)$$

Recall that B_m is given explicitly by

$$B_m = \{((\tilde{x};\tilde{p}),\xi)\in T^*\mathring{\mathbf{R}}^3\times P_m\,|\,\nu_{P_m}(\xi) = \tilde{x}\}. \quad (3.14)$$

The $U(1)$ action R_m on P_m naturally induces that on B_m, which is denoted by the same letter, R_m. In terms of (3.14), we define a bundle homomorphism $p_m: B_m\to P_m$ by $p_m((\tilde{x};\tilde{p}),\xi) = \xi$ and a bundle homomorphism $\Psi^{(m)}: B_1\to B_m$ by $\Psi^{(m)} = I\times\Phi^{(m)}$, where I is the identity map of $T^*\mathring{\mathbf{R}}^3$. This $\Psi^{(m)}$ satisfies

$$\nu_{B_m}\circ\Psi^{(m)} = \nu_{B_1},$$

$$p_m\circ\Psi^{(m)} = \Phi^{(m)}\circ p_1,$$

$$\Psi^{(m)}\circ R_1(e^{it}) = R_m(e^{it})^m\circ\Psi^{(m)}.$$

Combining these with (3.13), we get

$$\Psi^{(m)*}\alpha_m = \nu_{B_1}^*\tilde{\theta} + mp_1^*\beta_1. \quad (3.15)$$

Define a bundle isomorphism f: $F\to B_1$ by $f(x;p) = (\nu_F(x;p),x)$. We have

$$\nu_{B_1}\circ f = \nu_F, \quad p_1\circ f = \pi_{P_1}\circ\iota_F, \quad f\circ\hat{R}(e^{it}) = R_1(e^{it})\circ f,$$

so that

$$\Psi^{(m)}\circ f\circ\hat{R}(e^{it}) = R_m(e^{it})^m\circ\Psi^{(m)}\circ f.$$

Equations (3.7) and (3.15) yield

$$f^*\Psi^{(m)*}\alpha_m = \iota_F^*(\theta + m\pi_{P_1}^*\beta_1). \quad (3.16)$$

Set $\kappa_m: \psi^{-1}(m) \to B_m$ as $\kappa_m = \Psi^{(m)} \circ f \circ \tau_m$. Then, by means of (3.8) and (3.16), we get a relation between B_m, α_m and $T^*\mathring{\mathbb{R}}^4$, θ such as the following proposition.

Proposition 3.3: $\kappa_m: \psi^{-1}(m) \to B_m$ is a bundle homomorphism with the properties

$$\kappa_m \circ \hat{R}(e^{it}) = R_m(e^{it})^m \circ \kappa_m, \quad \kappa_m^* \alpha_m = \iota_m^* \theta.$$

IV. MASLOV CLASS

Let L be a Lagrangian submanifold of $(T^*\mathring{\mathbb{R}}^3, \sigma_m)$. Since $\nu_m^* \sigma_m = \iota_m^* d\theta$, $\hat{L} = \nu_m^{-1}(L)$ is also a Lagrangian submanifold of $(T^*\mathring{\mathbb{R}}^4, d\theta)$, which is contained in $\psi^{-1}(m)$. In this section, we investigate a relation between the Maslov classes of L and $\hat{L}$.[9] In Appendix A, we recall the definition of the Maslov class briefly.

We denote by μ_L and $\mu_{\hat{L}}$ the Maslov class of L and that of $\hat{L}$, respectively. We take a Riemannian metric $\tilde{g}$ on $\mathring{\mathbb{R}}^3$ such that $\nu:(\mathring{\mathbb{R}}^4, g_0) \to (\mathring{\mathbb{R}}^3, \tilde{g})$ is a Riemannian submersion, namely, $\tilde{g}(\tilde{X}, \tilde{Y}) = g_0(X^*, Y^*)$, where $\tilde{X}$, $\tilde{Y} \in T_{\tilde{x}}\mathring{\mathbb{R}}^3$ and X^*, Y^* are their horizontal lifts. For every $y = (x;p) \in T^*\mathring{\mathbb{R}}^4 = \mathring{\mathbb{R}}^4 \times \mathbb{R}^4$, we set $\mathscr{L}_F(y) = T_p T_x^* \mathring{\mathbb{R}}^4$, the Lagrangian subspace tangent to the fiber $T_x^*\mathring{\mathbb{R}}^4$. In what follows, we construct a Lagrangian subspace $\mathscr{L}_B(y) \subset T_y T^*\mathring{\mathbb{R}}^4$ transverse to $\mathscr{L}_F(y)$. We take orthonormal vectors $u_1, u_2, u_3, u_4 \in T_x \mathring{\mathbb{R}}^4$ such that $u_4 = \gamma(x)/|\gamma(x)|$. Suppose u_j ($j = 1,2,3,4$) are written as

$$u_j = \sum_{\alpha=1}^{4} u_{j\alpha} \frac{\partial}{\partial x_\alpha}.$$

We define orthonormal vectors $f_j \in \mathscr{L}_F(y)$ ($j = 1,2,3,4$) by

$$f_j = \sum_{\alpha=1}^{4} u_{j\alpha} \frac{\partial}{\partial p_\alpha}.$$

Set tangent vectors b_j ($j = 1,2,3,4$) of $T_y T^*\mathring{\mathbb{R}}^4$ as

$$b_4 = H_\psi(y)/|\gamma(x)| \quad [y = (x;p)],$$

$$b_j = u_j - d\theta(u_j, b_4) f_4 \quad (j = 1,2,3),$$

where H_ψ is the Hamiltonian vector field of ψ. Then $d\theta(f_4, b_4) = 1$ and $d\theta(f_4, u_j) = 0$ ($j = 1,2,3$). It follows that $\{f_j, b_j\ (j = 1,2,3,4)\}$ is a symplectic basis, i.e.,

$$d\theta(f_j, f_k) = d\theta(b_j, b_k) = 0,$$

$$d\theta(f_j, b_k) = \delta_{jk} \quad (j,k = 1,2,3,4). \tag{4.1}$$

Hence we define $\mathscr{L}_B(y) = \mathrm{span}\{b_j\ (j = 1,2,3,4)\}$, a Lagrangian subspace transverse to $\mathscr{L}_F(y)$. We take $y \in \psi^{-1}(m)$ and set $\tilde{y} = \nu_m(y)$. In what follows, we construct Lagrangian subspaces $\tilde{\mathscr{L}}_F(\tilde{y})$ and $\tilde{\mathscr{L}}_B(\tilde{y}) \subset T_{\tilde{y}} T^*\mathring{\mathbb{R}}^3$ out of $\mathscr{L}_F(y)$ and $\mathscr{L}_B(y)$, which are transverse to each other. Theorem 3.1 shows

$$d\nu_m(y): T_y \psi^{-1}(m) \to T_{\tilde{y}} T^*\mathring{\mathbb{R}}^3$$

is an onto map with the kernel spanned by $H_\psi(y)$. Set a subspace N_y of $T_y T^*\mathring{\mathbb{R}}^4$ as

$$N_y = \{Y \in T_y \psi^{-1}(m) | \pi_{P_1}^* \beta_1(Y) = 0\}.$$

Since $\pi_{P_1}^* \beta_1(H_\psi) = 1$, we have $T_y \psi^{-1}(m) = \mathbb{R} H_\psi(y) + N_y$ (a direct sum). Thus the restriction of $d\nu_m(y)$ to N_y yields a linear isomorphism between N_y and $T_{\tilde{y}} T^*\mathring{\mathbb{R}}^3$. Equations (4.1) indicate f_j ($j = 1,2,3$) and b_k ($k = 1,2,3$) belong to N_y. Note that b_4 spans the kernel of $d\nu_m(y)$. We then get a symplectic basis of $(T_{\tilde{y}} T^*\mathring{\mathbb{R}}, \sigma_m(\tilde{y}))$ such that

$$\tilde{f}_j = d\nu_m(y) f_j, \quad \tilde{b}_j = d\nu_m(y) b_j \quad (j = 1,2,3). \tag{4.2}$$

Thus we have that $\tilde{\mathscr{L}}_F(\tilde{y})$ is spanned by $\tilde{f}_j$ ($j = 1,2,3$) and we set $\tilde{\mathscr{L}}_B(\tilde{y}) = \mathrm{span}\{\tilde{b}_j\ (j = 1,2,3)\}$. Here $\tilde{\mathscr{L}}_B(\tilde{y})$ is a Largrangian subspace transverse to $\tilde{\mathscr{L}}_F(\tilde{y})$.

By means of $\{f_j, b_j\ (j = 1,2,3,4)\}$ and $\{\tilde{f}_j, \tilde{b}_j\ (j = 1,2,3)\}$, we identify $T_y T^*\mathring{\mathbb{R}}^4 \tilde{\to} \mathbb{C}^4$ and $T_{\tilde{y}} T^*\mathring{\mathbb{R}}^3 \tilde{\to} \mathbb{C}^3$ in the standard manner, respectively [see Appendix A, (A2)]. For $y \in \hat{L}$, $d\nu_m(y)$: $T_y \hat{L} \to T_{\tilde{y}} L$ is onto and the kernel is spanned by b_4. If a unitary matrix $W \in U(3)$ satisfies $T_{\tilde{y}} L = W \cdot \tilde{\mathscr{L}}_B(\tilde{y})$, then the unitary matrix $\hat{W} \in U(4)$ with $T_y \hat{L} = \hat{W} \cdot \mathscr{L}_B(y)$ is given by $\hat{W} = \left(\begin{smallmatrix} W & 0 \\ 0 & 1 \end{smallmatrix}\right)$ owing to (4.2). Let Det^2: $\Lambda(T^*\mathring{\mathbb{R}}^4) \to U(1)$ and $\widetilde{\mathrm{Det}}^2$: $\Lambda(T^*\mathring{\mathbb{R}}^3) \to U(1)$ be the maps given in (A3) of Appendix A, respectively. Also denote by λ: $\hat{L} \to \Lambda(T^*\mathring{\mathbb{R}}^4)$ and $\tilde{\lambda}$: $L \to \Lambda(T^*\mathring{\mathbb{R}}^3)$ the maps defined by (A4) in Appendix A, respectively. The above arguments show $\mathrm{Det}^2 \circ \lambda(y) = \widetilde{\mathrm{Det}}^2 \circ \tilde{\lambda}(\nu_m(y))$. Thus we get the following theorem.

Theorem 4.1: $\mu_{\hat{L}} = \nu_m^* \mu_L$, where we consider ν_m as a map of $\hat{L}$ to L.

V. QUASICLASSICAL EIGENVALUES OF THE MIC-KEPLER PROBLEM

For constants E (>0), $\bar{J}_m$, and $\bar{D}_m$, we set a level set

$$L(E, \bar{J}_m, \bar{D}_m) = \{(\tilde{x};\tilde{p}) \in T^*\mathring{\mathbb{R}}^3 | H_m(\tilde{x};\tilde{p}) = -E,$$
$$J_m(\tilde{x};\tilde{p}) = \bar{J}_m, \quad D_m(\tilde{x};\tilde{p}) = \bar{D}_m\}.$$

Recall that H_m, J_m, and D_m Poisson–commute (see Proposition 3.2). Owing to the Liouville–Arnold theorem,[12] $L(E, \bar{J}_m, \bar{D}_m)$ is a Lagrangian submanifold of $(T^*\mathring{\mathbb{R}}^3, \sigma_m)$ if dH_m, dJ_m, and dD_m are linearly independent at each point of $L(E, \bar{J}_m, \bar{D}_m)$. Set a level set of $T^*\mathring{\mathbb{R}}^4$:

$$\hat{L}(E, \bar{J}_m, \bar{D}_m) = \{(x;p) \in T^*\mathring{\mathbb{R}}^4 | H_c(x;p) = -E,$$
$$J(x;p) = \bar{J}_m, \quad D(x;p) = \bar{D}_m,$$
$$\psi(x;p) = m\}.$$

Because of Theorem 3.1 and Proposition 3.2,

$$\nu_m^{-1}(L(E, \bar{J}_m, \bar{D}_m)) = \hat{L}(E, \bar{J}_m, \bar{D}_m).$$

First of all, we investigate a condition on E, $\bar{J}_m$, and $\bar{D}_m$ under which $\hat{L}(E, \bar{J}_m, \bar{D}_m)$ [hence $L(E, \bar{J}_m, \bar{D}_m)$] becomes a Lagrangian submanifold. We introduce the complex variables ζ_j ($j = 1,2,3,4$):

$$\zeta_1 = \lambda(x_1 + x_3) - p_2 - p_4 + i(\lambda(x_2 + x_4) + p_1 + p_3),$$
$$\zeta_2 = \lambda(x_1 - x_3) - p_2 + p_4 + i(\lambda(x_2 - x_4) + p_1 - p_3),$$
$$\zeta_3 = \lambda(x_1 + x_3) + p_2 + p_4 + i(\lambda(x_2 + x_4) - p_1 - p_3),$$
$$\zeta_4 = \lambda(x_1 - x_3) + p_2 - p_4 + i(\lambda(x_2 - x_4) - p_1 + p_3), \tag{5.1}$$

where $\lambda = \sqrt{8E}$. It follows that

$$32|x|^2(H_c(x;p) + E) + 32k = |\zeta_1|^2 + |\zeta_2|^2 + |\zeta_3|^2 + |\zeta_4|^2,$$
$$J(x;p) = -(1/16\lambda)(|\zeta_1|^2 - |\zeta_2|^2 - |\zeta_3|^2 + |\zeta_4|^2),$$
$$D(x;p) + 2(x_1 x_3 + x_2 x_4)(H_c(x;p) + E)$$
$$= (1/32)(|\zeta_1|^2 - |\zeta_2|^2 + |\zeta_3|^2 - |\zeta_4|^2),$$
$$\psi(x;p) = (1/8\lambda)(-|\zeta_1|^2 - |\zeta_2|^2 + |\zeta_3|^2 + |\zeta_4|^2). \tag{5.2}$$

We set constants

$$\begin{aligned}\Lambda_1 &= 8k - 2\lambda m - 4\lambda \bar{J}_m + 8\bar{D}_m,\\ \Lambda_2 &= 8k - 2\lambda m + 4\lambda \bar{J}_m - 8\bar{D}_m,\\ \Lambda_3 &= 8k + 2\lambda m + 4\lambda \bar{J}_m + 8\bar{D}_m,\\ \Lambda_4 &= 8k + 2\lambda m - 4\lambda \bar{J}_m - 8\bar{D}_m.\end{aligned} \tag{5.3}$$

In accordance with (5.2), $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ is given by

$$|\zeta_j|^2 = \Lambda_j \quad (j = 1,2,3,4). \tag{5.4}$$

Proposition 5.1: $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ is a Lagrangian submanifold if and only if

$$\Lambda_j > 0, \quad j = 1,2,3,4. \tag{5.5}$$

Moreover, $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ is diffeomorphic to a four-dimensional torus T^4.

By means of (5.4), we parametrize $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ as $\zeta_j = \sqrt{\Lambda_j} e^{it_j}$, $t_j \in [0,2\pi]$ ($j = 1,2,3,4$), where the Λ_j are subject to (5.5). Since H_c, J, D, and ψ are $U(1)$ invariant, $\hat{R}(e^{it})$ acts on $\hat{L}(E,\bar{J}_m,\bar{D}_m)$. In terms of this parametrization, the $U(1)$ action is written as

$$\hat{R}(e^{it})(t_1,\ t_2,\ t_3,\ t_4) = (t_1 + t, t_2 + t, t_3 + t, t_4 + t) \pmod{2\pi}.$$

Hence, we introduce another parametrization $s = (s_1,s_2,s_3,s_4)$, for example, $t_j = s_j + s_4$ ($j = 1,2,3$), $t_4 = s_4$, namely,

$$\zeta_j = \sqrt{\Lambda_j} e^{i(s_j + s_4)} \quad (j = 1,2,3), \quad \zeta_4 = \sqrt{\Lambda_4} e^{is_4}. \tag{5.6}$$

By means of (5.6), the $U(1)$ action can be expressed as

$$\hat{R}(e^{it})(s_1,s_2,s_3,s_4) = (s_1,s_2,s_3,s_4 + t) \pmod{2\pi}.$$

Thus $L(E,\bar{J}_m,\bar{D}_m) \cong \hat{L}(E,\bar{J}_m,\bar{D}_m)/U(1)$ is parametrized by $(s_1,s_2,s_3) \pmod{2\pi}$, which shows $L(E,\bar{J}_m,\bar{D}_m) \cong T^3$. We define closed curves c_j ($j = 1,2,3,4$) in $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ by

$$c_j\colon\ s_j = t, \quad \text{others} = 0 \quad (j = 1,2,3,4). \tag{5.7}$$

Note that the c_j ($j = 1,2,3,4$) generate $H_1(\hat{L}(E,\bar{J}_m,\bar{D}_m);\mathbb{Z})$, and $[c_j]$ denotes the class of c_j. In terms of (5.6), we define a global section $q\colon L(E,\bar{J}_m,\bar{D}_m) \to \hat{L}(E,\bar{J}_m,\bar{D}_m)$ by $q(s_1,s_2,s_3) = (s_1,s_2,s_3,0)$. The κ_m given in Proposition 3.3 induces a bundle homomorphism of $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ to $\nu_{B_m}^{-1}(L(E,\bar{J}_m,\bar{D}_m))$. Then,

$$\kappa_m \circ q\colon\ L(E,\bar{J}_m,\bar{D}_m) \to \nu_{B_m}^{-1}(L(E,\bar{J}_m,\bar{D}_m))$$

provides a global section. Thus we have trivial bundles $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ and $\nu_{B_m}^{-1}(L(E,\bar{J}_m,\bar{D}_m))$ over $L(E,\bar{J}_m,\bar{D}_m)$, which are diffeomorphic to T^4. We denote by u the generator of $H_1(\nu_{B_m}^{-1}(L(E,\bar{J}_m,\bar{D}_m));\mathbb{Z})$ defined by the fiber $U(1)$, so that the first homology class is generated by $\kappa_{m*}[c_j]$ ($j = 1,2,3$) and u. We have $\kappa_{m*}[c_4] = mu$.

Now, in order to calculate quasiclassical eigenvalues of H_m we check the quantization condition

$$\frac{1}{2\pi}\int_{\kappa_{m*}[c_j]} \alpha_m - \frac{1}{4}\langle \nu_{B_m}^* \mu_L, \kappa_{m*}[c_j]\rangle = \text{integer} \qquad (j = 1,2,3),$$

$$\frac{1}{2\pi}\int_u \alpha_m - \frac{1}{4}\langle \nu_{B_m}^* \mu_L, u\rangle = \text{integer}, \tag{5.8}$$

where μ_L is the Maslov class of $L(E,\bar{J}_m,\bar{D}_m)$. Note that

$$\int_u \alpha_m = 2\pi, \quad \nu_{B_m*} u = 0,$$

so that the last equation trivially holds. Because of Proposition 3.3 and Theorem 4.1, (5.8) is equivalent to

$$\frac{1}{2\pi}\int_{[c_j]} \theta - \frac{1}{4}\langle \mu_{\hat{L}}, [c_j]\rangle = \text{integer} \quad (j = 1,2,3), \tag{5.9}$$

where $\mu_{\hat{L}}$ is the Maslov class of $\hat{L}(E,\bar{J}_m,\bar{D}_m)$. First, we compute action integrals along c_j ($j = 1,2,3$). Here, θ is expressed in terms of ζ_j as

$$\begin{aligned}\theta = (1/8\lambda)\operatorname{Im}(&\zeta_1\, d\bar{\zeta}_1 + \zeta_2\, d\bar{\zeta}_2 - \zeta_3\, d\bar{\zeta}_3 - \zeta_4\, d\bar{\zeta}_4\\ &+ \zeta_1\, d\bar{\zeta}_3 - \zeta_3\, d\bar{\zeta}_1 + \zeta_2\, d\bar{\zeta}_4 - \zeta_4\, d\bar{\zeta}_2).\end{aligned}$$

Hence we have

$$\theta|_{c_j} = -(\Lambda_j/8\lambda)dt + dF \quad (j = 1,2),$$

$$\theta|_{c_3} = (\Lambda_3/8\lambda)dt + dF,$$

where $F = (1/8\lambda)\operatorname{Im}(\zeta_1\bar{\zeta}_3 + \zeta_2\bar{\zeta}_4)$. Thus we get the following proposition.

Proposition 5.2:

$$\int_{c_j} \theta = -\frac{\Lambda_j}{4\lambda}\pi \quad (j = 1,2), \quad \int_{c_3} \theta = \frac{\Lambda_3}{4\lambda}\pi.$$

Next we compute Maslov indices along c_j ($j = 1,2,3$). Define

$$G(x;p) = {}^t(H(x;p),\psi(x;p),J(x;p),D(x;p))$$

and differentiate G by x_j and p_j ($j = 1,2,3,4$). Then we get 4×4 matrices $G_x = \partial_x G$ and $G_p = \partial_p G$. Because of Proposition 3.2 in Ref. 6, we have

$$\mu_{\hat{L}} = (1/\pi)d \arg\det(G_p + iG_x)\,|\hat{L}(E,\bar{J}_m,\bar{D}_m). \tag{5.10}$$

By means of (5.10), we can calculate Maslov indices for $[c_j]$ ($j = 1,2,3$) explicitly to get the next proposition.

Proposition 5.3:

$$\langle \mu_{\hat{L}}, c_j\rangle = -2 \quad (j = 1,2), \quad \langle \mu_{\hat{L}}, c_3\rangle = 2.$$

(For a proof, see Appendix B.)

Now, we are in a position to calculate the quasiclassical eigenvalues. By means of Propositions 5.2 and 5.3, (5.9) is quivalent to

$$(1/2\pi)(-\Lambda_j/4\lambda)\pi - \tfrac{1}{4}(-2) = -n_j \quad (j = 1,2),$$

$$(1/2\pi)(\Lambda_3/4\lambda)\pi - \tfrac{1}{4}2 = n_3,$$

namely,

$$\Lambda_j = 8\lambda(n_j + \tfrac{1}{2}) \quad (j = 1,2,3), \tag{5.11}$$

where n_j ($j = 1,2,3$) are integers. We set $n_4 = n_1 + n_2 - n_3 + m \in \mathbb{Z}$. According to (5.3), (5.11) is equivalent to

$$\begin{aligned}8k &= 2\lambda(n_1 + n_2 + n_3 + n_4 + 2),\\ 4\lambda\bar{J}_m &= 2\lambda(-n_1 + n_2 + n_3 - n_4),\\ 8\bar{D}_m &= 2\lambda(n_1 - n_2 + n_3 - n_4).\end{aligned} \tag{5.12}$$

It follows that $\Lambda_4 = 8\lambda(n_4 + \frac{1}{2})$. Substituting (5.11) and Λ_4 into (5.5), we have

$$n_j \geqslant 0 \quad (j = 1,2,3,4). \tag{5.13}$$

If we set $n = n_1 + n_2 + n_3 + n_4$, $\lambda = \sqrt{8E}$ and (5.12) give

$$\begin{aligned} E &= 2k^2/(n+2)^2 \quad [= -e_n, \text{ see } (1.8)]. \\ \bar{J}_m &= (-n_1 + n_2 + n_3 - n_4)/2, \\ \bar{D}_m &= k(n_1 - n_2 + n_3 - n_4)/(n+2). \end{aligned} \tag{5.14}$$

Note that $n_1 + n_2 = (n-m)/2$, $n_3 + n_4 = (n+m)/2$. Relation (5.13) implies

$$|m| \leqslant n, \quad n - m \text{ is even.} \tag{5.15}$$

Thus, in accordance with Proposition 5.1, we see that $L(E, \bar{J}_m, \bar{D}_m)$ satisfies the quantization condition (5.8) if and only if E, $\bar{J}_m$, and $\bar{D}_m$ are subject to (5.14) and (5.15). For each n obeying (5.15), the number of $L(-e_n, \bar{J}_m, \bar{D}_m)$'s that satisfy (5.8) is equal to the number of (n_1, n_2, n_3, n_4)'s that satisfy $n_1 + n_2 + n_3 + n_4 = n$, $-n_1 - n_2 + n_3 + n_4 = m$, $n_j \geqslant 0$. Then it coincides with the multiplicity of the eigenspace of $\hat{H}_m$ belonging to e_n, $(n - m + 2)(n + m + 2)/4$ [see (1.9)]. In conclusion, we get the following.

Theorem 5.3: The quasiclassical eigenvalues of H_m are just equal to the eigenvalues of $\hat{H}_m$. Moreover, for each eigenvalue e_n, the number of the Lagrangian submanifolds $L(-e_n, \bar{J}_m, \bar{D}_m)$ satisfying the quantization condition (1.1) is equal to the multiplicity of $\hat{H}_m$ belonging to e_n.

ACKNOWLEDGMENTS

The authors would like to thank Professor T. Iwai and Dr. Y. Uwano for stimulating discussions.

APPENDIX A: DEFINITION OF THE MASLOV CLASS

Let Q be an n-dimensional smooth manifold. We suppose Q is equipped with a Riemannian metric g. Let Ω_Q be a closed two-form on Q. We consider a symplectic manifold (T^*Q, σ_Q) with $\sigma_Q = d\theta_Q + \pi_Q^*\Omega_Q$, where θ_Q is the canonical one-form and $\pi_Q: T^*Q \to Q$ is the canonical projection. The Lagrangian Grassmannian manifold $\Lambda(T_yT^*Q)$ of T_yT^*Q is a collection of all Lagrangian subspaces of $(T_yT^*Q, \sigma_Q(y))$, where $y = (x;p) \in T^*Q$, $x \in Q$, $p \in T_x^*Q$. Now, $T_pT_x^*Q$, the tangent space to the fiber T_x^*Q at p, is an element of $\Lambda(T_yT^*Q)$. We denote it by $\mathscr{L}_F(y)$. One can choose a Lagrangian subspace $\mathscr{L}_B(y)$ transverse to $\mathscr{L}_F(y)$, i.e., $T_yT^*Q = \mathscr{L}_F(y) + \mathscr{L}_B(y)$ (a direct sum). One may assume $\mathscr{L}_B(y)$ depends smoothly on y. Owing to the linear structure of T_x^*Q, $\mathscr{L}_F(y)$ is endowed with an inner product induced from the fiber metric of T_x^*Q defined by g, which is denoted by $g^*(y)$. For an orthonormal basis $\varepsilon = \{f_1,\dots,f_n\}$ of $(\mathscr{L}_F(y), g^*(y))$, there exists uniquely a basis $\{b_1, b_2,\dots,b_n\}$ of $\mathscr{L}_B(y)$ such that

$$\begin{aligned} \sigma_Q(y)(f_j, f_k) &= \sigma_Q(y)(b_j, b_k) = 0, \\ \sigma_Q(y)(f_j, b_k) &= \delta_{jk} \quad (j,k = 1,\dots,n). \end{aligned} \tag{A1}$$

When $\{f_1, f_2,\dots,f_n, b_1, b_2,\dots,b_n\}$ satisfies (A1), we have a symplectic basis of $(T_yT^*Q, \sigma_Q(y))$. We introduce Euclidean inner product $\hat{g}(y)$ into T_yT^*Q with respect to which $\{f_1,\dots,f_n, b_1,\dots,b_n\}$ becomes an orthonormal basis. Through the map

$$\begin{aligned} f_j &\to e_j = {}^t(0,\dots,0,\overset{j}{1},0,\dots,0), \\ b_j &\to ie_j \quad (j = 1,\dots,n), \end{aligned} \tag{A2}$$

we identify the vector space T_yT^*Q with $\mathbb{C}^n$. We define a Hermitain inner product on T_yT^*Q by $h(y) = \hat{g}(y) + i\sigma_Q(y)$. For $v \in \Lambda(T_yT^*Q)$, there exists a unitary matrix $W \in U(n)$ such that $v = W \cdot \mathscr{L}_B(y)$. Then $U(n)$ acts on $\Lambda(T_yT^*Q)$ transitively and the isotropy subgroup is $O(n)$. We put $\mathrm{Det}^2(v) = (\det W)^2 \in U(1)$. Note that $\mathrm{Det}^2(v)$ is independent of the choice of ε. Then, Det^2 turns out to be a mapping

$$\mathrm{Det}^2 : \Lambda(T^*Q) \to U(1), \tag{A3}$$

where $\pi_\Lambda : \Lambda(T^*Q) \to T^*Q$ is a Lagrangian Grassmannian bundle defined by

$$\Lambda(T^*Q) = \cup_y \Lambda(T_yT^*Q)$$

and $\pi_\Lambda(v) = y$, for $v \in \Lambda(T_yT^*Q)$. Consider a Lagrangian submanifold L of (T^*Q, σ_Q). For every $l \in L$, T_lL is an element of $\Lambda(T_lT^*Q)$. We define a mapping $\lambda: L \to \Lambda(T^*Q)$ by

$$\lambda(l) = T_lL \in \Lambda(T_lT^*Q), \quad l \in L. \tag{A4}$$

Then we have an element of $H^1(L;\mathbb{Z})$ such that

$$\mu_L = [(\mathrm{Det}^2 \circ \lambda)^*(dz/2\pi iz)], \tag{A5}$$

where $[dz/2\pi iz] \in H^1(U(1);\mathbb{Z})$, $z \in \mathbb{C}$, $|z| = 1$. This μ_L is referred to as the Maslov class of L. We note μ_L is independent of the choice of g and $\mathscr{L}_B$ (see Ref. 2).

APPENDIX B: PROOF OF PROPOSITION 5.3

A direct calculation yields

$$\begin{aligned} \det(G_p + iG_x) &= -(1/32|x|^2)\{(H_1 + H_3)(Z_1 + Z_3) \\ &\quad + (H_2 + H_4)(Z_2 + Z_4)\} \\ &\quad \times \{(H_1 - H_3)(Z_1 - Z_3) \\ &\quad + (H_2 - H_4)(Z_2 - Z_4)\}, \end{aligned} \tag{B1}$$

where $H_k = p_k - i8Hx_k$, $Z_k = x_k - ip_k$ $(k = 1,2,3,4)$. We set $\zeta_j = \alpha_j + i\beta_j$, $\alpha_j, \beta_j \in \mathbb{R}$ $(j = 1,2,3,4)$. Putting $8H = -8E$ $(= -\lambda^2)$, we get

$$\begin{aligned} H_1 + H_3 &= \tfrac{1}{2}(\beta_1 - \beta_3) + (i\lambda/2)(\alpha_1 + \alpha_3), \\ H_2 + H_4 &= -\tfrac{1}{2}(\alpha_1 - \alpha_3) + (i\lambda/2)(\beta_1 + \beta_3), \\ H_2 - H_4 &= -\tfrac{1}{2}(\alpha_2 - \alpha_4) + (i\lambda/2)(\beta_2 + \beta_4), \\ Z_1 + Z_3 &= (1/2\lambda)(\alpha_1 + \alpha_3) - (i/2)(\beta_1 - \beta_3), \\ Z_1 - Z_3 &= (1/2\lambda)(\alpha_2 + \alpha_4) - (i/2)(\beta_2 - \beta_4), \\ Z_2 + Z_4 &= (1/2\lambda)(\beta_1 + \beta_3) + (i/2)(\alpha_1 - \alpha_3), \\ Z_2 - Z_4 &= (1/2\lambda)(\beta_2 + \beta_4) + (i/2)(\alpha_2 - \alpha_4). \end{aligned} \tag{B2}$$

Substituting (B2) into (B1), we have

$$\det(G_p + iG_x)|\hat{L}(E,\bar{J}_m,\bar{D}_m)$$
$$= (1/32|x|^2)[\{(\alpha_1\alpha_3+\beta_1\beta_3)(\alpha_2\alpha_4+\beta_2\beta_4) - ((1+\lambda^2)^2/4\lambda^2)(\alpha_3\beta_1-\alpha_1\beta_3)(\alpha_4\beta_2-\alpha_2\beta_4)\} - (i(1+\lambda^2)/2\lambda) \times\{(\alpha_1\alpha_3+\beta_1\beta_3)(\alpha_4\beta_2-\alpha_2\beta_4) + (\alpha_2\alpha_4+\beta_2\beta_4)(\alpha_3\beta_1-\alpha_1\beta_3)\}].$$

Since $1/32|x|^2>0$, $\arg\det(G_p+iG_x)=\arg(\mathrm{Re}+i\,\mathrm{Im})$, where

$$\mathrm{Re} = (\alpha_1\alpha_3+\beta_1\beta_3)(\alpha_2\alpha_4+\beta_2\beta_4) - ((1+\lambda^2)^2/4\lambda^2)(\alpha_3\beta_1-\alpha_1\beta_3)(\alpha_4\beta_2-\alpha_2\beta_4),$$

$$\mathrm{Im} = -((1+\lambda^2)/2\lambda)\{(\alpha_1\alpha_3+\beta_1\beta_3)(\alpha_4\beta_2-\alpha_2\beta_4) + (\alpha_2\alpha_4+\beta_2\beta_4)(\alpha_3\beta_1-\alpha_1\beta_3)\}.$$

Using the global parametrization of $\hat{L}(E,\bar{J}_m,\bar{D}_m)$ given in (5.6), we get the following: on c_1 and c_2,

$$\mathrm{Re} = \hat{\Lambda}\cos t,\quad \mathrm{Im} = -((1+\lambda^2)/2\lambda)\hat{\Lambda}\sin t,$$

where $\hat{\Lambda} = (\pi_{j=1}^4 \Lambda_j)^{1/2}$, and on c_3,

$$\mathrm{Re} = \hat{\Lambda}\cos t,\quad \mathrm{Im} = ((1+\lambda^2)/2\lambda)\hat{\Lambda}\sin t.$$

Thus we have

$$\int_{c_j} d\arg\det(G_p+iG_x) = -2\pi \quad (j=1,2),$$
$$\int_{c_3} d\arg\det(G_p+iG_x) = 2\pi, \tag{B3}$$

which proves Proposition 5.3.

[1] V. P. Maslov, *Theorie des Perturbations et Method Asymptotic* (Dunod, Paris, 1972).
[2] V. Guillemin and S. Sternberg, *Geometric Asymptotics*, Mathematical Surveys No. 14 (Am. Math. Soc., Providence, RI, 1977).
[3] J. Leray, *Analyse Lagrangienne et Mechanique Quantique*, Seminaire du College de France 1976–1977, R. C. P. 25 (Strasbourg, 1978).
[4] A. Weinstein, Symp. Math. **24**, 25 (1974).
[5] K. Ii, Tohoku Math. J. **30**, 517 (1978).
[6] A. Yoshioka, Tokyo J. Math. **9**, 415 (1986).
[7] W. Greub and H. R. Petry, J. Math. Phys. **16**, 1347 (1975).
[8] N. Woodhouse, Geometric Quantization (Clarendon, Oxford, 1980).
[9] A. Yoshioka, "The quasi-classical calculation of eigenvalues for the Bochner-Laplacian on a line bundle," in *Geometry of Manifolds, Perspectives in Mathematics*, Vol. 8, edited by K. Shiohama (Academic, Boston, 1989).
[10] T. Iwai and Y. Uwano, J. Math. Phys. **27**, 1523 (1986).
[11] T. Iwai and Y. Uwano, J. Phys. A: Math. Gen. **21**, 4083 (1988).
[12] V. I. Arnold, *Mathematical Methods of Classical Mechanics, Springer Graduate Texts in Mathematics*, Vol. 60 (Springer, New York, 1978).

Proof of the noninteraction theorem for a system of *N* relativistic particles in direct interaction

Fu-Bin Li
Department of Physics, Huaibei Coal Teacher's College, Huaibei, Anhui, People's Republic of China

(Received 13 September 1989; accepted for publication 13 December 1989)

This paper proves a noninteraction theorem for a system of N relativistic particles in direct interaction, and each particle has internal Grassmann degrees of freedom for describing spin.

I. INTRODUCTION

The noninteraction theorem of Currie, Jordan, and Sudarshan[1] proves the incompatibility of direct interaction with relativistic symmetry. Relativistic symmetry comprises two distinct requirements. The first is that the dynamics admits the Poincaré group as a symmetry. Further, it is also necessary that the particle world lines transform as is expected of them in special relativity. This last requirement is called the world line condition (WLC) and is an essential ingredient in the proof of the theorem.

There are several versions of this theorem in the literature. The first, due to Currie, Jordan, and Sudarshan,[1] was proved for a system of two-point particles. Then Cannon and Jordan[2] proved the result for three particles and Leutwyler[3] extended the proof to the general case. These proofs were set in the framework of Hamiltonian mechanics. Subsequently, it was possible to prove the theorem in a purely Lagrangian framework.[4–7] This approach is considerably more economical than the earlier ones and does not need to assume that the Lagrangian is nonsingular.

The above results were obtained for structureless point particles. In this paper, we allow for internal structure and show again that relativistic symmetry rules out direct particle interactions. The internal structure can be thought of as describing spin in a classical framework. The internal coordinates can be either commuting or Grassmann. In either case, our result applies. For definiteness, we treat the more novel situation where the internal coordinates take values in a Grassmann algebra. The case of bosonic internal variables is straightforward.

At the classical level, anticommuting variables correspond to Fermi degrees of freedom. In the last decade, Casalbuoni and others have developed the subject of pseudoclassical mechanics.[6,7] This provides a formal description of Fermi systems, which is purely classical. Upon quantization, these systems yield quantum Fermi systems. These developments provide a natural way to describe spin in a classical framework. (Other descriptions using bosonic internal coordinates are also possible.[8] For a review, see Ref. 1.)

There is another approach to the no-interaction theorem that is different from the original approach of Sudarshan and co-workers.[9] This approach uses the assumption that the action must be invariant under independent reparametrization of each subsystem. This approach has been extended to include Grassmann variables giving the dynamical independence of the subsystems.[10,11]

In this paper, we proceed along the lines of Sudarshan's approach. We first review pseudoclassical mechanics from a geometrical point of view in Sec. II. This involves working on supermanifolds. An excellent introduction to this subject is given in Ref. 11, whose notations we largely use. In Sec. III, we formulate the problem we are interested in and derive a WLC suitable for Grassmann variables. In Sec. IV, we prove the main results of this paper. Section V is a brief concluding discussion.

II. DYNAMICS ON SUPERMANIFOLDS

Let ϵ_i, $i = 1,...,L$, be anticommuting generators

$$\epsilon_i\epsilon_j + \epsilon_j\epsilon_i = 0\,, \tag{1}$$

of an algebra Λ_L. Elements of Λ_L are objects constructed from ϵ_i by multiplication and addition. A general element f of Λ_L has the form

$$f = \sum_{\nu=0}^{L} f_\nu \epsilon_\nu\,, \tag{2}$$

where $f_\nu = f_{i_1\cdots i_\nu}$ are ordinary real numbers and $\epsilon_\nu = \epsilon_{i_1}\cdots\epsilon_{i_\nu}$ are products of the basic generators. Obviously, since $\epsilon_i\epsilon_i = 0$ from (1), $\nu = L$ is the highest power that can appear in (2). Also $f_{i_1\cdots i_\nu}$ is completely antisymmetric in its indices. The real dimension of Λ_L is 2^L since this is the number of distinct ϵ's. A general element of Λ_L can be decomposed into

$$f = f_{\text{odd}} + f_{\text{even}} = \sum_{\nu=\text{odd}} f_\nu\epsilon_\nu + \sum_{\nu=\text{even}} f_\nu\epsilon_\nu\,, \tag{3}$$

its odd and even parts. Those elements that have vanishing odd or even parts will be called pure. In this work we need to only deal with pure objects.

B. Superspace

Consider a space $R_c^m \otimes R_a^n$ spanned by m even elements q^i and n odd elements s^α of Λ_L. Here, (q^i, s^α) are global coordinates on $R_c^m \otimes R_a^n$, whose real dimension is $(m+n)2^{L-1}$. Also, $R_c^m \otimes R_a^n$ is the model for the construction of supermanifolds, just as R^n is the model for ordinary manifolds. We will only need to work with the simple supermanifold $R_c^m \otimes R_a^n$, which admits a single global chart. For the more general definition, consult Ref. 7. In the future we use the symbol $Q = R_c^m \otimes R_a^n$ to save writing. We also set $L = \infty$ and drop the subscript on Λ.

C. Geometry on superspace

A function $f \in \mathscr{F}(Q)$ on Q is a map

$$f\colon Q \to \Lambda\,. \tag{4}$$

 0022-2488/90/061395-05$03.00

We suppose that all the functions we deal with are superanalytic.[12] This makes matters more restrictive and interesting, just like complex analysis is more restrictive than the analysis on R^2.

A vector field $X \in \mathscr{X}(Q)$ is a map

$$X:\ \mathscr{F}(Q) \to \mathscr{F}(Q)\,, \tag{5}$$

satisfying

$$X(f_\alpha) = X(f)_\alpha\,, \tag{6}$$

$$X(f+g) = X(f) + X(g)\,, \tag{7}$$

$$X(fg) = X(f)g + (-1)^{Xf} f X(g)\,, \tag{8}$$

where $f,g \in \mathscr{F}(Q)$, $\alpha \in \Lambda$, and the notation of Ref. 11 has been used. Obviously the coordinate functions (q^i, s^α) are elements of $\mathscr{F}(Q)$ and we can characterize any vector field by its action on these

$$X(q^i) = X^i,\quad X(s^\alpha) = X^\alpha\,. \tag{9}$$

Using the basis $(\partial/\partial q^i, \partial/\partial s^\alpha)$ of vector fields, any vector field X can be written

$$X = X^i \frac{\partial}{\partial q^i} + X^\alpha \frac{\partial}{\partial s^\alpha}\,. \tag{10}$$

A one-form $\alpha \in \mathscr{X}^*(Q)$ is a map

$$\alpha:\ \mathscr{X}(Q) \to \mathscr{F}(Q)\,, \tag{11}$$

satisfying

$$i_{x+y}\alpha = i_x \alpha + i_y \alpha\,, \tag{12}$$

where $i_x \alpha$ is the image of X under the map (11). We introduce the basis of one-forms aq^i, ds^α, and write

$$\alpha = dq^i \alpha_i + ds^\alpha \alpha_a\,. \tag{13}$$

Similarly from Ref. 11 one can define higher-order forms, the exterior derivative d and the Lie derivative L_x. While manipulating these objects it is important to keep track of the order, or else additional minus signs appear.

We now consider a simple dynamical system to illustrate dynamics on supermanifold Q of dimension (1,2) with global coordinates (q, s^1, s^2). Here, q is an even element of Λ and S^α are odd elements. Suppose that a Lagrangian function $\mathscr{L}$ is given on TQ, the tangent bundle over Q. Define

$$\theta_{\mathscr{L}} = dq \frac{\partial \mathscr{L}}{\partial q^i} + ds^\alpha \frac{\partial \mathscr{L}}{\partial \dot{s}^\alpha} \tag{14}$$

and

$$-\omega_{\mathscr{L}} = d\theta_{\mathscr{L}}\,. \tag{15}$$

The dynamical vector field Δ is a second-order vector field

$$\Delta = \dot{q} \frac{\partial}{\partial q} + \dot{s}^\alpha \frac{\partial}{\partial \dot{s}^\alpha} = a \frac{\partial}{\partial q} + b^\alpha \frac{\partial}{\partial s^\alpha} \tag{16}$$

$(\alpha = 1,2)$ on TQ, satisfying the Euler–Lagrange equations

$$L_s \theta_{\mathscr{L}} = d\mathscr{L}\,. \tag{17}$$

For instance if

$$\mathscr{L} = \tfrac{1}{2} m \dot{q}^2 + \dot{s}^1 \dot{s}^2 + u(q) s^1 s^2\,, \tag{18}$$

then

$$L_\Delta \theta_{\mathscr{L}} = m\, dq\, \dot{q} + ds'\, \dot{s}^2 - ds^2 \dot{s}^1 \tag{19}$$

and

$$L_\Delta \theta_{\mathscr{L}} = m\, d\dot{q}\dot{q} + m\, dq\, a + d\dot{s}^1 \dot{s}^2 + ds^1\, b^2 - d\dot{s}^2\, \dot{s}^1 - ds^2\, b^1\,, \tag{20}$$

where a, b^α are the coefficients appearing in expression (16). The Euler–Lagrange equations (17) yield

$$ma = \frac{du}{dq} - s^1 s^2\,, \tag{21}$$

$$b^\alpha = u(q) s^\alpha,\quad \alpha = 1,2\,; \tag{22}$$

that is,

$$m\ddot{q} = \frac{du}{dq} s^1 s^2,\quad \bar{s}^\alpha = u(q) s^\alpha\,. \tag{23}$$

II. THE *N*-PARTICLE PROBLEM AND THE WORLD LINE CONDITION

We now address the problem of N relativistic particles. Each particle is described by special spatial coordinates q_a^i ($a = 1,\dots,N$ is a particle label, $i = 1,2,3$ is a Cartesian vector index, we add q_a^α to deal with time components) and internal Grassmann coordinates s_a^α, $\alpha = 1,\dots,1$. We collectively denote (q_a^i, s_a^α) by the single symbol x_a^μ, where χ_a^μ are global coordinates on the supermanifold Q, which is the configuration space of the system. Thus Q has dimension $(3N, lN)$. Under Lorentz transformations, the q transform as the spatial components of a four-vector and s transform according to some representation of the Lorentz group. Let us denote the corresponding generators by $(\Sigma_{\mu\nu})^\alpha_\beta$. These are not dynamical objects, but just a collection of numbers. The physical quantities describing spin would be even objects like $s_\alpha (\Sigma_{\mu\nu})^\alpha_\beta s^\beta$. Note that we use the summation convention for the indices α and i, but not for the particle labels a,b,c. We work in the instant form of dynamics (like Refs. 1–4) and our use of q as coordinates is, therefore, appropriate. The evolution parameter is t, the physical time of some inertial observer.

Let us suppose that a Lagrangian function $\mathscr{L}(x,\dot{x})$ is given on TQ.

From $\mathscr{L}$ we construct the one-form

$$\theta_{\mathscr{L}} = \sum_a dx_a^\mu \frac{\partial \mathscr{L}}{\partial \dot{x}_a^\mu} \tag{24}$$

and the two-form

$$-\omega_{\mathscr{L}} = d\theta_{\mathscr{L}}\,. \tag{25}$$

Explicitly written out, $\omega_{\mathscr{L}}$ has the form

$$-\omega_{\mathscr{L}} = \sum_{a,b} \left[dx_a^\mu\, dx_b^\nu \frac{\partial^2 \mathscr{L}}{\partial x_b^\nu\, \partial \dot{x}_a^\mu} + dx_a^\mu \Lambda\, d\dot{x}_b^\nu \frac{\partial^2 \mathscr{L}}{\partial \dot{x}_b^\nu\, \partial \dot{x}_a^\mu} \right]. \tag{26}$$

Note that, by construction,

$$i_{\partial/\partial \dot{x}_a^\mu} \cdot i_{\partial/\partial \dot{x}_b^\nu} \omega_{\mathscr{L}} = 0\,, \tag{27}$$

which fact we will use in the next section to prove the theorem forbidding interaction. Also

$$i_{\partial/\partial x_a^\mu} \cdot i_{\partial/\partial \dot{x}_b^\nu} \omega_{\mathscr{L}} = -i_{\partial/\partial \dot{x}_a^\mu} \cdot i_{\partial/\partial x_b^\nu} \omega_{\mathscr{L}_x} = \frac{\partial^2 \mathscr{L}}{\partial \dot{x}_b^\nu\, \partial \dot{x}_a^\mu}\,, \tag{28}$$

as can be easily seen if one is careful with signs.

We suppose that dynamics is an even second-order vector field on TQ (otherwise even objects will become odd under time evolution). With no loss of generality we write

$$\Delta = \sum_a \Delta a , \qquad (29)$$

$$\Delta_a = \dot{x}^{\mu}_a \frac{\partial}{\partial x^{\mu}_a} + A^{\mu}_a \frac{\partial}{\partial \dot{x}^{\mu}_a} = \dot{q}^i_a \frac{\partial}{\partial q^i_a} + \dot{s}^{\alpha}_a \frac{\partial}{\partial s^{\alpha}_a} + A^{i}_a \frac{\partial}{\partial \dot{q}^i_a} + A^{\alpha}_a \frac{\partial}{\partial \dot{s}^{\alpha}_a} . \qquad (30)$$

The accelerations A^{μ}_a have to satisfy the Euler-Lagrange equations (17), which can also be written

$$i_a \omega_{\mathscr{L}} = dE_{\mathscr{L}} , \qquad (31)$$

where

$$E_{\mathscr{L}} = i_{\Delta} \theta_{\mathscr{L}} - \mathscr{L} , \qquad (32)$$

is the energy function on TQ.

Relativistic symmetry demands the existence of a ten Poincaré vector field on TQ: P_i, J_i, K_i, Δ, which reflect the Lie algebra structure of the Poincaré group in their Lie bracket relations. For example,

$$[K_j,\Delta] = P_j . \qquad (33)$$

These ten vector fields must generate canonical transformation on TQ and so

$$L_{P_i}\omega_{\mathscr{L}} = L_{J_i}\omega_{\mathscr{L}} = L_{K_i}\omega_{\mathscr{L}} = 0 , \qquad (34)$$

$$L_{\Delta}\omega_{\mathscr{L}} = 0 . \qquad (35)$$

The last is of course immediate from (31). Since we would like the gradings of the basic variables to be preserved under Lorentz transformations, these ten vector fields must be even.

As mentioned before, we work in the instant form of dynamics, where the evolution parameter is the physical time of an inertial observer. The ten Poincaré group generators can be divided into two classes according to whether or not they preserve the $t=$ const surface. The generators J_i, P_i of spatial rotations and translations are called "simple" or kinematical generators,[14] while the boosts K_i and the time translation Δ are called "complicated" or "dynamical." The kinematical generators retain their free-particle forms even in the presence of interaction[14]

$$J_i = \sum_a \epsilon_{ijk} \left[q^j_a \frac{\partial}{\partial q^k_a} + \dot{q}^j_a \frac{\partial}{\partial \dot{q}^k_a} + \left(\sum_{jk}\right)^{\alpha}_{\beta} \left(s^{\beta}_a \frac{\partial}{\partial s^{\alpha}_a} + \dot{s}^{\beta}_{\alpha} \frac{\partial}{\partial \dot{s}^{\alpha}_a} \right) \right] , \qquad (36)$$

$$P_i = \sum_a \frac{\partial}{\partial q^i_a} . \qquad (37)$$

The remaining generators cannot have their free-particle forms if there is interaction. For instance, if K_j had its free-particle form, (33) would imply that Δ too had free-particle form and there would be no interaction. The forms of K_j will be determined below by using the WLC.

We now formulate a WLC that expresses the objectivity of the world lines $(q^i_a(t), s^{\alpha}_a(t))$ of the particles in superspace. Consider particle a located at q^i_a at time t with internal coordinates s^{α}_a. Under an infinitesimal boost with Lorentz parameters ϵ^{0i}, the quantities (t, q^i_a, s^{α}_a) transform as follows:

$$t \to t' = t + \delta t = t + \epsilon^0_i q^i_a , \qquad (38)$$

$$q^i_a \to q'^i_a = q^i_a + \delta q^i_a = q^i_a + \epsilon^i_{0t} , \qquad (39)$$

$$s^{\alpha}_a \to s'^{\alpha}_a = s^{\alpha}_a + \delta s^{\alpha}_a = s^{\alpha}_a + \left(\epsilon^{0j} \sum_{0j}\right)^{\alpha}_{\beta} s^{\beta}_a . \qquad (40)$$

Thus the world line $(q^i_a(t), s^{\alpha}_a(t))$ in superspace transforms to $(q'^i_a(t'), s'^{\alpha}_a(t'))$. To first order in the infinitesimal parameter ϵ^{0i} we can write

$$q'^i_a(t') = q'^i_a(t') + \dot{q}^i_a \delta t , \qquad (41)$$

$$s'^{\alpha}_a(t') = s'^{\alpha}_a + \dot{s}^{\alpha}_a \delta t , \qquad (42)$$

or at $t=0$,

$$\delta q^i_a(t) = \dot{q}^i_a \epsilon^0_j q^j_a , \qquad (43)$$

$$\delta s^{\alpha}_a(t) = \left(\epsilon^{0j} \sum_{\Delta j}\right)^{\alpha}_{\beta} s^{\beta}_a + \dot{s}^{\beta}_a + \dot{s}^{\alpha}_a \epsilon^0_j q^j_a . \qquad (44)$$

"Peeling off" the ϵ^{0j}, we can write the WLC as a condition on the boost generators:

$$L_{K_j} q^i_a = \dot{q}^i_a q^j_a , \qquad (45)$$

$$L_{K_j} S^{\alpha}_a = \left(\sum_{0j}\right)^{\alpha}_{\beta} s^{\beta}_a + \dot{s}^{\alpha}_a q^j_a . \qquad (46)$$

These equations express the objectivity of the world lines under Lorentz transformations. They fix the "horizontal" parts of the boost generators. Let us now apply L_{Δ} to (45) and (46) and use the Lie bracket relations (33) and (37). This yields

$$I_{K_j} \dot{q}^i_a = (\dot{q}^i_a \dot{q}^j_a - \delta_{ij}) + A^i_a q^j_a , \qquad (47)$$

$$L_{K_j} \dot{s}^{\alpha}_a = \left(\sum_{0j}\right)^{\alpha}_{\beta} \dot{s}^{\beta}_{\alpha} + A^{\alpha}_a q^j_a + \dot{s}^{\alpha}_a \dot{q}^j_a . \qquad (48)$$

This fixes the "vertical" part of K_j as well in terms of the dynamics Δ. The explicit form of K_j thus derived is

$$K_j = \sum_a q^j_a \Delta_a + \left(\sum_{0j}\right)^{\alpha}_{\beta} \left(s^{\beta}_a \frac{\partial}{\partial s^{\alpha}_a} + \dot{s}^{\alpha}_a \frac{\partial}{\partial \dot{s}^{\alpha}_a} \right) + (\dot{q}^j_a \dot{q}^i_a - \delta_{ij}) \frac{\partial}{\partial q^i_a} + \dot{s}^{\alpha}_a \dot{q}^j_a \frac{\partial}{\partial \dot{s}^{\alpha}_a} . \qquad (49)$$

This form of the boost generators, derived from the WLC, will be made use of in the next section to prove the no-interaction theorem.

IV. PROOF OF THE MAIN THEOREM

We now set about proving the main result of this paper. The proof is in three steps and entirely parallel to that of Ref. 4.

Step 1: Apply L_{K_j} to (27) and use (34). This gives

$$i_{[K_j,\partial/\partial x^{\mu}_a]} i_{\partial/\partial \dot{x}^{\nu}_b} \omega_{\mathscr{L}} + i_{\partial/\partial \dot{x}^{\mu}_a} i_{[K_j,\partial/\partial \dot{x}^{\nu}_b]} \omega_{\mathscr{L}} = 0 . \qquad (50)$$

Note that no extra minus signs appear here since K_j is an even vector field. Using the form (49) for K_j and remembering (27),

$$- q^j_a i_{\partial/\partial x^{\mu}_a} i_{\partial/\partial \dot{x}^{\nu}_b} \omega_{\mathscr{L}} - q^j_b i_{\partial/\partial \dot{x}^{\mu}_a} i_{\partial/\partial \dot{x}^{\nu}_b} \omega_{\mathscr{L}} = 0 . \qquad (51)$$

Or, by (28),

$$(q^j_a - q^j_b) i_{\partial/\partial \dot{x}^\mu_a} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} = 0. \tag{52}$$

For $a \neq b$ this implies

$$i_{\partial/\partial \dot{x}^\mu_a} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} = \frac{\partial^2 \mathscr{L}}{\partial \dot{x}^\mu_a \, \partial \dot{x}^\nu_b} = 0, \tag{53}$$

which means that $\mathscr{L}$ is of the form

$$\mathscr{L}(x,\dot{x}) = \sum_a \mathscr{L}^a(x,\dot{x}_a), \tag{54}$$

which is completely separated in the velocities.

Step 2: Let us apply L_Δ to (53) where $a \neq b$. This gives

$$i_{[\Delta,\partial/\partial \dot{x}^\mu_a]} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} + i_{\partial/\partial \dot{x}^\mu_a} i_{[\Delta,\partial/\partial \dot{x}^\nu_b]} \omega_{\mathscr{L}} = 0. \tag{55}$$

The second term drops out by virtue of (27) and the form (29), (30) of Δ. The first term then yields

$$i_{\partial/\partial x^\mu_a} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} = -\frac{\partial}{\partial \dot{x}^\mu_a} A^a_b i_{\partial/\partial x^a_b} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} = 0. \tag{56}$$

Next, apply L_{K_j} to (53). This gives

$$i_{[K_j,\partial/\partial \dot{x}^\mu_a]} \omega_{\mathscr{L}} + i_{\partial/\partial \dot{x}^\mu_a} i_{[K_j,\partial/\partial \dot{x}^\nu_b]} \omega_{\mathscr{L}} = 0. \tag{57}$$

The second term again drops out and the first yields by use of (56)

$$q^j_a i_{\partial/\partial x^\mu_a} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} = q^j_b i_{\partial/\partial x^\mu_a} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}}. \tag{58}$$

Since $a \neq b$ this implies

$$i_{\partial/\partial x^\mu_a} i_{\partial/\partial \dot{x}^\nu_b} \omega_{\mathscr{L}} = 0. \tag{59}$$

Using the velocities separated from (54) of the Lagrangian we can write

$$\theta_{\mathscr{L}} = \sum_b dx^\nu_b \frac{\partial \mathscr{L}^b}{\partial \dot{x}^\nu_b} \tag{60}$$

and

$$\omega_{\mathscr{L}} = \sum_{ab} dx^\nu_b \wedge dx^\mu_a \frac{\partial^2 \mathscr{L}^b}{\partial \dot{x}^\nu_b \, \partial x^\mu_a} + \sum_b dx^\nu_b \wedge d\dot{x}^\mu_b \frac{\partial^2 \mathscr{L}^b}{\partial x^\mu_b \, \partial \dot{x}^\mu_a}. \tag{61}$$

The use of (59) then shows that

$$\frac{\partial^2 \mathscr{L}^b}{\partial \dot{x}^\nu_b \, \partial x^\mu_a} = (-1)^{\mu\nu} \frac{\partial^2 \mathscr{L}^a}{\partial x^\nu_b \, \partial \dot{x}^\mu_a}, \tag{62}$$

where $(-1)^{\mu\nu} = -1$ if both μ and ν are odd variables, and $(-1)^{\mu\nu} = +1$ otherwise. Equation (62) implies that the dependence of $\mathscr{L}^a$ on x_b can be at most linear in the velocities

$$\mathscr{L}^a(x,\dot{x}_a) = \mathscr{L}^a(x_a,\dot{x}_a) + \dot{x}^\mu_a f_{a\mu}(x) - V_a(x), \tag{63}$$

where an abuse of notation has been made by giving the same symbol to $\mathscr{L}^a(x,\dot{x}_a)$ and $\mathscr{L}^a(x_a,\dot{x}_a)$. Using (62) once more we find that $f_{a\mu}$ satisfies

$$\frac{\partial}{\partial x^\mu_a} f_{b\nu} = (-1)^{\mu\nu} \frac{\partial}{\partial x^\nu_b} f_{a\mu}, \tag{64}$$

for $a \neq b$. We eliminate the linear terms in the velocities in (63) by the following procedure:

$$\alpha = \sum_a dx^\mu_a f_{a\mu}, \tag{65}$$

$$d\alpha = \sum_{ab} dx^\mu_a \wedge dx^\nu_b \frac{\partial f_{au}}{\partial x^\nu_b} = \sum_a dx^\mu_a \wedge dx^\nu_a \frac{\partial f_{a\mu}}{\partial x^\nu_a}, \tag{66}$$

by use of (64). Let us define

$$F_{a\mu\nu} = \frac{\partial f_{a\mu}}{\partial x^\nu_a} - (-1)^{\mu\nu} \frac{\partial f_{a\nu}}{\partial x^\mu_a}. \tag{67}$$

Then

$$d\alpha = \sum_a dx^\mu_a \wedge dx^\nu_a F_{a\mu\nu}. \tag{68}$$

By Poincaré lemma $dd\alpha = 0$ and so

$$dd\alpha = \sum_{ab} dx^\mu_a \wedge dx^\nu_a \wedge dx^\sigma_b \frac{\partial F_{a\mu\nu}}{\partial x^\sigma_b} = 0. \tag{69}$$

Since only one term with ax_b, $b \neq a$ appears here, its coefficient must be zero:

$$\frac{\partial F_{a\mu\nu}}{\partial x^\sigma_b} = 0. \tag{70}$$

This means that $F_{a\mu\nu}(x_a)$ depends only on x_a and (69) reads

$$\sum_a ax^\mu_a \wedge dx^\nu_a \wedge dx^\sigma_a \frac{\partial F_{a\mu\nu}}{\partial x^\sigma_a}. \tag{71}$$

The only way this can be true is that each term in the sum vanishes, or the forms

$$F_a = dx^\mu_a \wedge dx^\nu_a F_{a\mu\nu}(x^a) \tag{72}$$

are closed

$$dF_a = 0, \tag{73}$$

and hence exact

$$F_a = dA_a(x_a). \tag{74}$$

Now, letting $\beta = \alpha - \Sigma_a A_a$, we find from (68) that β is closed

$$d\beta = 0, \tag{75}$$

and hence exact

$$\beta = dF. \tag{76}$$

Thus

$$\alpha = dF + \sum_a A_a(x_a). \tag{77}$$

Thus (63) assumes the form

$$\mathscr{L}^a(x,\dot{x}_a) = \mathscr{L}^a(x_a,\dot{x}_a) + \dot{x}^\mu_a \frac{\partial F}{\partial x^\mu_a} + \dot{x}^\mu_a A_{a\mu} - V(x). \tag{78}$$

The second term on the right is a total time derivative and can therefore be dropped. The third term is in separated form and so can be absorbed in the first. Thus at this stage, the Lagrangian has the form

$$\mathscr{L} = \sum_a \mathscr{L}^a(x_a,\dot{x}_a) - V(x). \tag{79}$$

The symplectic two-form $\omega_{\mathscr{L}}$ is now completely separated

$$\omega_{\mathscr{L}} = \sum_a \omega^a_{\mathscr{L}}. \tag{80}$$

$$\omega^a_{\mathscr{L}} = ax^{\mu}_a \wedge d\left[\frac{\partial \mathscr{L}^a}{\partial \dot{x}^{\mu}_a}\right]. \tag{81}$$

This completes the argument of step 2.

Step 3: Apply L_Δ to (59). This gives

$$i_{[\Delta,\partial/\partial x^{\mu}_b]} i_{\partial/\partial x^{\nu}_a} \omega_{\mathscr{L}} + i_{\partial/\partial x^{\mu}_a} i_{[\Delta,\partial/\partial x^{\nu}_b]} \omega_{\mathscr{L}} = 0. \tag{82}$$

From the form (30) of Δ,

$$\left[\Delta,\frac{\partial}{\partial x^{\mu}_a}\right] = -\sum_0 \frac{dA^{\sigma}_c}{\partial x^{\mu}_a} \frac{\partial}{\partial \dot{x}^{\sigma}_c}, \tag{83}$$

when this is put into (82) all the terms with $c=b$ drop out by (53) and we find

$$\frac{\partial A^{\sigma}_b}{\partial x^{\mu}_a} i_{\partial/\partial \dot{x}^{\sigma}_b} i_{\partial/\partial x^{\nu}_b} \omega^b_{\mathscr{L}} = \frac{\partial A^{\sigma}_b}{\partial x^{\nu}_b} i_{\partial/\partial \dot{x}^{\sigma}_a} i_{\partial/\partial x^{\mu}_a} \omega^a_{\mathscr{L}}. \tag{84}$$

Next, apply L_{K_j} to (59). We find that

$$i_{[K_j,\partial/\partial x^{\mu}_a]} i_{\partial/\partial x^{\nu}_b} \omega_{\mathscr{L}} + i_{\partial/\partial x^{\mu}_a} i_{[K_j,\partial/\partial x^{\nu}_b]} \omega_{\mathscr{L}} = 0. \tag{85}$$

Using the form of K_j (49) in (85) we find

$$q^j_b \frac{\partial A^{\sigma}_b}{\partial x^{\mu}_a} i_{\partial/\partial \dot{x}^{\sigma}_b} i_{\partial/\partial x^{\nu}_b} \sum \omega^b_{\mathscr{L}} = q^j_a \frac{\partial A^{\sigma}_a}{\partial x^{\nu}_b} i_{\partial/\partial \dot{x}^{\sigma}_a} i_{\partial/\partial x^{\mu}_a} \omega^a_{\mathscr{L}}, \tag{86}$$

which by (84) is

$$q^j_b \frac{\partial A^{\sigma}_b}{\partial x^{\mu}_a} i_{\partial/\partial \dot{x}^{\sigma}_b} i_{\partial/\partial x^{\nu}_b} \omega^b_{\mathscr{L}} = q^j_a \frac{\partial A^{\sigma}_b}{\partial x^{\mu}_a} i_{\partial/\partial \dot{x}^{\sigma}_b} i_{\partial/\partial x^{\nu}_b} \omega^b_{\mathscr{L}}. \tag{87}$$

For $a \neq b$ we have

$$\frac{\partial A^{\sigma}_b}{\partial x^{\mu}_a} i_{\partial/\partial \dot{x}^{\sigma}_b} i_{\partial/\partial x^{\nu}_b} \omega^b_{\mathscr{L}} = 0. \tag{88}$$

Or since $\omega_{\mathscr{L}}$ is in separated form (80), (81),

$$\frac{\partial}{\partial x^{\mu}_b}\left[A^{\sigma}_a i_{\partial/\partial \dot{x}^{\sigma}_a} i_{\partial/\partial x^{\nu}_a} \omega^a_{\mathscr{L}}\right] = 0. \tag{89}$$

This is the main result of step 3.

With form (79) of the Lagrangian, the Euler–Lagrange equations (17) assume the form

$$\dot{x}^{\nu}_a \frac{\partial^2 \mathscr{L}^a}{\partial x^{\nu}_a \partial \dot{x}^{\mu}_a} + A^{\nu}_a \frac{\partial^2 \mathscr{L}^a}{\partial \dot{x}^{\nu}_a \partial \dot{x}^{\mu}_a} - \frac{\partial \mathscr{L}^a}{\partial x^{\mu}_a} = \frac{-\partial v}{\partial x^{\mu}_a}. \tag{90}$$

Clearly the lhs of (90) has no dependence on x_b by (79) and (89). If these Euler–Lagrange equations have solutions all over TQ, we must then have

$$\frac{\partial^2 V}{\partial x^{\nu}_b \partial x^{\mu}_a} = 0, \tag{91}$$

which implies that V has the separated form

$$V = \sum_a V^a(x_a). \tag{92}$$

This implies that the original Lagrangian separates

$$\mathscr{L}(x,\dot{x}) = \sum_a \mathscr{L}'^a(x_a,\dot{x}_a), \tag{93}$$

and so each particle is governed by its own dynamics and moves independently of the others. Consequently, there is no interaction and the theorem is proved.

V. CONCLUSIONS

We have presented a proof of the no-interaction theorem that applies even if the particles have internal structure. For definiteness, we have presented the case where the internal coordinates are odd Grassmann variables. But, in fact, the arguments go through trivially even in the case of commuting internal variables. The assumptions made were: (1) the existence of a (possibly singular) Lagrangian; (2) Poincaré invariance as expressed by the existence of ten Hamiltonian vector fields on *TQ*; (3) objectivity of the world lines; (4) the existence of a second-order dynamics all over *TQ*.

One might wonder about the physical content of the WLC for the Grassmann variables. As explained in Refs. 5 and 6, these do not possess direct experimental significance. We perform purely formal manipulations with the odd variables. It is only after quantization that pseudoclassical mechanics makes contact with experiment. Then one would require that the analog of the WLC helped in the quantum theory. The WLC derived here is a classical version of this requirement. This is why we impose the WLC on odd variables.

In all this we have nowhere assumed that the Lagrangian is nonsingular. This is unlike the proofs of Refs. 1–3 that do assume that dynamics are described by a nonsingular Lagrangian. This point is emphasized by Sudarshan and Mukunda.[15] The Lagrangian proof of the no-interaction theorem generalizes easily to accommodate Grassmann variables.[4–5] This is particularly important because the standard Lagrangians written for Grassmann variables are singular.[16]

[1]D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).
[2]J. T. Cannon and T. F. Jordan, J. Math. Phys. **5**, 299 (1964).
[3]H. Leutwyler, Nuovo Cimento **37**, 556 (1965).
[4]G. Marmo, N. Mukunda, and E. C. G. Sudarshan, Phys. Rev. D **30**, 2110 (1984).
[5]A. P. Balachandran, G. Marmo, and A. Stern, Nuovo Cimento A **69**, 195 (1982).
[6]R. Casalbuoni, Nuovo Cimento **33**, 115 (1976).
[7]F. A. Berezin and M. S. Marinov, Ann. Phys. (NY) **104**, 336 (1977).
[8]M. V. Atre and N. Mukunda, J. Math. Phys. **28**, 792 (1987).
[9]R. Giachetti and E. Sorace, Lett. Nuovo Cimento **26**, 1 (1979).
[10]A. Barducci, R. Giachetti, and E. Sorace, Lett. Math. Phys. **8**, 105 (1984).
[11]A. Barducci, R. Giachetti, J. Gomis, and E. Sorace, J. Phys. A **17**, 3277 (1984).
[12]B. De Witt, *Supermanifolds* (Cambridge U.P., Cambridge, 1984).
[13]A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[14]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[15]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics, A Modern Perspective* (Wiley, New York, 1976).
[16]A. P. Balachandran *et al.*, *Gauge Symmetries and Fibre Bundles, Application to Particle Dynamics* (Springer, Berlin, 1983).

On equations of type $u_{xt}=F(u,u_x)$ which describe pseudospherical surfaces

Mauro L. Rabelo and Keti Tenenblat
Institute de Ciencias Exatas, Departamento de Matematica, Universidade de Brasilia, C. P. 153101, 70910 Brasilia D. F., Brazil

(Received 28 March 1989; accepted for publication 3 January 1990)

The equations $u_{xt} = F(u,u_x)$, which describe η-pseudospherical surfaces, are characterized. In particular, when F does not depend on u_x, the sine–Gordon, sinh–Gordon, and Liouville equations are essentially obtained. Moreover, it is shown that an equation $u_{xt} = F(u)$ has a self-Bäclund transformation if and only if it describes an η-pseudospherical surface.

I. INTRODUCTION

In 1979, Sasaki[1] observed that a class of nonlinear differential equations, which can be solved by the inverse scattering method, was related to hyperbolic surfaces. Other results associating nonlinear equations with Riemannian manifolds of constant curvature were obtained in Refs. 2–8. Ablowitz, Beals, and Tenenblat[9] obtained solutions of the generalized wave equation and the generalized sine–Gordon equation using the inverse scattering method. These results were extended in Ref. 10. Solutions for these equations are orthogonal matrices with n-independent variables that correspond, respectively, to flat n-dimensional submanifolds of the unit sphere S^{2n-1} and to hyperbolic n-dimensional submanifolds of the Euclidean space $\Re^{2n-1}$.

In order to apply the inverse scattering method it is necessary to obtain a one-parameter linear problem associated to the nonlinear differential equation. In Ref. 4, Chern and Tenenblat began a systematic procedure to obtain such a linear problem. The notion of a differential equation, for a real function $u(x,t)$, which describes a pseudospherical surface (p.s.s.) was introduced. A generic solution of such an equation provides a metric defined on an open subset of $\Re^2$, whose Gaussian curvature is -1. We say that such an equation describes an η-p.s.s., where η is a parameter, if the length of the vector field $\partial/\partial x$ satisfies $|\partial/\partial x|^2 \geqslant \eta^2$ (see Sec. II for definitions). Under these conditions one obtains a one-parameter (η) linear problem (6) whose compatibility condition is the differential equation for $u(x,t)$. These compatibility conditions are the structure equations of a hyperbolic surface.

Equations related to the AKNS system[11] are included in this class of equations. Equations of the type $u_t = F(u,u_x,\ldots,\partial^k u/\partial x^k)$, which describe η-p.s.s., were studied in Ref. 4. Similar results were obtained in Ref. 5 for the equations $u_{tt} = F(u,u_x,u_{xx},u_t)$. Motivated mainly by the existence of important examples such as sine–Gordon, sinh–Gordon, and Liouville equations, we are interested in studying equations of the type

$$u_{xt} = F\left(u,u_x,\ldots,\frac{\partial^k u}{\partial x^k}\right). \tag{1}$$

In this paper, extending results obtained in Ref. 12, we characterize the equations $u_{xt} = F(u,u_x)$ that describe an η-p.s.s. for $\eta\in\mathscr{P} = \Re - S$, where S is a set of isolated points and F is a differentiable (C^∞) function independent of η. Equations of the form (1), for $k\geqslant 2$, will be considered in another paper.

We obtain the following results.

Theorem 1: Let F be a differentiable function defined on an open connected subset $U\subset\Re^2$. An equation

$$u_{xt} = F(u,u_x)$$

describes an η-p.s.s. for $\eta\in\mathscr{P}$ and F independent of η if and only if F satisfies one of the following:

(i) $F''(u) + \alpha F(u) = 0,$

where F is independent of u_x, $U = \Re^2$, $\mathscr{P} = \Re - \{0\}$, and α is a nonzero real constant;

(ii) $F = \nu e^{\delta u}\sqrt{\beta + \gamma u_x^2},$

where $U = \{(u,z)\in\Re^2;\ \beta + \gamma z^2 > 0\}$, $\mathscr{P} = \Re$, δ,γ,β,ν are real constants, with δ,γ,ν nonzero and $\beta = 0$ when $\gamma = 1$; or

(iii) $F = \lambda u + \zeta u_x + \tau,$

where $U = \Re^2$, $\mathscr{P} = \Re - \{0\}$, and λ,τ,ζ are real constants.

In particular, from (i) and (iii) of the above theorem one gets the following result obtained in Ref. 12 which shows that the nonlinear differential equations $u_{xt} = F(u)$ that describe an η-p.s.s. are essentially the sine–Gordon, sinh–Gordon, and Liouville equations.

Theorem 2: An equation

$$u_{xt} = F(u)$$

describes an η-p.s.s, for $\eta\in\mathscr{P}$, with F independent of η if and only if

$$F''(u) + \alpha F(u) = 0, \tag{2}$$

where $\alpha\in\Re$ and $\mathscr{P} = \Re - \{0\}$.

Finally, we conclude with an interesting result that relates the property of η-p.s.s. with the existence of Bäcklund transformations, In Refs. 13 and 14 it was shown that an equation $u_{xt} = F(u)$ has a self-Bäcklund transformation if only if F satisfies (2). Therefore as an immediate consequence of Theorem 2 we have the following.

Corollary 1: An equation $u_{xt} = F(u)$ has a self-Bäcklund transformation if and only if it describes an η-p.s.s.

In Sec. II we consider the basic concepts, and we fix our notation. In Sec. III we prove Theorem 1. The one-parameter linear problems associated with the equations of Theorem I are given in Remarks 2 and 3.

II. PRELIMINARIES

Let M be a pseudospherical surface (p.s.s.), i.e., a two-dimensional Riemannian manifold with constant Gaussian curvature -1. We consider a local orthonormal frame e_1,e_2

 0022-2488/90/061400-08$03.00

and its dual coframe ω_1,ω_2. We denote by ω_3 the connection form (usually denoted by ω_{12}). Then the following structure equations are satisfied:

$$d\omega_1 = \omega_3 \wedge \omega_2, \quad d\omega_2 = \omega_1 \wedge \omega_3, \quad d\omega_3 = \omega_1 \wedge \omega_2 . \qquad (3)$$

We consider the following definition introduced in Ref. 4: A differential equation for a real function $u(x,t)$ *describes a p.s.s.* if it is the necessary and sufficient condition for the existence of differential functions f_{ij}, $1 \leqslant i \leqslant 3$, $1 \leqslant j \leqslant 2$, depending on u and its derivatives, such that the one-forms

$$\omega_i = f_{i1}\, dx + f_{i2}\, dt \qquad (4)$$

satisfy the structure equations (3) of a p.s.s.

Examples of such differential equations are given by the Korteweg–de Vries equation, the Modified Korteweg–de Vries equation, sine–Gordon equation, sinh–Gordon equation, wave equation, Burgers equation, etc. See Refs. 4,5, and 12 for more examples.

It follows from the above definition that each generic solution (for which $\omega_1 \wedge \omega_2 \neq 0$) of a differential equation (E) that describes a p.s.s. provides a metric on open subsets of $\mathfrak{R}^2$, whose Gaussian curvature is constant equal to -1. Moreover, taking Ω given by

$$\Omega = \frac{1}{2}\begin{pmatrix} \omega_2 & \omega_1 - \omega_3 \\ \omega_1 + \omega_3 & -\omega_2 \end{pmatrix}, \qquad (5)$$

the definition is equivalent to saying that (E) is the integrability condition for the linear problem

$$dv = \Omega v, \qquad (6)$$

where $v = \binom{v_1}{v_2}$. In fact, the compatibility condition for this system is

$$d\Omega - \Omega \wedge \Omega = 0, \qquad (7)$$

which is equivalent to (3).

We note that the property of an equation (E) to describe a p.s.s. is invariant by a change of independent variables. Moreover, there is a linear problem (6) associated to (E). For the purpose of obtaining solutions for (E) by the inverse scattering method (see Refs. 11,15, and 16 for this method) applied to a one-parameter family of linear problems [for which (E) is the integrability condition], one would like the functions f_{ij}, which define ω_i, to depend not only on u and its derivatives, but also on a parameter η. As an example we consider the one-forms

$$\omega_1 = (1/\eta)\sin u\, dt,$$
$$\omega_2 = \eta\, dx + (1/\eta)\cos u\, dt, \qquad (8)$$
$$\omega_3 = u_x\, dx .$$

Then the ω_i satisfy (3) if and only if

$$u_{xt} = \sin u,$$

i.e., u is a solution of the sine–Gordon equation.

In our main results the parameter $\eta \in \mathscr{P} = \mathfrak{R} - S$, where S is a set of isolated points. In the above example [(8)] we have $\eta \in \mathfrak{R} - \{0\}$.

We say that a differential equation *describes an η-p.s.s.* if it describes a p.s.s. with $f_{21} = \eta$. This definition was motivated by the equations associated to the AKNS system.[11] Such equations describe an η-p.s.s. However, they have an additional condition that corresponds to the requirement that f_{11} and f_{31} do not depend on the parameter η.

We observe that the property of describing an η-p.s.s. is not preserved by a change of independent variables. Moreover, it does not exclude the possibility of the differential equation depending on η. The following result provides a geometrical interpretation for the condition $f_{21} = \eta$.

Proposition 1: Let (E) be a differential equation that describes a p.s.s. Then (E) describes an η-p.s.s if and only if generic solutions of (E) define a metric for which the length of the vector field $\partial/\partial x$ satisfies

$$\left|\frac{\partial}{\partial \xi}\right|^2 \geqslant \eta^2 . \qquad (9)$$

Proof: The equation (E) describes a p.s.s., i.e., there exist one-forms ω_i, defined as in (4), that satisfy (3).

Suppose $f_{21} = \eta$. Then the first fundamental form, defined by $I = \omega_1^2 + \omega_2^2$, gives

$$\left|\frac{\partial}{\partial x}\right|^2 = f_{11}^2 + \eta^2 \geqslant \eta^2 .$$

Conversely, suppose the metric $\langle\ ,\ \rangle$, defined by generic solutions of (E), satisfies (9). We show that there exists a frame field $\bar{e}_1\bar{e}_2$ and its dual coframe $\bar{\omega}_1,\bar{\omega}_2$ such that $\bar{\omega}_2(\partial/\partial x) + \eta$. In fact, using the notation

$$g_{11} = \left\langle\frac{\partial}{\partial x},\frac{\partial}{\partial x}\right\rangle, \quad g_{12} = \left\langle\frac{\partial}{\partial x},\frac{\partial}{\partial t}\right\rangle, \quad g_{22} = \left\langle\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right\rangle,$$

we define

$$\bar{e}_2 = a\frac{\partial}{\partial x} + b\frac{\partial}{\partial t},$$

where

$$b^2 = (g_{11} - \eta^2)/(g_{11}g_{22} - g_{12}^2), \quad a = (\eta - bg_{12})/g_{11}.$$

Then $|\bar{e}_2| = 1$ and $\langle \bar{e}_2, \partial/\partial x\rangle = \eta$. Hence

$$\frac{\partial}{\partial x} = c\bar{e}_1 = \eta\bar{e}_2 ,$$

where c is a differential function and $\bar{e}_1,\bar{e}_2$ is an orthonormal frame field. Therefore the dual coframe satisfies $\bar{\omega}_2(\partial/\partial x) = \eta$.

As we mentioned above, we are interested in differential equations (E) that are the integrability condition of a one-parameter family of linear problems. Hence we want to obtain equations (E) that describe an η-p.s.s. such that (E) is independent of the parameter η. This is done in the following section for equations of type $u_{xt} = F(u,u_x)$.

III. PROOF OF THE MAIN RESULTS

In this section we consider solutions of an equation $u_{xt} = F(u,u_x)$ as integral manifolds of an exterior differential system. In Lemma 1 we give necessary conditions on the functions f_{ij} for the one-forms $\omega_i = f_{i1}\, dx + f_{i2}\, dt$ to be associated to a differential equation $u_{xt} = F(u,u_x)$, which describes an η-p.s.s. In particular, we obtain that f_{11} and f_{31} are functions of u_x. Denoting by L the Wronskian of f_{11} and f_{31} we characterize F, when $L \neq 0$ in Theorem 3. The case $L = 0$ is considered in Theorem 5. In those two results we consider η to be fixed.

In Theorems 4 and 6, assuming that f_{ij} and F depend differentiably on $\eta \in I$, where I is an open interval, we characterize the functions F, obtained in the previous results, which are independent of η.

Finally, we prove Theorem 1, by imposing the parameter η to vary in $\mathscr{P}$, where $\mathscr{P} = \Re - S$, and S is a set of isolated points. The linear problems associated to the differential equations of Theorem 1 are obtained from the proofs of the above-mentioned theorems.

Cartan–Kähler theory relates solutions of differential equations with integral manifolds of exterior differential systems.[17] In particular, for differential equations of the type

$$u_{xt} = F(u,u_x),$$

for a function $u(x,t)$, we obtain the following result, which we prove for the sake of completeness.

Proposition 2: Let $\mathscr{I}$ be the ideal generated in the space of variables x,t,u,z, by the two-forms

$$\Omega_1 = du \wedge dt - z\, dx \wedge dt,$$
$$\Omega_2 = dz \wedge dx + F(u,z) dx \wedge dt, \tag{10}$$

where $F(u,z)$ is a real differentiable function. Then $\mathscr{I}$ is a closed differential ideal. If $u(x,t)$ is a solution of

$$u_{xt} = F(u,u_x), \tag{11}$$

then the map

$$\phi(x,t) = (x,t,u(x,t),z(x,t)), \tag{12}$$

with $z = u_x$, defines an integral manifold of $\mathscr{I}$. Conversely, any two-dimensional integral manifold of $\mathscr{I}$ given by

$$\phi(s,r) = (x(s,r),t(s,r),u(s,r),z(s,r)), \tag{13}$$

with dx and dt linearly independent, determines a local solution of the equation (11).

Proof: We have that $\mathscr{I}$ is a closed differential ideal, since

$$d\Omega_1 = -\Omega_2 \wedge dt,$$
$$d\Omega_2 = -F_u dx \wedge \Omega_1 + F_z \Omega_2 \wedge dt.$$

Suppose u is a solution of (11). We need to show that, for ϕ defined by (12), we have $\phi^*\Omega_i = 0, i = 1,2$. From the definition of ϕ^* and (10) it follows that

$$\phi^*\Omega_1 = u_x\, dx \wedge dt - z\, dx \wedge dt,$$
$$\phi^*\Omega_2 = -z_t\, dx \wedge dt = F\, dx \wedge dt.$$

Since $z = u_x$ and u satisfies (11), we get $\phi^*\Omega_i = 0$.

Conversely, if ϕ given by (13) is an integral manifold of $\mathscr{I}$, such that $dx \wedge dt \neq 0$, then we locally have $(s,r) = h(x,t)$. Taking $\bar{\phi} = \phi \circ h$, we get $\bar{\phi}^*\Omega_i = h^*\phi^*\Omega_i = 0$. Therefore

$$0 = (u_x - z) dx \wedge dt,$$
$$0 = (-u_{xt} + F(u,z)) dx \wedge dt.$$

We conclude that $z = u_x$ and $u_{xt} = F(u,z)$, i.e., u is a solution of (11).

The following result gives necessary conditions on the functions $f_{ij}(u,z)$ for the one-forms $\omega_i = f_{i1}\, dx + f_{i2}\, dt$ to be associated to a differential equation $z_t = F(u,z)$, which describes an η-p.s.s. From now on the variables appearing in the lower indices will denote partial differentiation. Unless explicitly stated, the functions f_{ij} and F may depend on η.

Lemma 1: Let

$$z_t = F(u,z) \tag{14}$$

be a differential equation that describes an η-p.s.s. with associated one-forms $\omega_i = f_{i1}\, dx + f_{i2}\, dt$, where f_{ij} and F are real differentiable (C^∞) functions on an open connected set $U \subset \Re^2$. Then

$$f_{11,u} \equiv f_{31,u} \equiv 0, \tag{15}$$
$$f_{12,z} \equiv f_{22,z} \equiv f_{32,z} \equiv 0, \tag{16}$$
$$f_{11,z}^2 + f_{31,z}^2 \neq 0 \quad \text{in } U. \tag{17}$$

Moreover,

$$\begin{aligned} &-Ff_{11,z} + zf_{12,u} + \eta f_{32} - f_{22}f_{31} \equiv 0, \\ &zf_{22,u} - f_{11}f_{32} + f_{12}f_{31} \equiv 0, \\ &-Ff_{31,z} + zf_{32,u} + \eta f_{12} - f_{22}f_{11} \equiv 0. \end{aligned} \tag{18}$$

Proof: In the space of variables x,t,u,z, we consider the ideal $\mathscr{I}$ generated by Ω_i defined by (10) where F is given by (14). It follows from Proposition 2 that $\Omega_i = 0$ when restricted to each integral manifold of $\mathscr{I}$. Hence, for u,z satisfying (14), we have

$$du \wedge dt = z\, dx \wedge dt, \quad dz \wedge dx = -F\, dx \wedge dt. \tag{19}$$

The one-forms ω_i satisfy the structure equations (3). Therefore

$$df_{11} \wedge dx + df_{12} \wedge dt + (\eta f_{32} - f_{22}f_{31}) dx \wedge dt = 0,$$
$$df_{22} \wedge dt + (f_{12}f_{31} - f_{11}f_{32}) dx \wedge dt = 0,$$
$$df_{31} \wedge dx + df_{32} \wedge dt + (\eta f_{12} - f_{22}f_{11}) dx \wedge dt = 0.$$

Substituting

$$df_{ij} = f_{ij,u}\, du + f_{ij,z}\, dz$$

and (19) into the above equations and equating to zero the coefficients of the independent two-forms, we obtain (15), (16), and (18). Relation (17) follows from the fact that (14) is the necessary and sufficient condition for (18) to be satisfied.

In Lemma 1, we showed that a necessary condition for (14) to describe an η-p.s.s. is that f_{ij} satisfy (15)–(17). Therefore we will assume these conditions in order to characterize such equations. We introduce the notation

$$L = \begin{vmatrix} f_{11} & f_{31} \\ f_{11,z} & f_{31,z} \end{vmatrix}. \tag{20}$$

Assuming the condition $L \neq 0$, we obtain Theorem 3, which shows that the function F in (14) is algebraically determined by f_{11}, f_{31}, and f_{22}. The case $L = 0$ is considered in Theorem 5. In both theorems, $\eta \neq 0$ is considered to be fixed.

Remark 1: An equation $z_t = F(u,z)$, where F is independent of u, can be considered as an ordinary differential equation on z. Therefore in the following results we will require F to depend on u. However, for the sake of completeness, we note that any equation $z_t = F(z)$, where F is a differentiable function defined on an interval I, with $F(z) \neq 0$, describes an η-p.s.s. In fact, consider

$$\omega_1 = f_{11}(z) dx + (1/\eta) dt, \quad \omega_2 = \eta dx, \quad \omega_3 = \omega_1, \tag{21}$$

where $f_{11,z} = 1/F(z)$. Then (21) satisfy (3) if and only if $z_t = F(z)$.

The wave equation $z_t = 0$ is the necessary and sufficient condition for the one-forms

$$\omega_1 = z\, dx, \quad \omega_2 = \eta\, dx + e^u\, dt, \quad \omega_3 = \eta\, dx + e^u\, dt$$

to satisfy (3).

Theorem 3: Let f_{ij}, $1\leqslant i\leqslant 3$, $1\leqslant j\leqslant 2$, and F be differentiable (C^∞) functions defined on an open connected set $U\subset\mathfrak{R}^2$, such that (15)–(17) hold. Suppose $L\neq 0$ in U and $F_u\neq 0$ in a dense subset of U. Then $z_t = F(u,z)$ describes an η-p.s.s. with associated one-forms $\omega_i = f_{i1}\,dx + f_{i2}\,dt$, if and only if,

$$\text{(i)}\quad F = Ee^{Du}[Af_{11} - Bf_{31} + \eta D(B^2 - A^2)] \tag{22}$$

and

$$f_{22} = Ee^{Du},\quad f_{12} = Af_{22,u},\quad f_{32} = Bf_{22,u}, \tag{23}$$

f_{11} and f_{31} are algebraically determined by

$$Bf_{11} - Af_{31} = z,$$
$$f_{11}^2 - f_{31}^2 + 2\eta D(Bf_{31} - Af_{11}) - D^2z^2 + G = 0, \tag{24}$$

where A, B, D, E, and G are constants, which may depend on η, with $1 - D^2(B^2 - A^2)\neq 0$ and $DE\neq 0$; or

$$\text{(ii)}\quad F = \eta(B^2 - A^2)f_{22,u} + Qf_{22} \tag{25}$$

and

$$f_{11} = (Bz - AQ)/(B^2 - A^2),$$
$$f_{31} = (Az - BQ)/(B^2 - A^2), \tag{26}$$
$$f_{12} = Af_{22,u},\quad f_{32} = Bf_{22,u},$$

f_{22} satisfies the equation

$$f_{22,uu} + f_{22}/(A^2 - B^2) = 0, \tag{27}$$

where A, B, and Q are constants that may depend on η and $A^2 - B^2\neq 0$, $Q\neq 0$.

Proof: Suppose that $z_t = F(u,z)$ describes an η-p.s.s. Then, from Lemma 1, we have that (18) is satisfied. This is equivalent to the system

$$z(f_{31,z}f_{12,u} - f_{11,z}f_{32,u}) + \eta(f_{32}f_{31,z} - f_{12}f_{11,z}) + f_{22}H\equiv 0, \tag{28}$$

$$zf_{22,u} - f_{11}f_{32} + f_{31}f_{12}\equiv 0, \tag{29}$$

$$FL + z(f_{31}f_{12,u} - f_{11}f_{32,u}) + \eta(f_{31}f_{32} - f_{11}f_{12}) - f_{22}(f_{31}^2 - f_{11}^2)\equiv 0, \tag{30}$$

where

$$H = \begin{vmatrix} f_{11} & f_{31} \\ f_{31,z} & f_{11,z}\end{vmatrix}. \tag{31}$$

Taking the derivative of (29) with respect to z, it follows from (15) and (16) that

$$f_{22,u} - f_{32}f_{11,z} + f_{12}f_{31,z}\equiv 0. \tag{32}$$

From (29) and (32) we obtain

$$f_{12} = f_{22,u}(zf_{11,z} - f_{11})/L,$$
$$f_{32} = f_{22,u}(zf_{31,z} - f_{31})/L. \tag{33}$$

Let $U^1 = \{(u,z)\in U;\ f_{22,u}\neq 0\}$. It follows from (30), (33), and the hypothesis on F_u that U^1 is an open dense subset of U. Now, in each connected component U_c^1 of U^1, from (15), (16), and (33) we obtain

$$(zf_{11,z} - f_{11})/L = A, \tag{34}$$

$$(zf_{31,z} - f_{31})/L = B, \tag{35}$$

$$f_{12} = Af_{22,u},\quad f_{32} = Bf_{22,u}. \tag{36}$$

Using (34) and (35) we have, in U_c^1,

$$Bf_{11} - Af_{31} = z. \tag{37}$$

Observe that (37) implies that $A^2 + B^2\neq 0$. Now, using (36) and (37), Eq. (28) in U_c^1 reduces to

$$-zf_{22,uu} + \eta f_{22,u}(Bf_{31,z} - Af_{11,z}) + f_{22}H = 0. \tag{38}$$

Taking the derivative of this equation with respect to z, multiplying the result by $-z$, and adding to (38), we get

$$\eta f_{22,u}[z(Bf_{31} - Af_{11})_{zz} - (Bf_{31} - Af_{11})_z] + f_{22}(zH_z - H) = 0. \tag{39}$$

We denote

$$Y = z(Bf_{31} - Af_{11})_{zz} - (Bf_{31} - Af_{11})_z.$$

Let $V^1 = \{(u,z)\in U_c^1;\ Y\neq 0\}$ and $V^0 = U_c^1 - V^1$. In each connected component V_c^1 of V^1 we have, from (39), (15), and (16),

$$f_{22} = Ee^{Du}, \tag{40}$$

where $ED\neq 0$, since $V^1\subset U^1$. Substituting (40) and H defined by (31) in (38), we obtain, by integrating on z,

$$f_{11}^2 - f_{31}^2 + 2\eta D(Bf_{31} - Af_{11}) - z^2D^2 + G = 0,$$

where G is an integration constant. This last equation and (37) give f_{11} and f_{31} algebraically determined by (24) in terms of z, η, A, B, D, and G. Moreover, since $Y\neq 0$, we get $1 - D^2(B^2 - A^2)\neq 0$. In order to obtain F we note that (34) and (35) imply

$$f_{31}^2 - f_{11}^2 = L(Af_{11} - Bf_{31}) - zH,$$
$$Af_{11} - Bf_{31} = (B^2 - A^2)L + z(Af_{11,z} + Bf_{31,z}).$$

Using these relations and (38) in Eq. (30) we obtain F given by (22) in V_c^1.

Suppose int $V^0\neq\varnothing$. In a connected subset V_c^0 of int V^0, we have $Y\equiv 0$, i.e.,

$$((Bf_{31} - Af_{11})_z/z)_{zz} = 0.$$

Integrating twice with respect to z, we get

$$Bf_{31} - Af_{11} + Cz^2/2 + Q = 0,$$

where C and Q are constants, that may depend on η. Using (37) and the last equation, we obtain $B^2 - A^2\neq 0$ and

$$f_{11} = (Bz - ACz^2/2 - AQ)/(B^2 - A^2),$$
$$f_{31} = (Az - BCz^2/2 - BQ)/(B^2 - A^2). \tag{41}$$

From (39) we have

$$f_{22}(zH_z - H) = 0.$$

Since $f_{22,u}\neq 0$ in U_c^1, we must have $zH_z - H = 0$, i.e., $C = 0$. Now substituting (41) (with $C = 0$) into (38) we obtain (27). Moreover, (26) follows from (41) and (36). Since $L\neq 0$, from (26) we have $Q\neq 0$. Substituting (26) into (30) we get (25).

We observe that, as a consequence of the continuity of the function f_{22} and its derivatives, we have int $V^0 = 0$ or $V^1 = \varnothing$. Otherwise, in the boundary $\mathrm{Fr}\,V^1\cap U_c^1 = \mathrm{Fr}(\mathrm{int}\,V^0)\cap U_c^1\neq\varnothing$, Eqs. (23) and (27) imply $1 - D^2(B^2 - A^2) = 0$, which is a contradiction. Therefore, $V^1 = \varnothing$ and therefore $U_c^1 = V^0$ or else int $V^0 = \varnothing$ and hence V^1 is dense in U_c^1 and the constants E, D, and G do not differ in the connected components of U_c^1, since f_{ij} and its derivatives are continuous. From this, we conclude that in each connected component U_c^1 of U^1 the functions are given by (i) or (ii).

Now we show that (i) or (ii) occur in all U. In fact, if there exists a connected component U_c^1 where (i) occurs then it follows from (23) that Fr $U_c^1 \cap U^0 = \varnothing$; therefore $U = U^1 = U_c^1$. Otherwise, in each connected component of U^1 the functions are given as in (ii). Since U^1 is dense in U, it follows by the continuity of the functions that (ii) defines F and f_{ij} in U.

The converse in both cases is a straightforward computation.

In the following theorem we consider the dependence of the functions F and f_{ij} on the parameter η.

Theorem 4: Let f_{ij} and F be differentiable functions defined on $U\times I$, where U is an open connected set of $\Re^2$ and I is an open interval. Suppose, for each $\eta \in I$, $z_t = F(u,z,\eta)$ describes an η-p.s.s. as in Theorem 3. Then F is independent of η if and only if

(i) $F = \nu e^{\delta u}\sqrt{\beta + \gamma z^2}$,

where $U = \{(u,z)\in\Re^2;\ \beta + \gamma z^2 > 0\}$, $I = \Re$, δ, β, γ, ν are real constants with δ, γ, ν nonzero and $\beta = 0$ when $\gamma = 1$; or

(ii) $F''(u) + \alpha F(u) = 0$,

where F is independent of z, $U = \Re^2$, $I = \Re^+$ or $\Re^-$, and α is a nonzero real constant.

Proof: Since $z_t = F$ describes an η-p.s.s. as in Theorem 3, there exist A and B, differentiable functions of η such that

$$Bf_{11} - Af_{31} = z.$$

Denote

$$Y(u,z,\eta) = z(Bf_{31} - Af_{11})_{zz} - (Bf_{31} - Af_{11})_z.$$

Let $W^1 = \{(u,z,\eta)\in U\times I;\ Y \neq 0\}$ and $W^0 = (U\times I) - W^1$. In each connected component W_c^1 of W^1 we have F given by (22), where f_{11} and f_{31} are determined by (24). Therefore

$$F = \pm E e^{Du}\sqrt{\Delta},$$

where

$$\Delta = \eta^2 D^2 (B^2 - A^2)^2 + G(B^2 - A^2) + (1 - D^2(B^2 - A^2))z^2.$$

Since F is independent of η, we have

$$D = \delta, \quad \pm E = \nu, \quad 1 - \delta^2(B^2 - A^2) = \gamma,$$

where ν, δ, γ are nonzero real constants. It follows that

$$\Delta = \beta + \gamma z^2,$$

where β is independent of η and satisfies

$$\beta = [(1-\gamma)G + \eta^2(1-\gamma)^2]/\delta^2. \tag{42}$$

This equation determines G when $\gamma \neq 1$ and shows that $\beta = 0$ when $\gamma = 1$. In the latter case we can choose $G \equiv 0$. We conclude that F is given by (i) in W_c^1.

If int $W^0 \neq \varnothing$, in each connected component W_c^0 of int W^0 we have F given by (25), where f_{22} satisfies (27). From these two equations we obtain that F is a solution of (ii), where $\alpha = 1/(A^2 - B^2)$ is independent of η.

It follows from the smoothness of F that F is defined on U by (i) or (ii). Its associated functions f_{ij} are defined on $U\times I$ (see Remark 2), where

$$U = \{(u,z)\in\Re^2;\ \beta + \gamma z^2 > 0\}, \quad I = \Re \text{ for the case (i)},$$

and

$$U = \Re^2, \quad I = \Re^+ \text{ or } \Re^-, \qquad \text{for the case (ii)}.$$

The converse in both cases is straightforward.

Remark 2: The one-parameter linear problem (6) associated to the equation $z_t = F$, where F is given by Theorem 4(i) or 4(ii) is obtained from the previous proof. In fact, it is not difficult to see that

$$\omega_1 = (\eta A\delta + \delta^2(Bu_z \mp A\sqrt{\Delta})/(1-\gamma))dx \pm A\nu\delta e^{\delta u}\,dt,$$

$$\omega_2 = \eta\,dx \pm \nu e^{\delta u}\,dt,$$

$$\omega_3 = (\eta B\delta + \delta^2(Au_x \mp B\sqrt{\Delta})/(1-\gamma))dx \pm B\nu\delta e^{\delta u}\,dt,$$

where $\Delta = \beta + \gamma u_x^2$, $B^2 - A^2 = (1-\gamma)/\delta^2$, and $\gamma \neq 1$ satisfy the structure equations (3) if and only if

$$u_{xt} = \nu e^{\delta u}\sqrt{\beta + \gamma u_x^2}.$$

When $\gamma = 1$ we have $\beta = 0$ and F reduces to $F = \nu e^{\delta u} z$. Choosing $G = 0$ in (42) we have

$$\omega_1 = (\pm(1/A + \delta^2 A)u_x/2 + \eta\delta A)dx \pm A\nu\delta e^{\delta u}\,dt,$$

$$\omega_2 = \eta\,dx \pm \nu e^{\delta u}\,dt,$$

$$\omega_3 = (\pm(-1/A + \delta^2 A)u_z/2 \pm \eta\delta A)dx + A\nu\delta e^{\delta u}\,dt,$$

where $A \neq 0$, satisfy (3) if and only if

$$u_{xt} = \nu e^{\delta u} u_x.$$

When F is given by Theorem 4(ii), we have $\alpha = 1/(A^2 - B^2)$. From (25) and (27) we obtain

$$f_{22} = \eta(F_u + (\alpha/\eta)QF)/(Q^2\alpha + \eta^2).$$

Therefore using (26) we have

$$\omega_1 = -\alpha(Bu_x - AQ)dx + [A\alpha/(Q^2\alpha + \eta^2)](QF_u - \eta F)dt,$$

$$\omega_2 = \eta\,dx + [(\eta F_u + \alpha QF)/(Q^2\alpha + \eta^2)]dt,$$

$$\omega_3 = -\alpha(Au_x - BQ)dx + [B\alpha/(Q^2\alpha + \eta^2)](QF_u - \eta F)dt,$$

where η, Q, and α are nonzero. These forms satisfy (3) if and only if

$$u_{xt} = F(u).$$

In the following theorem we consider the case $L = 0$ in U, i.e., f_{11} and f_{31} are linearly independent in U; hence there exist A and B that do not vanish simultaneously such that $Af_{11} + Bf_{31} = 0$. As in Theorem 3 we consider $\eta \neq 0$ to be fixed and A, B may depend on η.

Theorem 5: Let f_{ij}, $1 \leqslant i \leqslant 3$, $1 \leqslant j \leqslant 2$, and F be differentiable functions defined on an open connected set $U \subset \Re^2$, such that (15)–(17) hold. Suppose there exist A, B satisfying $Af_{11} + Bf_{31} = 0$ in U, $A^2 + B^2 \neq 0$, and $F_u \neq 0$ in a dense subset of U. Then $z_t = F(u,z)$ describes an η-p.s.s. with associated one-forms $\omega_i = f_{i1}dx + f_{i2}dt$ if and only if

$$\text{(i)}\quad F = \eta(B^2 - A^2)f_{22,u}/E^2 \tag{43}$$

and

$$f_{11} = Dz, \quad f_{31} = Cz,$$
$$f_{12} = -Af_{22,u}/E, \quad f_{32} = Bf_{22,u}/E, \tag{44}$$

f_{22} satisfies the equation

$$(B^2 - A^2)f_{22,uu} - E^2 f_{22} = 0, \tag{45}$$

where A, B, C, D, and E are constants, which may depend on η, such that $A^2 - B^2 \neq 0$, $E = AC + BD \neq 0, AD + BC = 0$; or

$$\text{(ii)} \quad F = (zf_{12,u} \pm \eta f_{12} \mp Kf_{11})/f_{11,z} \tag{46}$$

and

$$f_{22} = K, \quad f_{31} = \pm f_{11}, \quad f_{32} = \pm f_{12}, \tag{47}$$

where K is a constant that may depend on η and $f_{12,u} \neq 0$ in a dense subset of U.

Proof: Suppose $z_t = F(u,z)$ describes an η-p.s.s. Then from Lemma 1 we have (18) satisfied. This is equivalent to the system

$$z(Af_{12,u} + Bf_{32,u}) + \eta(Af_{32} + Bf_{12}) - f_{22}(Af_{31} + Bf_{11}) \equiv 0, \tag{48}$$

$$zf_{22,u} - f_{11}f_{32} + f_{31}f_{12} \equiv 0, \tag{49}$$

$$-F(Bf_{11,z} + Af_{31,z}) + z(Bf_{12,u} + Af_{32,u}) + \eta(Bf_{32} + Af_{12}) \equiv 0, \tag{50}$$

where we have used the hypothesis on A and B. Multiplying (49) by A and B and using $Af_{11} + Bf_{31} = 0$, we get, respectively,

$$Azf_{22,u} + (Bf_{32} + Af_{12})f_{31} \equiv 0, \tag{51}$$

$$Bzf_{22,u} - (Bf_{32} + Af_{12})f_{11} \equiv 0. \tag{52}$$

(i) Suppose $A^2 - B^2 \neq 0$. We consider

$$U^1 = \{(u,z); \quad Bf_{32} + Af_{12} \neq 0\}$$

and $U^0 = U - U^1$. We claim that U^1 is an open, dense subset of U. Otherwise, suppose V is an open subset contained in U^0; then, from (51) and (52), we get $f_{22,u} = 0$ on V. Combining this result with the derivative of (48) with respect to u, we get $Af_{32,u} + Bf_{12,u} = 0$ on V. Therefore, it follows from (50) that $F_u = 0$ on V, which contradicts the hypothesis on F. Therefore, U^1 is dense in U. Now, in each connected component U^1_c of U^1, it follows from (51), (52) and (15), (16) that

$$f_{31} = Cz, \quad f_{11} = Dz, \tag{53}$$

where C and D are constants, which may depend on η. Since U^1 is dense in U, from the continuity of f_{11}, f_{31}, and their derivatives, we have f_{11} and f_{31} defined by (53) in U. Observe that by hypothesis $Af_{11,z} + Bf_{31,z} \equiv 0$ and (17) holds; therefore $AD + BC = 0$, $Bf_{11,z} + Af_{31,z} \neq 0$ in U, and $E \equiv AC + BD \neq 0$. Taking the derivatives of (48) with respect to z, we get

$$Af_{12,u} + Bf_{32,u} - f_{22}(AC + BD) \equiv 0. \tag{54}$$

Therefore it follows from (48) that

$$Af_{32} + Bf_{12} \equiv 0.$$

This relation and (49) imply that

$$f_{12} = -Af_{22,u}/(AC + BD), \quad f_{32} = Bf_{22,u}/(AC + BD). \tag{55}$$

Hence (44) follows from (53) and (55). Taking these results into (54) we obtain (45). Substituting (44) and (45) into (50), we obtain F given by (43).

(ii) Suppose $A^2 - B^2 = 0$. It follows from the hypothesis $Af_{11} + Bf_{31} \equiv 0$ that $Bf_{11,z} + Af_{31,z} \equiv 0$. Using (50) we obtain

$$z(Bf_{12,u} + Af_{32,u}) + \eta(Bf_{32} + Af_{12}) \equiv 0.$$

From this equation we have

$$Bf_{32} + Af_{12} \equiv 0. \tag{56}$$

Since $A^2 + B^2 \neq 0$, it follows from (51) and (52) that $f_{22,u} \equiv 0$, i.e.,

$$f_{22} = K. \tag{57}$$

Since $A/B = \pm 1$, combining (56), (57), and the first equation of (18), we obtain F given by (46). The hypothesis $Af_{11} + Bf_{31} \equiv 0$ with (56) and (57) imply (47). From (46), we see that $f_{12,u} \neq 0$ in a dense subset of U.

The converse in both cases is a straightforward computation.

As in Theorem 4, in the next theorem we consider the dependence of the functions F and f_{ij} on the parameter η.

Theorem 6: Let f_{ij} and F be differentiable functions defined on $U \times I$, where U is an open connected set of $\Re^2$ and I is an open interval, such that $0 \notin I$. Suppose, for each $\eta \in I$, $z_t = F(u,z,\eta)$ describes an η-p.s.s. as in Theorem 5. Then F is independent of η, if and only if

$$\text{(i)} \quad F''(u) + \alpha F(u) = 0,$$

where F is independent of z, $U = \Re^2$, $I = \Re^+$ or $\Re^-$, and α is a nonzero real constant; or

$$\text{(ii)} \quad F = \nu e^{\theta u + h(z)},$$

where $U = \Re \times J$, J is an open interval, h is a differentiable function on J, $\nu, \theta \in \Re - \{0\}$, and $U \times I$ does not intersect one of the planes $\{(u,z,\eta) \in \Re; \ \pm \eta + \theta z = 0\}$; or

$$\text{(iii)} \quad F = \lambda u + \zeta z + \tau,$$

where $U = \Re^2$, $I = \Re^+$ or $\Re^-$, λ, ζ, τ are real constants, and $\lambda \neq 0$.

Proof: Since $z_t = F$ describes an η-p.s.s. as in Theorem 5, there exist A and B differentiable functions in η such that $Af_{11} + Bf_{31} \equiv 0$ in $U \times I$. Let $I^1 = \{\eta \in I; A^2 - B^2 \neq 0\}$, and $I^0 = I - I^1$.

In each connected component of $U \times I^1$ it follows immediately from (43) and (45) that F satisfies Theorem 6(i), with $\alpha = E^2/(A^2 - B^2)$ independent of η.

Suppose int $I^0 \neq 0$. In each connected component $U \times J$ of $U \times \text{int } I^0$, from Theorem 5(ii) we have F given by (46) and $f_{11,z} \neq 0$. Taking the derivative of F with respect to η, we get, in $U \times J$,

$$\frac{f_{11,z\eta}}{f_{11,z}}[zf_{12,u} \pm \eta f_{12} \mp Kf_{11}] - [zf_{12,u\eta} \pm (\eta f_{12})_\eta \mp (kf_{11})_\eta] \equiv 0. \tag{58}$$

Taking the derivative with respect to u and then to z we obtain

$$\frac{f_{11,z\eta}}{f_{11,z}}[zf_{12,uu} \pm \eta f_{12,u}] - [zf_{12,uu\eta} \pm (\eta f_{12,u})_\eta] \equiv 0, \tag{59}$$

$$\left(\frac{zf_{11,z\eta}}{f_{11,z}}\right)_z f_{12,uu} \pm \eta\left(\frac{f_{11,z\eta}}{f_{11,z}}\right)_z f_{12,u} - f_{12,uu\eta} \equiv 0. \tag{60}$$

Let $W = \{(u,z,\eta) \in U \times J; f_{12,u} \neq 0\}$. From Theorem 5, for each $\eta_0 \in J$, we have $f_{12,u}(u,\eta_0) \neq 0$ in a dense subset of U. Therefore W is dense in $U \times J$.

Now we restrict ourselves to W. Dividing (60) by $f_{12,u}$ and taking the derivative with respect to z,

$$\left(\frac{zf_{11,z\eta}}{f_{11,z}}\right)_{zz} \frac{f_{12,u}}{f_{12,u}} \pm \eta\left(\frac{f_{11,z\eta}}{f_{11,z}}\right)_{zz} \equiv 0. \tag{61}$$

We denote

$$Z(z,\eta) = \left(\frac{zf_{11,z\eta}}{f_{11,z}}\right)_{zz}.$$

Let $W^1 = \{(u,z,\eta) \in W; Z \neq 0\}$ and $W^0 = W - W^1$.

(a) In each connected component W_c^1 of W^1 from (61) we obtain

$$f_{12,u} = N(\eta)e^{M(\eta)u}, \tag{62}$$

where $N(\eta) \neq 0$, and

$$M\left(\frac{zf_{11,z\eta}}{f_{11,z}}\right)_z \pm \eta\left(\frac{f_{11,z\eta}}{f_{11,z}}\right)_z = P_\eta(\eta). \tag{63}$$

Using these results in (60) we get

$$M = \theta = \text{const}, \tag{64}$$

$$P_\eta = \theta(\log N)_\eta. \tag{65}$$

Integrating (63) on z, we obtain

$$(\theta z \pm \eta)\frac{f_{11,z\eta}}{f_{11,z}} = P_\eta z + Q_\eta. \tag{66}$$

Substituting into (59) we have

$$Q_\eta = \pm[1 + \eta(\log N)_\eta].$$

Combining this equation with (66) and integrating on η, we obtain, in W_c^1,

$$f_{11,z} = \gamma N(\theta z \pm \eta)e^{-h(z)}, \tag{67}$$

where γ is a nonzero real constant, and h is a differentiable function on an open interval J. The constant $\theta \neq 0$, otherwise $Z \equiv 0$, which is a contradiction since we are in W_c^1. Moreover, since $f_{11,z} \neq 0$ we must have $W_c^1 \cap \Pi = 0$, where Π is one of the planes $\theta z \pm \eta = 0$. From now on we fix such a plane. It follows from (62) and (64) that

$$f_{12} = (N/\theta)e^{\theta u} + R(\eta). \tag{68}$$

Using (67) and (68) in (46), we get, on W_c^1,

$$F = \frac{e^{h(z)}}{\gamma}\left(\frac{e^{\theta u}}{\theta} \pm \frac{\eta R - Kf_{11}}{N(\theta z \pm \eta)}\right).$$

Since F is independent of η we must have

$$\eta R - Kf_{11} = N(\theta z \pm \eta)m(z), \tag{69}$$

where m is a function of z only. Taking the derivative of (69) with respect to z, using (67), and derivating the result twice with respect to η, we easily obtain

$$K = m = R = 0. \tag{70}$$

Hence we get F given by Theorem 6(ii) on W_c^1, where $\nu = 1/(\gamma\theta)$.

(b) If int $W^0 \neq \varnothing$, in each connected component W_c^0 contained in int W^0, we have $Z \equiv 0$. Using this relation in (61) we get

$$\left(\frac{f_{11,z\eta}}{f_{11,z}}\right)_z = 0,$$

which, substituted in (60), provides

$$\frac{f_{11,z\eta}}{f_{11,z}} f_{12,uu} - f_{12,uu\eta} = 0. \tag{71}$$

Now (59) reduces to

$$\eta\frac{f_{11,z\eta}}{f_{11,z}} f_{12,u} - (\eta f_{12,u})_\eta = 0. \tag{72}$$

Taking the derivative of this expression with respect to u and using (71), we obtain

$$f_{12} = Tu + R, \tag{73}$$

where $T \neq 0$ and R depend only on η. It follows from (72) and (73) that

$$f_{11,z} = \eta T g(z), \tag{74}$$

where $g(z) \neq 0$. Taking the derivative of (58) with respect to z and using (73) and (74), we get

$$g = \pm 1/\lambda, \quad K = \lambda/\eta \mp \zeta, \tag{75}$$

where $\lambda \neq 0$ and ζ are real constants. Therefore

$$f_{11} = \pm(1/\lambda)\eta T z + Q(\eta), \tag{76}$$

and it follows from (46) and (73) that F is given by Theorem 6(iii), where $R = \tau T/\lambda$, τ is a real constant, and without loss of generality we choose $Q \equiv 0$.

We observe that it follows from the smoothness of F that F is defined on U by Theorem 6(i) or 6(ii) or 6(iii), with its associated functions f_{ij} defined on $U \times I$, where $I = \Re^+$ or $\Re^-$, for (i) and (iii).

For F given by (ii), we have $U = \Re \times J$, $(U \times I) \cap \Pi = \varnothing$, where Π is one of the planes $\{(u,z,\eta) \in \Re^3, \pm\eta + \theta z = 0\}$. The converse is a straightforward computation.

Remark 3: The one-parameter linear problem (6) associated to the equation $z_t = F$, for F as in Theorem 6, is obtained from Theorem 5 and the above proof.

Consider F given as in Theorem 6(i). Then we have $\alpha = (AC + BD)^2/(A^2 - B^2) \neq 0, A^2 - B^2 \neq 0$. From (43) and (45) we obtain

$$f_{22} = (1/\eta)F_u.$$

Therefore using (44) we have

$$\omega_1 = Du_x\, dx + [A\alpha/\eta(AC + BD)]F\, dt,$$

$$\omega_2 = \eta\, dx + (1/\eta)F_u\, dt,$$

$$\omega_3 = Cu_x\, dx - [B\alpha/\eta(AC + BD)]F\, dt,$$

where $\eta \neq 0$ and $AD + BD = 0$. These forms satisfy (3) if and only if

$$u_{xt} = F(u).$$

In particular, in the case of the sine–Gordon equation, $F = \sin u$. Taking $B = D = 0, A = C = 1$, we obtain the one-forms (8). Similarly, one gets the corresponding one-forms in the case of sinh–Gordon and Liouville equations.

When F is given by Theorem 6(ii), the functions f_{ij} are determined by (47), (67), (68), and (70). We obtain

$$\omega_1 = f_{11}\,dx + (N/\theta)e^{\theta u}\,dt, \quad \omega_2 = \eta\,dx, \quad \omega_3 = \pm\omega_1,$$

where

$$f_{11,u_x} = (N/\nu\theta)(\theta u_x \pm \eta)e^{-h(u_x)},$$

$N \neq 0, \theta \neq 0, \nu \neq 0$, and the sign $\pm$ is chosen according to the plane $\theta z \pm \eta = 0$, which does not intersect $U \times I$. The one-forms above satisfy (3) if and only if

$$u_{xt} = \nu e^{\theta u + h(u_x)},$$

For F given as in Theorem 6(iii), we obtain f_{ij} from (73), (75), (76), and (47). Choosing $Q = 0$ in (76) and $T = 1$, we have

$$\omega_1 = \pm(\eta/\lambda)u_x\,dx + (u + \tau/\lambda)dt,$$

$$\omega_2 = \eta\,dx + (\lambda/\eta \mp \zeta)dt,$$

$$\omega_3 = \pm\omega_1,$$

where $\lambda \neq 0$, $\eta \neq 0$. These forms satisfy (3) if and only if

$$u_{xt} = \lambda u + \zeta u_x + \tau.$$

Remark 4: In case (ii) of Theorem 6, $z_t = F$ cannot describe an η-p.s.s. for $\eta \in \mathscr{P} = \mathfrak{R} - S$, where S is a set of isolated points. In fact, since the constant θ is nonzero and U is an open set in $\mathfrak{R}^2$, there exists $(u_0, z_0) \in U$ such that $\pm z_0\theta$ belongs to $\mathscr{P}$. For such a point and $\eta_0 = \pm z_0\theta$, we have that (u_0, z_0, η_0) belongs to the intersection of $U \times \mathscr{P}$ and the plane $\theta z \pm \eta = 0$. This is a contradiction.

Finally we prove Theorem 1, which gives a complete characterization of the equations $u_{xt} = F(u, u_x)$, which describes an η-p.s.s. where F is independent of η and $\eta \in \mathscr{P}$.

Proof of Theorem 1: Suppose $z_t = F(u,z)$ describes an η-p.s.s. with associated differentiable functions $f_{ij}(u,z,\eta)$, defined on $U \times \mathscr{P}$. Consider L given by (20). Let

$$W^1 = \{(u,z,\eta) \in U \times \mathscr{P}; \quad L \neq 0\}$$

and $W^0 = (U \times \mathscr{P}) - W^1$.

If int $W^0 \neq \varnothing$, there exists an open connected set $V \times I \subset$ int W^0 with $V \subset U$ and $I \subset \mathscr{P}$. On the set $V \times I$ the conditions of Theorem 6 hold and we have two cases, since, from Remark 4, case (ii) cannot occur. From Theorem 6(i), F satisfies case (i) of Theorem 1 with $V = \mathfrak{R}^2$, and the f_{ij} are defined on $\mathfrak{R}^2 \times (\mathfrak{R} - \{0\}) = \text{int } W^0$; from Theorem 6(iii), F is given on $V = \mathfrak{R}^2$ by

$$F = \lambda u + \zeta z + \tau,$$

where $\lambda \neq 0$ and the functions f_{ij} are defined on $\mathfrak{R}^2 \times (\mathfrak{R} - \{0\}) = \text{int } W^0$. Therefore, using Remark 1 for $\lambda = 0$, we obtain case (iii) of Theorem 1.

If $W^1 \neq \varnothing$, there exists an open connected set $V \times I \subset W^1$, with $V \subset U$ and $I \subset \mathscr{P}$. On the set $V \times I$ the conditions of Theorem 4 hold. Therefore we have two cases: from Theorem 4(ii), we obtain again case (i) of Theorem 1; from Theorem 4(i), F is given on $V = U$ as in Theorem 1(ii), with the functions f_{ij} defined on $U \times \mathfrak{R} = W^1$. This concludes the proof of Theorem 1.

Theorem 2 and Corollary 1 follow immediately from Theorem 1, as mentioned in the Introduction.

ACKNOWLEDGMENT

Both authors were partially supported by CNPq.

[1]R. Sasaki, "Soliton equations and pseudospherical surfaces," Nucl. Phys. B **154**, 343 (1979).

[2]J. A. Cavalcante, and K. Tenenblat, "Conservation laws for nonlinear evolution equations," J. Math. Phys. **29**, 1044 (1988).

[3]S. S. Chern and K. Tenenblat, "Foliations on a surface of constant curvature and the modified Korteweg–de Vries equations," J. Diff. Geom. **16**, 347 (1981).

[4]S. S. Chern and K. Tenenblat, "Pseudo-spherical surfaces and evolution equations," Stud. Appl. Math. **74**, 55 (1986).

[5]L. P. Jorge and K. Tenenblat, "Linear problems associated to evolution equations of type $u_{tt} = F(u, u_x, u_{xx}, u_t)$," Stud. Appl. Math. 77, 103 (1987).

[6]K. Tenenblat, "Backlund's theorem for submanifolds of space forms on a generalized wave equation," Bol. Soc. Bras. Mat. **16**, 65 (1985).

[7]K. Tenenblat and C. L. Terng, "Backlund's theorem for n-dimensional submanifolds of R^{2n-1}," Ann. Math. **111**, 477 (1980).

[8]C. L. Terng, "A higher dimension generalization of the sine–Gordon equation and its soliton theory," Ann. Math. **111**, 491 (1980).

[9]M. J. Ablowitz, R. Beals, and K. Tenenblat, "On the solution of the generalized wave and generalized sine–Gordon equations," Stud. Appl. Math. **74**, 177 (1986).

[10]R. Beals and K. Tenenblat, "Inverse scattering and the Backlund transformation for the generalized wave and generalized sine–Gordon equations," Stud. Appl. Math. **78**, 227 (1988).

[11]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, "The inverse scattering transform—Fourier analysis for nonlinear problems," Stud. Appl. Math. **53**, 249 (1974).

[12]M. L. Rabelo, "Characterization of differential equations of type $u_{xt} = F(u, \partial u/\partial x, \ldots, \partial^k u/\partial x^k)$, which describe the pseudo-spherical surfaces," An. Acad. Bras. Cien. **60**, 119 (1988).

[13]R. K. Dodd and R. Bullough, "Backlund transformations for the sine–Gordon equations," Proc. R. Soc. London Ser. A **351**, 499 (1976).

[14]D. W. McLaughtin, and A. C. Scott, "A restricted Backlund transformation," J. Math. Phys. **14**, 1817 (1973).

[15]R. Beals and R. R. Coifman, "Inverse scattering and evolution equations," Comm. Pure Appl. Math. **88**, 29 (1985).

[16]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, "Method for solving the Korteweg–de Vries equation," Phys. Rev. Lett. **19**, 1095 (1967).

[17]E. Cartan, *Les Systèmes Différentielles Extérieurs et Leurs Applications Géometriques* (Herman, Paris, 1945).

Stochastic mechanics of a relativistic spinless particle

Gian Fabrizio De Angelis
Dipartimento di Matematica, Università di Roma I, Piazzale Aldo Moro, 2, 00185 Rome, Italy

(Received 29 November 1988; accepted for publication 24 January 1990)

An extension of Nelson's stochastic mechanics to the relativistic domain is proposed. To each pure state of a spinless relativistic quantum particle corresponds a Markov process $t \mapsto \xi_t$, where the random variable ξ_t represents, at every time t, the space position of the particle in the sense of Newton and Wigner. The process $t \mapsto \xi_t$ is not a diffusion but the usual Nelson's theory is restored in the nonrelativistic limit.

I. INTRODUCTION

Stochastic mechanics of relativistic spinless particles is not a novelty in literature. Besides a few considerations in Caubet's book[1] there exists two interesting related papers[2,3] on a possible probabilistic scenario for the Klein–Gordon equation. Both articles are inspired by Feynman's path integral approach to the Klein–Gordon propagator[4] and make use of diffusions $\tau \mapsto x^\mu_\tau$, $\mu = 0,...,D$ in space-time, where τ must be interpreted as some kind of proper time. As the path $\tau \mapsto x^\mu_\tau$ wanders over Minkowski space it crosses many times any spacelike hypersurface $x^0 = ct$ creating a cloud of points, an observer watching space at time t perceives cross points as particles if the path goes forward in time and as antiparticles when it runs backward. By constructing the diffusions $\tau \mapsto x^\mu_\tau$ from classical relativistic mechanics revistited according to some version of Nelson's bible,[5–7] it is possible to reconstruct, from each of them, complex solutions of the Klein–Gordon equation whose electric current $J^\mu(x^0,x^1,...,x^D)$ gives the correct average flux of charges across hypersurfaces. The many particles picture related to Feynman's path suggests some link between this stochastic treatment of the Klein–Gordon equation and the second quantized version of the theory. My approach is different. I consider only positive frequency solutions of the Klein–Gordon equation and I associate to each of them a Markov process $t \mapsto \xi_t$ in space with a single particle meaning: The random variable ξ_t represents the space position of the particle at time t in the well-known sense of Newton and Wigner.[8,9] Of course this stochastic framework is not manifestely covariant as it is based on the noncovariant concept of localization in space. It happens that the processes $t \mapsto \xi_t$ are jump Markov processes and not diffusions. This fact deserves some physical explanation. Picturesquely one can identify the single physical particle with the center of mass of the cloud generated by a Feynman's path and it is quite clear that this center of mass undergoes random jumps when "particles" and "antiparticles" annihilate in pairs inside the cloud. The jump character of processes is a relativistic feature and it disappears in the nonrelativistic limit.

II. MATHEMATICAL FACTS

For any $M > 0$, the function

$$L(\cdot):\ \mathbf{p} \in \mathbb{R}^D \mapsto L(\mathbf{p}) = (Mc^2/\hbar)\left(1 - \sqrt{(\hbar^2/M^2c^2)\|\mathbf{p}\|^2 + 1}\right)$$

is conditionally positive definite and therefore the Levy–Khintchine formula[10] holds:

$$L(\mathbf{p}) = \int_{\mathbb{R}^D}\left(\exp i\mathbf{p}\cdot\mathbf{y} - 1 - \frac{i\mathbf{p}\cdot\mathbf{y}}{1 + \|\mathbf{y}\|^2}\right)\nu(d\mathbf{y}), \qquad (1)$$

where $\nu(d\mathbf{y})$ is a Levy measure invariant under orthogonal transformations. The measure $\nu(d\mathbf{y})$ is not finite but $\nu(\{\mathbf{y}:\ \|\mathbf{y}\| > r\}) < +\infty$ for any $r > 0$. For future needs, I call $\nu_r(d\mathbf{y})$ the finite measure in $\mathbb{R}^D$ given by $\nu_r(d\mathbf{y}) = \chi_{\{\mathbf{y}:\|\mathbf{y}\|>r\}}(\cdot)\nu(d\mathbf{y})$, where $\chi_B(\cdot)$ is the characteristic function of the subset B. The Levy–Khintchine formula (1) allows me to obtain some control on the pseudodifferential operator

$$\mathbb{L} = L(-i\nabla) = (Mc^2/\hbar)\left(\mathbb{I} - \sqrt{-(\hbar^2/M^2c^2)\Delta + \mathbb{I}}\right), \qquad (2)$$

as it is clear, from (1), that

$$(\mathbb{L}f)(\mathbf{x}) = \int_{\mathbb{R}^D}\left(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - \frac{\mathbf{y}\cdot\nabla f(\mathbf{x})}{1 + \|\mathbf{y}\|^2}\right)\nu(d\mathbf{y}), \qquad (3)$$

for instance if $f(\cdot):\ \mathbb{R}^D \to \mathbb{C}$ is a bounded C^2 function. If M is interpreted as the mass of a relativistic particle, then (3) gives a useful representation of the free relativistic quantum Hamiltonian $H_0 = \sqrt{-\hbar^2c^2\Delta + M^2c^4\mathbf{I}}$, which is related to $\mathbb{L}$ by

$$H_0 = -\hbar\mathbb{L} + Mc^2\mathbb{I}. \qquad (4)$$

It is interesting to observe that $\mathbb{L}$ is the generator or infinitesimal operator of a time homogeneous Markovian family whose transition functions are obtained from the "heat equation" $\partial u/\partial t = \mathbb{L}u$.

This fact has been exploited by Ichinose and Tamura[11] in constructing a Feynman–Kac formula for Hamiltonians $H = H_0 + V(\cdot)$ for a suitable class of potentials $V(\cdot)$. Any Markov process with generator $\mathbb{L}$ has the form

$$\xi_t = \xi_0 + \int_{\mathbb{R}^D\times[0,t]} \mathbf{y}\widetilde{N}(d\mathbf{y}\,ds), \qquad (5)$$

where $\xi_0 \in \mathbb{R}^D$, $N(d\mathbf{y}\,ds)$ is the Poisson measure in $\mathbb{R}^D \times [0,+\infty)$ with average $\mathbf{E}(N(d\mathbf{y}\,ds)) = \nu(d\mathbf{y})ds$, and $\widetilde{N}(d\mathbf{y}\,ds)$ is the martingale $N(d\mathbf{y}\,ds) - \nu(d\mathbf{y})ds$. More generally, if $(t,\mathbf{x},\mathbf{y}) \mapsto \gamma(t,\mathbf{x},\mathbf{y})$ is a bounded non-negative function continous in $(t,\mathbf{x})$, sufficiently smooth in $\mathbf{y}$, and such that $\gamma(t,\mathbf{x},0) = 1$, operators $\mathbb{L}_t$ of the form

 0022-2488/90/061408-05$03.00

$$(\mathbb{L}_t f)(\mathbf{x}) = \int_{\mathbb{R}^D} \Big(\gamma(t,\mathbf{x},\mathbf{y})(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x})) - \frac{\mathbf{y}\cdot\nabla f(\mathbf{x})}{1+\|\mathbf{y}\|^2}\Big)\nu(d\mathbf{y})$$
$$= (\mathbb{L}_f)(\mathbf{x}) + \int_{\mathbb{R}^D} (\gamma(t,\mathbf{x},\mathbf{y}) - 1) \times (f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}))\nu(d\mathbf{y}) \qquad (6)$$

are infinitesimal operators of Markov processes obtained as solutions of the stochastic differential equation

$$d\,\xi_t = \int_{\mathbb{R}^D}\Big(\gamma(t,\xi_t,\mathbf{y}) - \frac{1}{1+\|\mathbf{y}\|^2}\Big)\mathbf{y}\nu(d\mathbf{y})dt + \int_{\mathbb{R}^D} \mathbf{y}\widetilde{N}(\gamma(t,\xi_t,\mathbf{y})d\mathbf{y}\,dt). \qquad (7)$$

The meaning of $\gamma(t,\mathbf{x},\mathbf{y})$ is clear: It represents a change of the rate of jumping which depends on the time t, the point $\mathbf{x}$ reached at time t, and the jump amplitude $\mathbf{y}$. It is better dealing with finite Levy measures, so I will give an elementary useful lemma.

Lemma: If $f(\cdot):\mathbb{R}^D\to\mathbb{C}$ is a bounded C^2 function, then

$$(\mathbb{L}f)(\mathbf{x}) = \lim_{r\to 0}\int_{\mathbb{R}^D}(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}))\nu_r(d\mathbf{y}) = \lim_{r\to 0}(\mathbb{L}_r f)(\mathbf{x}).$$

Proof: It is sufficient to exploit the identity

$$\int_{\mathbb{R}^D}(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}))\nu_r(d\mathbf{y}) = \int_{\mathbb{R}^D}\Big(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - \frac{\mathbf{y}\cdot\nabla f(\mathbf{x})}{1+\|\mathbf{y}\|^2}\Big)\nu_r(d\mathbf{y})$$

and Lebesque's theorem on dominated convergence as $f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - \mathbf{y}\cdot\nabla f(\mathbf{x})/(1+\|y\|^2)$ is $O(\|\mathbf{y}\|^2)$ in the neighborhood of $\mathbf{y}=\mathbf{0}$, hence $\nu(d\mathbf{y})$ integrable.

The operator $\mathbb{L}_r$ is the generator of a regular jump Markov process[12] with jump probability per unit time given by

$$q(t,\mathbf{x},B) = \lim_{s\to t+}\frac{P(t,\mathbf{x},s,B) - \chi_B(\mathbf{x})}{s-t} = \int_{\mathbb{R}^D}(\chi_B(\mathbf{x}+\mathbf{y}) - \chi_B(\mathbf{x}))\nu_r(d\mathbf{y}).$$

By the way, regular jump processes with

$$q(t,\mathbf{x},B) = \int_{\mathbb{R}^D}\gamma(t,\mathbf{x},\mathbf{y})(\chi_B(\mathbf{x}+\mathbf{y}) - \chi_B(\mathbf{x}))\nu_r(d\mathbf{y})$$

have generators $\mathbb{L}_t'$ of the form

$$(\mathbb{L}_t' f)(\mathbf{x}) = \int_{\mathbb{R}^D}\gamma(t,\mathbf{x},\mathbf{y})(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}))\nu_r(d\mathbf{y}),$$

and vice versa. After this long discussion I will come back to quantum mechanics.

III. RELATIVISTIC SCHRÖDINGER EQUATION

Let me consider now the relativistic Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = H\psi, \qquad (8)$$

where $H = H_0 + V(\cdot)$ for some potential $V(\cdot):\mathbb{R}^D\to\mathbb{R}$.

From the standpoint of stochastic mechanics, the following problem is quite natural: given a normalized solution $t\in\mathbb{R}\mapsto\psi(t,\cdot)\in L^2(\mathbb{R}^D)$ of (8), find a Markov process $t\mapsto\xi_t$ in $\mathbb{R}^D$ such that

$$\text{Prob}(\xi_t\in B) = \int_{B^D}|\psi(t,\mathbf{x})|^2\,d^D\mathbf{x} \qquad (9)$$

at every time t and for each Borel subset B of $\mathbb{R}^D$. The trouble with the relativistic Hamiltonian

$$H = \sqrt{-\hbar^2c^2\Delta + M^2c^4\mathbb{I}} + V(\cdot)$$

lies in the quantum mechanical continuity equation for $\rho(t,\mathbf{x}) = |\psi(t,\mathbf{x})|^2$, namely,

$$\frac{\partial\rho}{\partial t} = i(\bar\psi\,\mathbb{L}\psi - \psi\,\mathbb{L}\,\bar\psi) = 2\Im(\psi\,\mathbb{L}\,\bar\psi), \qquad (10)$$

because $\mathbb{L}$ is not $(\hbar/2M)\Delta$ as in the nonrelativistic case but the pseudodifferential operator

$$(Mc^2/\hbar)(\mathbb{I} - \sqrt{-(\hbar^2/M^2c^2)\Delta + \mathbb{I}}).$$

Now I follow the general strategy introduced in Ref. 13 and I try to find a forward Kolmogorov equation that is satisfied by the quantum mechanical probability density $\rho(t,\mathbf{x}) = |\psi(t,\mathbf{x})|^2$. In order to simplify the search for such a good Kolmogorov equation, I replace, for the moment, the operator $\mathbb{L}$ with the better looking $\mathbb{L}_r$. That leaves me with the (approximate) continuity equation $\partial\rho/\partial t = 2\Im(\psi\mathbb{L}_r\bar\psi)$ and, if I define $m_t(B)$ as $\int_B\rho(t,\mathbf{x})d^D\mathbf{x}$, I obtain the following chain of equalities:

$$\frac{dm_t(B)}{dt} = 2\int_{\mathbb{R}^D}\chi_B(\mathbf{x})\Im(\psi\mathbb{L}_r\bar\psi)d^D\mathbf{x}$$
$$= 2\int_{\mathbb{R}^D\times\mathbb{R}^D}\chi_B(\mathbf{x})\Im(\psi(t,\mathbf{x})\bar\psi(t,\mathbf{x}+\mathbf{y}))\times d^D\mathbf{x}\nu_r(d\mathbf{y})$$

[if $\mathbf{x}\mapsto\psi(t,\mathbf{x})$ nowhere vanishes at every time t]

$$= \Big(\int_{\mathbb{R}^D\times\mathbb{R}^D}\chi_B(\mathbf{x})|\psi(t,\mathbf{x}+\mathbf{y})|^2\Im\frac{\psi(t,\mathbf{x})}{\psi(t,\mathbf{x}+\mathbf{y})}\times d^D\mathbf{x}\,\nu_r(d\mathbf{y}) - \int_{\mathbb{R}^D\times\mathbb{R}^D}\chi_B(\mathbf{x})|\psi(t,\mathbf{x})|^2\Im\times[\psi(t,\mathbf{x}+\mathbf{y})/\psi(t,\mathbf{x})]d^D\mathbf{x}\,\nu_r(d\mathbf{y})\Big)$$

[by translational invariance of $d^D\mathbf{x}$ and reflection invariance of $\nu_r(d\mathbf{y})$]

$$= \int_{\mathbb{R}^D}|\psi(t,\mathbf{x})|^2\,d^D\mathbf{x}\int_{\mathbb{R}^D}\Im\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}\times(\chi_B(\mathbf{x}+\mathbf{y}) - \chi_B(\mathbf{x}))\nu_r(d\mathbf{y}).$$

In this form the (approximate) continuity equation looks almost Kolmogorov for a regular jump process[12] except for a little detail: Since $\Im[\psi(t,x+y)/\psi(t,x)]$ is not necessarily positive, nobody can guarantee that

$$q(t,\mathbf{x},B) = \int_{\mathbb{R}^D}\Im\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}(\chi_B(\mathbf{x}+\mathbf{y}) - \chi_B(\mathbf{x}))\times\nu_r(d\mathbf{y})$$

is a true jump probability per unit time. I notice, however, the identity

$$\int_{\mathbb{R}^D\times\mathbb{R}^D} |\psi(t,\mathbf{x}+\mathbf{y})\psi(t,\mathbf{x})|(\chi_B(\mathbf{x}+\mathbf{y}) - \chi_B(\mathbf{x}))d^D\mathbf{x}\,\nu_r(d\mathbf{y}) = 0,$$

and this fact allows me to rewrite the (approximate) continuity equation as

$$\frac{dm_t(B)}{dt} = \int_{\mathbb{R}^D} q_r(t,\mathbf{x},B)m_t(d^D\mathbf{x}),$$

where now

$$q_r(t,\mathbf{x},B) = \int_{\mathbb{R}^D}\left(\left|\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}\right| + \Im\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}\right) \times(\chi_B(\mathbf{x}+\mathbf{y}) - \chi_B(\mathbf{x}))\nu_r(d\mathbf{y}), \tag{11}$$

which is the jump probability per unit time of a regular jump Markov process whose infinitesimal operators are given by

$$(\mathbb{L}_t^r f)(\mathbf{x}) = \int_{\mathbb{R}^D} \gamma(t,\mathbf{x},\mathbf{y})(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}))\nu_r(d\mathbf{y}), \tag{12}$$

where

$$\gamma(t,\mathbf{x},\mathbf{y}) = \left|\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}\right| + \Im\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})} \geqslant 0. \tag{13}$$

In order to get a true jump probability, a similar device has been used before.[13,14] If $\mathbf{x}\mapsto\psi(t,\mathbf{x})$ is bounded and sufficiently smooth, there exists

$$\lim_{r\to 0}(\mathbb{L}_t^r f)(\cdot)$$

for each bounded C^2 function $f(\cdot)$ and this limit is given by

$$(\mathbb{L}_t f)(\mathbf{x}) = \int_{\mathbb{R}^D}\left(\gamma(t,\mathbf{x},\mathbf{y})(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x})) - \frac{\mathbf{y}\cdot\nabla f(\mathbf{x})}{1+\|\mathbf{y}\|^2}\right)\nu(d\mathbf{y})$$
$$= (\mathbb{L}f)(\mathbf{x}) + \int_{\mathbb{R}^D}(\gamma(t,\mathbf{x},\mathbf{y}) - 1)(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}))\nu(d\mathbf{y}). \tag{14}$$

Now the true quantum mechanical continuity equation contains $\mathbb{L}$ and not $\mathbb{L}_r$ but, recalling my lemma and by taking care of all limits involved, I obtain, eventually, the forward Kolmogorov equation for ρ.

Theorem: If $t\mapsto\psi(t,\cdot)\in L^2(\mathbb{R}^D)$ is a solution of (8) bounded, sufficiently smooth, and nowhere vanishing for every t, then the quantum mechanical probability density $\rho(t,\mathbf{x}) = |\psi(t,\mathbf{x})|^2$ obeys the forward Kolmogorov equation

$$\frac{\partial\rho}{\partial t} = \mathbb{L}_t^*\rho, \tag{15}$$

where $\mathbb{L}_t^*$ is the adjoint of the operator $\mathbb{L}_t$ given by (14).

Therefore there exists a jump Markov process $t\mapsto\xi_t$ in $\mathbb{R}^D$ such that (9) holds. The process $t\mapsto\xi_t$ is obtained as a solution of the stochastic differential equation (7) where $\gamma(t,\mathbf{x},\mathbf{y})$ is related to the wave function by (13) and the random variable ξ_0 is distributed according to $\rho(0,\cdot)d^D\mathbf{x} = |\psi(0,\cdot)|^2 d^D\mathbf{x}$. I end this section with three remarks.

Remark 1: Hypotheses of smoothness and absence of zeros for the wave function were made by Nelson in his original treatment of Schrödinger's equation. These hypotheses were relaxed considerably by Carlen[15] and it would be interesting to perform something similar in the present context.

Remark 2: Suppose that $H = H_0 + V(\cdot)$ has a ground state $\Omega(\cdot)>0$ with energy E_0. Under transition to ground state representation: $\psi(\cdot)\in L^2(\mathbb{R}^D)\mapsto f(\cdot) = \psi(\cdot)/\Omega(\cdot)\in L^2(\mathbb{R}^D,\Omega^2(\cdot)d^D\mathbf{x})$ the operator $(1/\hbar)(E_0\mathbb{I} - H)$ becomes $\mathbb{L}_\Omega$, where

$$(\mathbb{L}_\Omega f)(\mathbf{x}) = \int_{\mathbb{R}^D}\left(\frac{\Omega(\mathbf{x}+\mathbf{y})}{\Omega(\mathbf{x})}(f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x})) - \frac{\mathbf{y}\cdot\nabla f(\mathbf{x})}{1+\|\mathbf{y}\|^2}\right)\nu(d\mathbf{y}),$$

which is, precisely, the generator of ground state process. So the semigroup $t\mapsto\exp-(t/\hbar)(H - E_0\mathbb{I})$ is unitarily equivalent to the Markovian semigroup $t\mapsto\exp t\mathbb{L}_\Omega$. Moreover,

$$\langle f(\cdot),\mathbb{L}_\Omega f(\cdot)\rangle_{L^2(\mathbb{R}^D,\Omega^2(\cdot)d^Dx)} = -\frac{1}{2}\int_{\mathbb{R}^D\times\mathbb{R}^D}\Omega(\mathbf{x}+\mathbf{y})\Omega(\mathbf{x})|f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x})|^2 d^D\mathbf{x}\,\nu(d\mathbf{y}).$$

This expression reminds us of the theory of Dirichlet forms.[16,17]

Remark 3: It would be perfectly possible to define $\gamma(t,\mathbf{x},\mathbf{y})$ in (13) as

$$\gamma(t,\mathbf{x},\mathbf{y}) = k\left|\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}\right| + \Im\frac{\psi(t,\mathbf{x}+\mathbf{y})}{\psi(t,\mathbf{x})}$$

provided that $k\geqslant 1$. I want to justify the choice which I made, namely $k=1$. By observing that $L(\mathbf{p}) = \int_{\mathbb{R}^D}(\cos\mathbf{p}\cdot\mathbf{y} - 1)\nu(d\mathbf{y})$ and $\lim_{c\to+\infty}L(\mathbf{p}) = -(\hbar/2M)\|\mathbf{p}\|^2$ it is clear that the measure $\nu(d\mathbf{y})$ concentrates its mass around $\mathbf{y}=\mathbf{0}$ as c increases and that

$$\lim_{c\to+\infty}\int_{\{y:\,\|y\|<R\}} y_i y_j\nu(d\mathbf{y}) = \frac{\hbar}{M}\delta_{ij},$$

for any $R>0$. If $S(t,\mathbf{x})$ and $R(t,\mathbf{x})$ are defined by $\psi(t,\mathbf{x}) = \exp(R(t,\mathbf{x}) + iS(t,\mathbf{x}))$ and I make the choice $k=1$, then $\gamma(t,\mathbf{x},\mathbf{y}) = 1 + (M/\hbar)\mathbf{y}\cdot\mathbf{b}(t,\mathbf{x}) + O(\|\mathbf{y}\|^2)$ in the neighborhood of $\mathbf{y}=\mathbf{0}$, where $\mathbf{b}(t,\mathbf{x}) = (\hbar/M)(\nabla R + \nabla S)$. It is clear, now, (at least formally) that

$$\lim_{c\to+\infty}(\mathbb{L}_t f)(\cdot) = (\hbar/2M)(\Delta f)(\cdot) + \mathbf{b}\cdot\nabla f)(\cdot),$$

which is exactly the generator of the diffusion associated to $(t,\mathbf{x})\mapsto\psi(t,\mathbf{x})$ by Nelson's theory.

IV. KLEIN–GORDON EQUATION

Let $(t,\mathbf{x})\mapsto\varphi(t,\mathbf{x})$ be a normalized positive frequency solution of the Klein–Gordon equation

$$\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} - \Delta\varphi + \frac{M^2c^2}{\hbar^2}\varphi = 0. \tag{16}$$

This means that

$$\varphi(t,\mathbf{x}) = (2\pi)^{-D/2}\int_{\mathbf{R}^D} e^{i(\mathbf{p}\cdot\mathbf{x}-ct\omega(\mathbf{p}))}\widetilde{\varphi}(\mathbf{p})\frac{d^D\mathbf{p}}{\omega(\mathbf{p})}$$

$$= (2\pi)^{-D/2}\int_{\mathbf{R}^D} e^{i\mathbf{p}\cdot\mathbf{x}}\widetilde{\varphi}_t(\mathbf{p})\frac{d^D\mathbf{p}}{\omega(\mathbf{p})}, \qquad (17)$$

where

$$\omega(\mathbf{p}) = \sqrt{\|\mathbf{p}\|^2 + \frac{M^2c^2}{\hbar^2}}$$

and

$$\widetilde{\varphi}(\cdot)\in\mathscr{H} = L^2\left(\mathbf{R}^D, \frac{d^D\mathbf{p}}{\omega(\mathbf{p})}\right)$$

with $\|\widetilde{\varphi}(\cdot)\|_{\mathscr{H}} = 1$. The Hilbert space $\mathscr{H}$ describes precisely the pure states of a relativistic spinless particle of mass M and the fact that $\mathscr{H}$ carries an (irreducible) unitary representation of the Poincaré group that gives us the self-adjoint operators representing physical observables as energy, momentum, and angular momentum. What about the components of space position of the particle in a given inertial frame? The right answer was given many years ago by Newton and Wigner.[8,9] Under a few mild regularity assumptions, there exists only one choice of D mutually commuting self-adjoint operators q_α that transform in the proper manner under orthogonal group and space displacements and they are given by

$$(q_\alpha\widetilde{\varphi})(\mathbf{p}) = i\frac{\partial\widetilde{\varphi}(\mathbf{p})}{\partial p_\alpha} - i\frac{p_\alpha}{\|\mathbf{p}\|^2 + M^2c^2/\hbar^2}\widetilde{\varphi}(\mathbf{p})$$

$$= \sqrt{\omega(\mathbf{p})}\,i\frac{\partial}{\partial p_\alpha}\frac{\widetilde{\varphi}(\mathbf{p})}{\sqrt{\omega(\mathbf{p})}}. \qquad (18)$$

Let $B\subseteq\mathbf{R}^D\mapsto E(B)$ be the joint spectral measure of q_α's and let $\widetilde{\varphi}(\cdot)$ be a normalized vector of $\mathscr{H}$. The quantum mechanical probability of finding the particle localized inside B when the state is $\widetilde{\varphi}(\cdot)$, is given by

$$\langle\widetilde{\varphi}(\cdot), E(B)\widetilde{\varphi}(\cdot)\rangle_{\mathscr{H}} = \int_B |\psi(\mathbf{x})|^2\, d^D\mathbf{x},$$

where $\psi(\cdot)$ is related to $\widetilde{\varphi}(\cdot)$ by

$$\psi(\mathbf{x}) = (2\pi)^{-D/2}\int_{\mathbf{R}^D}\frac{\widetilde{\varphi}(\mathbf{p})}{\sqrt{\omega(\mathbf{p})}}e^{i\mathbf{p}\cdot\mathbf{x}}\, d^D\mathbf{p}, \qquad (19)$$

a well-defined Fourier–Plancherel transformation as $\widetilde{\varphi}(\cdot)/\sqrt{\omega(\cdot)}\in L^2(\mathbf{R}^D)$. Of course $\psi(\cdot)$ is the wave function of the particle in the representation that "diagonalizes" all q_α's. Coming back to our positive frequency solution $(t,\mathbf{x})\mapsto\varphi(t,\mathbf{x})$ of Klein–Gordon equation, we see that the probability of finding the particle inside $B\subseteq\mathbf{R}^D$ at time t is $\int_B|\psi(t,\mathbf{x})|^2\, d^D\mathbf{x}$, where

$$\psi(t,\mathbf{x}) = \left(\left(-\Delta + \frac{M^2c^2}{\hbar^2}\right)^{1/4}\varphi\right)(t,\mathbf{x}). \qquad (20)$$

Obviously, $t\mapsto\psi(t,\cdot)$ obeys the relativistic Schrödinger equation (8) with $H = H_0$ (and I can apply the stochastic description of Sec. III). The conclusion is the following: if $(t,\mathbf{x})\mapsto\varphi(t,\mathbf{x})$ is a normalized positive frequency solution of the Klein–Gordon equation, there exists a jump Markov process $t\mapsto\xi_t$ in $\mathbf{R}^D$ such that

$$\text{Prob}(\xi_t\in B) = \int_B\left|\left(\left(-\Delta + \frac{M^2c^2}{\hbar^2}\right)^{1/4}\varphi\right)(t,\mathbf{x})\right|^2 d^D\mathbf{x}$$

at every time t and for each Borel subset B of $\mathbf{R}^D$, provided that $\mathbf{x}\mapsto((-\Delta + M^2c^2/\hbar^2)^{1/4}\varphi)(t,\mathbf{x})$ is sufficiently smooth, bounded, and nowhere vanishing at each time t. By looking at the stochastic differential equation for the process $t\mapsto\xi_t$ we can imagine that the praticle travels in space guided by a smooth velocity filed on which are superimposed, at random times, jumps of random magnitude. In the nonrelativistic limit, the probability of jumping concentrates more and more on vanishing jump amplitudes and the process will approach to a diffusion. The Newton–Wigner theory of localization in space for elementary systems applies to all spins and therefore the present treatment can be extended to all relativistic wave equation. In particular, it can be extended to Dirac equation that, after a Foldy–Wouthysen transformation, becomes a relativistic Schrödinger equation for a multicomponent wave function. Of course other approaches are feasible.[18]

V. CONCLUSIONS AND OUTLOOK

The choice of space localization á la Newton and Wigner is not arbitrary. In a recent paper[19] Blanchard, Carlen, and Dell'Antonio analyzed the configurations of the free scalar quantum field at fixed time. Since quantum fields operators commute at fixed time, expectations of their products on arbitrary quantum states can be interpreted as correlation functions of some (state dependent) space random field. The typical configuration of this random field is rather rough but, after a suitable filtering procedure that eliminates vacuum fluctuations, a smooth field configuration comes out and it shows a bump near $\mathbf{x}$ for quantum states describing a particle localized near $\mathbf{x}$ in the sense of Newton and Wigner. It is clear that a full stochastic treatment of the Klein–Gordon equation requires an infinite-dimensional version of Nelson's theory namely a stochastic field theory in space-time. Starting from a single particle state of the quantum field it will then be possible, by the filtering procedure of Blanchard, Carlen, and Dell'Antonio, to reconstruct a Markov process for the space position of the particle. When someone will do that, I can reasonably bet that this Markov process will not be a diffusion but exactly the kind of stochastic process which I just discussed.

[1] J. P. Caubet, *Le Mouvement Brownien Relativiste*, Lectures Notes in Mathematics, Vol. 559 (Springer, Berlin, 1976).

[2] F. Guerra and P. Ruggiero, "A note on relativistic Markov processes," Lett. Nuovo Cimento 23, 529 (1978).

[3] M. Serva, "Relativistic stochastic processes associated to Klein–Gordon equation," Ann. Inst. Henri Poincaré 49, 415 (1988).

[4] R. P. Feynman, "Space-time approach to nonrelativistic quantum mechanics," Rev. Mod. Phys. 20, 367 (1948).

[5] E. Nelson, "Derivation of the Schrödinger equation from Newtonian mechanics," Phys. Rev. 150, 1079 (1966).

[6] E. Nelson, *Dynamical Theories of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967).

[7] E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1985).

[8] T. D. Newton and E. P. Wigner, "Localized states for elementary systems," Rev. Mod. Phys. 21, 400 (1949).

[9]A. S. Wightman, "On the localizability of quantum mechanical systems," Rev. Mod. Phys. **34**, 845 (1962).
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. IV, Analysis of Operators (Academic, New York, 1978).
[11]T. Ichinose and H. Tamura, "Imaginary-time path integral for a relativistic spinless particle in an electromagnetic field," Commun. Math. Phys. **105**, 239 (1986).
[12]I. I. Gihman and A. V. Skorohod, *Introduction to the Theory of Random Processes* (Saunders, Philadelphia, 1969).
[13]G. F. De Angelis and G. Jona-Lasinio, "A stochastic description of a spin 1/2 particle in a magnetic field," J. Phys. A: Math. Gen. **15**, 2053 (1982).
[14]Ph. Blanchard and M. Sirugue, "Large deviations form classical paths. Hamiltonian flows as classical limits of quantum flows," Commun. Math. Phys. **101**, 173 (1985).
[15]E. A. Carlen, "Conservative diffusions," Commun. Math. Phys. **94**, 293 (1984).
[16]S. Albeverio, R. Høegh-Krohn, and L. Streit," Energy forms, Hamiltonians, and distorted Brownian paths," J. Math Phys. **18**, 907 (1977).
[17]M. Fukushima, *Dirichlet Forms and Markov Processes* (North-Holland, New York, 1980).
[18]G. F. De Angelis, G. Jona-Lasinio, M. Serva, and N. Zanghi, "Stochastic mechanics of a Dirac particle in two space time dimensions," J. Phys. A: Math Gen. **19**, 865 (1986).
[19]Ph. Blanchard, E. A. Carlen, and G. F. Dell'Antonio, "Particles and bumps in quantum field configurations," J. Stat. Phys. **55**, 769 (1989).

Rotations, squeezing, and the unitary transformation operator from individual particles to Jacobi variables

F. C. Khanna
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta T6G 2J1, Canada and TRIUMF, 4004 Wesbrook Mall, Vancouver, British Columbia V6T 2A3, Canada

A. Mann[a] and Y. Takahashi
Theoretical Physics Institute, University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 6 October 1989; accepted for publication 31 January 1990)

A unitary operator for the transformation from individual particles to Jacobi variables is constructed explicitly for particles of arbitrary masses. It is expressed as a product of rotation and squeezing operators using only canonical variables.

I. INTRODUCTION

It is often necessary to have an explicit form for the unitary operator effecting the transformation between the canonical variables of individual particles to those of the center of mass (cm) and the relative coordinates, and, more generally, to those of the Jacobi variables. Since such an operator for the general case of arbitrary single-particle masses does not seem to be available in the literature, a method is developed to construct such an operator. The special case of equal mass particles with specific dynamical assumptions has been the subject of a recent paper.[1] The generalized form reported here could be useful in several physical situations. Furthermore, the unitary operator is independent of any dynamical assumptions about the system of particles.

In the following section we exhibit the construction of the unitary operator transforming the coordinates and momenta of two particles with mass m_1 and m_2, into those of the cm and relative motion variables. In Sec. III we show how to generalize it for N particles, treating explicitly the three-particle case. It should be stressed that only the canonical variables are used and the treatment depends neither on the single-particle Hamiltonians and two-particle interactions, nor on their wave functions. Finally, in Sec. IV we conclude by mentioning the use of these transformations for harmonic oscillator states for the two-particle case.

II. UNITARY TRANSFORMATION: THE TWO-PARTICLE CASE

Let the individual two particles have masses m_1, m_2, position operators $\mathbf{x}_1$, $\mathbf{x}_2$, and momentum operators $\mathbf{p}_1$, $\mathbf{p}_2$, respectively. We define the reduced mass parameters as

$$\mu_1=m_1/(m_1+m_2), \quad \mu_2=m_2/(m_1+m_2). \tag{2.1}$$

The cm and relative coordinate dynamical variables are given by the following well-known formulas;

$$\mathbf{X}_{\rm cm}=\mu_1\mathbf{x}_1+\mu_2\mathbf{x}_2, \tag{2.2a}$$

$$\mathbf{P}_{\rm cm}=\mathbf{p}_1+\mathbf{p}_2, \tag{2.2b}$$

$$\mathbf{x}_r=\mathbf{x}_2-\mathbf{x}_1, \tag{2.2c}$$

$$\mathbf{p}_r=\mu_1\mathbf{p}_2-\mu_2\mathbf{p}_1. \tag{2.2d}$$

The characteristics of the unitary transformation (U) between the individual dynamical variables to those of the cm and relative motion are best exhibited by separating the transformation into three parts.

Part 1: Define an angle α by

$$\cos\alpha=\sqrt{\mu_1}, \quad \sin\alpha=\sqrt{\mu_2}, \tag{2.3}$$

and a set of intermediate variables between the two systems:

$$\mathbf{X}'=\sqrt{\mu_1}\mathbf{x}_1+\sqrt{\mu_2}\mathbf{x}_2=\mathbf{x}_1\cos\alpha+\mathbf{x}_2\sin\alpha, \tag{2.4a}$$

$$\mathbf{x}'=-\sqrt{\mu_2}\mathbf{x}_1+\sqrt{\mu_1}\mathbf{x}_2=-\mathbf{x}_1\sin\alpha+\mathbf{x}_2\cos\alpha, \tag{2.4b}$$

$$\mathbf{P}'=\sqrt{\mu_1}\mathbf{p}_1+\sqrt{\mu_2}\mathbf{p}_2=\mathbf{p}_1\cos\alpha+\mathbf{p}_2\sin\alpha, \tag{2.5a}$$

$$\mathbf{p}'=-\sqrt{\mu_2}\mathbf{p}_1+\sqrt{\mu_1}\mathbf{p}_2=-\mathbf{p}_1\sin\alpha+\mathbf{p}_2\cos\alpha. \tag{2.5b}$$

Obviously the connection between the coordinates $(\mathbf{X}',\mathbf{x}')$ and $(\mathbf{x}_1,\mathbf{x}_2)$ and between the momenta $(\mathbf{P}',\mathbf{p}')$ and $(\mathbf{p}_1,\mathbf{p}_2)$ is given by a rotation through an angle α in the "(1, 2) plane." This is generated by the operator

$$L=\mathbf{x}_1\cdot\mathbf{p}_2-\mathbf{x}_2\cdot\mathbf{p}_1. \tag{2.6}$$

The finite rotation is given by the unitary operator

$$R_1=e^{-i\alpha L}, \tag{2.7}$$

and therefore we have in an obvious notation

$$R_1\{\mathbf{x}_1, \mathbf{x}_2; \mathbf{p}_1, \mathbf{p}_2\}R_1^\dagger=\{\mathbf{X}', \mathbf{x}'; \mathbf{P}', \mathbf{p}'\}. \tag{2.8}$$

Of course the results may be verified directly.

Part 2: To obtain the transformation from the intermediate variables $\mathbf{X}'(\mathbf{P}')$ to the center of mass variables $\mathbf{X}_{\rm cm}(\mathbf{P}_{\rm cm})$, we need to have (for i = 1,2) $\mathbf{x}_i(\mathbf{p}_i)$ multiplied (divided) by $\sqrt{\mu_i}$. This is achieved by the unitary squeeze operator (see, e.g., Ref. 2)

$$S_2=e^{(r_1+r_2)/2}\exp[i(r_1\mathbf{x}_1\cdot\mathbf{p}_1+r_2\mathbf{x}_2\cdot\mathbf{p}_2)], \tag{2.9}$$

where

[a]Permanent address: Department of Physics, Technion—Israel Institute of Technology, Haifa 32000, Israel.

 0022-2488/90/061413-03$03.00

$e^{r_i}=\sqrt{\mu_i},\ i=1,2.$

We then have immediately, by combining the two unitary transformations, that

$$(S_2R_1)\{\mathbf{x}_1,\mathbf{x}_2,\mathbf{p}_1;\mathbf{p}_2\}(S_2R_1)^\dagger=\{\mathbf{X}_{\rm cm},\mathbf{x}'';\mathbf{P},\mathbf{p}''\}, \quad (2.10)$$

with the new intermediate variables related to the initial variables for the particles and to the relative coordinate variables by

$$\mathbf{x}''=\sqrt{\mu_1\mu_2}(\mathbf{x}_2-\mathbf{x}_1)=\sqrt{\mu_1\mu_2}\,\mathbf{x}_r, \quad (2.11a)$$

$$\mathbf{p}''=(1/\sqrt{\mu_1\mu_2})(\mu_1\mathbf{p}_2-\mu_2\mathbf{p}_1)=(1/\sqrt{\mu_1\mu_2})\mathbf{p}_r. \quad (2.11b)$$

Part 3: The final component of the transformation consists now in passing from $(\mathbf{x}'',\mathbf{p}'')$ to $(\mathbf{x}_r,\mathbf{p}_r)$, by applying the unitary squeeze operator in the relative variables (which leaves the CM variables unaffected). This unitary squeeze operator has the form

$$\begin{aligned}S_3&=e^{-(r_1+r_2)/2}\exp[-i(r_1+r_2)\mathbf{x}''\cdot\mathbf{p}'']\\&=e^{-(r_1+r_2)/2}\exp[-i(r_1+r_2)\mathbf{x}_r\cdot\mathbf{p}_r]\\&=e^{-(r_1+r_2)/2}\exp[-i(r_1+r_2)(\mu_2\mathbf{x}_1\cdot\mathbf{p}_1+\mu_1\mathbf{x}_2\cdot\mathbf{p}_2\\&\quad-\mu_1\mathbf{x}_1\cdot\mathbf{p}_2-\mu_2\mathbf{x}_2\cdot\mathbf{p}_1)].\end{aligned} \quad (2.12)$$

The total unitary operator (U) effecting the transformation is obtained by combining the three separate unitary transformations and is given as

$$U=S_3S_2R_1. \quad (2.13)$$

The transformation assumes a particularly simple form for the case of equal masses, i.e., $\mu_1=\mu_2=\frac{1}{2}$. In that case $r=r_1=r_2=-(\ln 2)/2$, $\alpha=\pi/4$, and we have

$$\begin{aligned}S_3&=e^{-r}\exp[-ir(\mathbf{x}_1\cdot\mathbf{p}_1+\mathbf{x}_2\cdot\mathbf{p}_2-\mathbf{x}_1\cdot\mathbf{p}_2-\mathbf{x}_2\cdot\mathbf{p}_1)]\\&=e^{-r}\exp[-ir(\mathbf{x}_1\cdot\mathbf{p}_1+\mathbf{x}_2\cdot\mathbf{p}_2)]\\&\quad\times\exp[+ir(\mathbf{x}_1\cdot\mathbf{p}_2+\mathbf{x}_2\cdot\mathbf{p}_1)],\end{aligned} \quad (2.14)$$

and finally

$$\begin{aligned}U&=S_3S_2R_1\\&=\exp[-i(\ln 2/2)(\mathbf{x}_1\cdot\mathbf{p}_2+\mathbf{x}_2\cdot\mathbf{p}_1)]\\&\quad\times\exp[-i(\pi/4)(\mathbf{x}_1\cdot\mathbf{p}_2-\mathbf{x}_2\cdot\mathbf{p}_1)].\end{aligned} \quad (2.15)$$

In terms of creation and annihilation operators, defined by

$$x_i=\frac{1}{\sqrt{2m\omega}}(\mathbf{a}_i+\mathbf{a}_i^\dagger),\quad p_i=\frac{1}{i}\sqrt{\frac{m\omega}{2}}(\mathbf{a}_i-\mathbf{a}_i^\dagger),\quad i=1,2,$$

(for some ω) this may be written as

$$\begin{aligned}U&=\exp[-i(\ln 2/2)(\mathbf{a}_1\cdot\mathbf{a}_2-\mathbf{a}_1^\dagger\cdot\mathbf{a}_2^\dagger)]\\&\quad\times\exp[-i(\pi/4)(\mathbf{a}_1^\dagger\cdot\mathbf{a}_2-\mathbf{a}_2^\dagger\cdot\mathbf{a}_1)],\end{aligned}$$

which is recognized as a rotation followed by a two-mode squeezing. To end this section we note that by using standard techniques the general transformation operator U [Eq. (2.13)] may be written as a single exponent,

$$U=e^{-iG},$$

with

$$G=(\theta_0/2\sin\theta_0)(\mu_2\mathbf{x}_1\cdot\mathbf{p}_1+2\mathbf{x}_1\cdot\mathbf{p}_2-2\mu_2\mathbf{x}_2\cdot\mathbf{p}_1-\mu_2\cdot\mathbf{x}_2\cdot\mathbf{p}_2),$$

where $\cos\theta_0=(1+\mu_1)/2$.

III. UNITARY TRANSFORMATION: GENERALIZATION TO $N>2$

The Jacobi coordinate and momentum operators for N particles of masses m_i $(i=1,\ldots,N)$ are defined as follows:

$$\begin{aligned}&\zeta_1=\mathbf{x}_2-\mathbf{x}_1, && \pi_1=\frac{m_1\mathbf{p}_2-m_2\mathbf{p}_1}{m_1+m_2},\\&\zeta_2=\mathbf{x}_3-\frac{m_1\mathbf{x}_1+m_2\mathbf{x}_2}{m_1+m_2}, && \pi_2=\frac{(m_1+m_2)\mathbf{p}_3-m_3(\mathbf{p}_1+\mathbf{p}_2)}{m_1+m_2+m_3},\\&\zeta_{N-1}=\mathbf{x}_N-\frac{\sum_{i=1}^{N-1}m_i\mathbf{x}_i}{\sum_{i=1}^{N-1}m_i}, && \pi_{N-1}=\frac{(\sum_{i=1}^{N-1}m_i)\mathbf{p}_M-m_N\sum_{i=1}^{N-1}\mathbf{p}_i}{\sum_{i=1}^{N}m_i},\\&\zeta_N=\mathbf{X}_c=\frac{\sum_{i=1}^{N}m_i\mathbf{x}_i}{\sum_{i=1}^{N}m_i}, && \pi_N=\mathbf{P}_c=\sum_{i=1}^{N}\mathbf{p}_i,\end{aligned}$$

From this structure it is obvious that the required unitary operator for the N-particle system may be constructed by using successive unitary transformations between pairs of "particles." This is illustrated explicitly for the special case of three-particle system, i.e., $N=3$.

We denote now the center of mass coordinate and momentum of the subsystem of particles 1 and 2 by $\mathbf{x}_{12}$, $\mathbf{p}_{12}$, respectively, so that the Jacobi variables for the three-particle system are

$$\zeta_1=\mathbf{x}_2-\mathbf{x}_1, \quad (3.1a)$$

$$\zeta_2=\mathbf{x}_3-\mathbf{x}_{12}, \quad (3.1b)$$

$$\zeta_3=X=\frac{(m_1+m_2)\mathbf{x}_{12}+m_3\mathbf{x}_3}{m_1+m_2+m_3}=\mu_{12}\mathbf{x}_{12}+\mu_3\mathbf{x}_3, \quad (3.1c)$$

$$\pi_1=\mu_1\mathbf{p}_2-\mu_2\mathbf{p}_1, \quad (3.2a)$$

$$\pi_2=\mu_{12}\mathbf{p}_3-\mu_3\mathbf{p}_{12}, \tag{3.2b}$$

$$\pi_3=\mathbf{p}_{12}+\mathbf{p}_3, \tag{3.2c}$$

with

$$\mu_{12}=(m_1+m_2)/(m_1+m_2+m_3), \tag{3.3a}$$

$$\mu_3=m_3/(m_1+m_2+m_3), \tag{3.3b}$$

and the constants μ_1 and μ_2 are defined in Eq. (2.1). We first transform within the subsystem of particles 1 and 2 to the variables ξ_1, $\mathbf{x}_{12}$, π_1, $\mathbf{p}_{12}$ (as carried out in Sec. II). This is followed by another two-"particle" transformation between the center of mass of particles 1 and 2 and the particle 3. This is clear since by looking at Eqs. (3.1b), (3.1c), (3.2b), and (3.2c) we see that the transformation from $(\mathbf{x}_{12}, \mathbf{x}_3, \mathbf{p}_{12}, \mathbf{p}_3)$ to $(\xi_3, \xi_2, \pi_3, \pi_2)$ is now the same as for the two-particle system carried out in Sec. II with the appropriate renaming of the variables. Indeed, defining intermediate variables

$$\xi_3'=\sqrt{\mu_{12}}\mathbf{x}_{12}+\sqrt{\mu_3}\mathbf{x}_3, \tag{3.4a}$$

$$\xi_2'=-\sqrt{\mu_3}\mathbf{x}_{12}+\sqrt{\mu_{12}}\mathbf{x}_3, \tag{3.4b}$$

$$\pi_3'=\sqrt{\mu_{12}}\mathbf{p}_{12}+\sqrt{\mu_3}\mathbf{p}_3, \tag{3.5a}$$

$$\pi_2'=-\sqrt{\mu_3}\mathbf{p}_{12}+\sqrt{\mu_{12}}\mathbf{p}_3, \tag{3.5b}$$

the transformation $(\mathbf{x}_{12}, \mathbf{x}_3; \mathbf{p}_{12}, \mathbf{p}_3)\to(\xi_3', \xi_2'; \pi_3', \pi_2')$ is again a rotation, effected by $e^{-\beta L_3}$, where

$$L_3=\mathbf{x}_{12}\cdot\mathbf{p}_3-\mathbf{x}_3\cdot\mathbf{p}_{12}=\mu_1\mathbf{x}_1\cdot\mathbf{p}_3+\mu_2\mathbf{x}_2\cdot\mathbf{p}_3-\mathbf{x}_3\cdot\mathbf{p}_1-\mathbf{x}_3\cdot\mathbf{p}_2, \tag{3.6}$$

and the angle β is defined by

$$\cos\beta=\sqrt{\mu_{12}}, \quad \sin\beta=\sqrt{\mu_3}. \tag{3.7}$$

To obtain ξ_3 and π_3 from ξ_3' and π_3' we need to have $\mathbf{x}_{12}$ $(\mathbf{p}_{12})$ and $\mathbf{x}_3$ $(\mathbf{p}_3)$ multiplied (divided) by $\sqrt{\mu_{12}}$ and $\sqrt{\mu_3}$, respectively. This is achieved by the unitary squeeze operator

$$e^{(r_{12}+r_3)/2}\exp[i[r_{12}(\mathbf{x}_{12}\cdot\mathbf{p}_{12})+r_3(\mathbf{x}_3\cdot\mathbf{p}_3)]], \tag{3.8}$$

where

$$e^{r_{12}}=\sqrt{\mu_{12}}, \quad e^{r_3}=\sqrt{\mu_3}. \tag{3.9}$$

This transformation results also in

$$\xi_2'\to\xi_2''=\sqrt{\mu_{12}\mu_3}\,\xi_2, \tag{3.10a}$$

$$\pi_2'\to\pi_2''=\pi_2/\sqrt{\mu_{12}\mu_3}, \tag{3.10b}$$

so that now the unitary squeeze operator (affecting only ξ_2 and π_2),

$$e^{-(i/2)(r_{12}+r_3)}\exp[-i(r_{12}+r_3)\xi_2\cdot\pi_2],$$

will effect the final transformation

$$(\xi_2'',\pi_2'')\to(\xi_2,\pi_2).$$

IV. CONCLUSIONS

The transformation of the canonical variables of individual particles to those of the center of mass and the relative coordinates is expressed in an operator form. It goes without saying that applying the unitary operator to any state of the individual particles turns it into the corresponding state in the Jacobi variables. Application to the case of a dynamical system with harmonic potential have been considered many years ago, and the explicit form of the transformation matrix elements has been given, e.g., by Smirnov.[3] Recently Fan[1] has considered the case of a Hamiltonian with a harmonic potential that can be diagonalized using a unitary operator. The present work provides a general unitary operator for any number of particles with arbitrary masses.

ACKNOWLEDGMENTS

A.M. acknowledges the kind hospitality of the Theoretical Physics Institute at the University of Alberta.

This work is supported in part by the Natural Sciences and Engineering Research Council of Canada and in part by the Fund for Promotion of Research in Technion.

[1] Hong-yi Fan, J. Math. Phys. **30**, 1273 (1989).
[2] Y. Ben-Aryeh and A. Mann, Phys. Rev. A **32**, 552 (1985), Eq. (2.6).
[3] Yu. F. Smirnov, Nucl. Phys. **39**, 346 (1962).

Sternberg construction and reduction

A. V. Razumov
Institute for High Energy Physics, Protvino, Moscow Region 142284, USSR

(Received 13 July 1989; accepted for publication 3 January 1990)

A connection between the Sternberg construction, which allows one to introduce a symplectic structure on an associated fiber bundle with the base and the fiber being symplectic manifolds, and the reduction of symplectic manifolds is considered. It is shown that the Sternberg construction commutes with the reduction.

I. INTRODUCTION

In this paper we consider a connection between the two methods of constructing new symplectic manifolds from the given ones. The first method is called the Sternberg construction. The essence of this method is as follows.

Let us have two symplectic manifolds[1,2] M and F with the symplectic two-forms ω^M and ω^F. It is clear that on the manifold $E \equiv M \times F$ there exists a natural symplectic structure given by the two-form

$$\omega^E \equiv \mathrm{pr}_M^* \omega^M + \mathrm{pr}_F^* \omega^F,$$

where pr_M and pr_F are the projections of the direct product $M \times F$ onto M and F, respectively. A generalization of this construction was considered by Sternberg[3,4] when he described the motion of a particle in the Yang–Mills field. In his approach the symplectic manifold M was the base of a principal fiber bundle P with the structure group G, F was a left Hamiltonian G-space with an Ad*-equivariant momentum,[1] and E was a fiber bundle with the fiber F, associated with the principal fiber bundle P. A symplectic two-form ω^E on E was given by introducing a connection in P. In this, it was supposed that the symplectic manifold M was a cotangent bundle with a natural symplectic structure,[1,2] and the connection in P was chosen in a special way.

We consider the generalization of the Sternberg construction to the case when M is an arbitrary symplectic manifold. In this case it appears that for the Sternberg construction to give a symplectic structure on E, it is necessary to require some additional conditions to be satisfied. We formulate these conditions explicitly.

The second method to construct new symplectic manifolds we consider, is the reduction of symplectic manifolds. The initial object in this method is a left Hamiltonian H-space F with a momentum mapping Ψ^F. Considering the level F'_λ of some fixed value λ of the momentum mapping Ψ^F, we see that the restriction of the symplectic two-form ω^F to this level is a degenerate two-form. However after the factorization of F'_λ with respect to the action of the isotropy subgroup of this momentum value we get the reduced symplectic manifold F_λ with a natural symplectic structure.[5]

We suppose that the left Hamiltonian G-space F, used in the Sternberg construction is also a right Hamiltonian H-space with a momentum mapping Ψ^F. In this case the fiber bundle E may also be considered as a right Hamiltonian H-space with a momentum mapping Ψ^E. Perform the reduction of a level λ of the momentum mapping Ψ^F. As a result we get the reduced symplectic manifold F_λ. This manifold is a left Hamiltonian G-space and can be used in the Sternberg construction. We may also perform the reduction of the level λ of the momentum mapping Ψ^E. It appears that as a result of this we get one and the same symplectic manifold. In other words, one can say that the Sternberg construction commutes with the reduction. The proof of this fact is the main content of the present paper which is organized as follows.

In Secs. II and III we recall necessary definitions and facts on Lie groups, Lie symplectic transformation groups, and discuss the reduction of symplectic manifolds. In Sec. IV we consider the generalization of the Sternberg construction and formulate conditions under which it gives a symplectic structure on the corresponding associated fiber bundle. In Sec. V it is proved that the Sternberg construction commutes with the reduction.

Some notations not explained in the text are: $X(\varphi)$ is the Lie derivative of the k-form φ with respect to the vector field X, $i(X)\varphi$ is the inner product of the vector field X and the k-form φ, and pr_{M_i} is the projection of the direct product $M_1 \times \cdots \times M_n$ onto the ith factor M_i, $i = 1,...,n$. As usual, we consider all the manifolds to be of class C^∞.

II. LIE TRANSFORMATION GROUPS

A Lie group G is called a Lie transformation group on a manifold M, if a mapping R^M: $M \times G \to M$, satisfying the following conditions is given.

(i) If for each $g \in G$ we define a mapping R_g^M: $M \to M$ by the relation

$$R_g^M(m) \equiv R^M(m,g),$$

then for any $g_1, g_2 \in G$ the following equality is valid:

$$R_{g_1}^M \circ R_{g_2}^M = R_{g_2 g_1}^M.$$

(ii) The mapping R_e^M, where e is the unit element of the group G, is an identity map:

$$R_e^M = \mathrm{id}_M.$$

In this case we also say that a right action of the group G on M is given, or that M is a right G-space. Often the following notation is used:

$$m \cdot g \equiv R^M(m,g).$$

Construct a homomorphic mapping from the Lie algebra lG of the Lie group G to the Lie algebra of the vector

 0022-2488/90/061416-06$03.00

fields on M, assigning to an element $u\in lG$ the vector field U^M, given by the relation

$$U_m^M(f)\equiv\frac{d}{dt}f(m\cdot\exp(tu))|_{t=0},$$

for any $m\in M, f\in C^\infty(M)$.

A left action of a Lie group G on a manifold M is specified by a mapping $L^M: G\times M\to M$, which satisfies the conditions

$$L_{g_1}^M\circ L_{g_2}^M=L_{g_1g_2}^M,\quad L_e^M=\mathrm{id}_M,$$

where

$$L_g^M(m)\equiv g\cdot m\equiv L^M(g,m).$$

In this case one constructs an antihomomorphic mapping from the Lie algebra lG to the Lie algebra of the vector fields on M, assigning to an element $u\in lG$ the vector field $\bar{U}^M$, given by the relation

$$\bar{U}^M(f)\equiv\frac{d}{dt}f(\exp(tu)\cdot m)|_{t=0}.$$

Let a left action of a Lie group G on a manifold M be given. Introduce on M an equivalence relation, considering the points belonging to one and the same orbit to be equivalent. Denote the corresponding quotient set M/G by N. Suppose that it is possible to introduce on N the structure of a manifold in such a way that the canonical projection τ^M: $M\to N$ be a submersion on M.[1]

Proposition 2.1: If a k-form φ on M (with the values in a linear space L) satisfies the conditions

$$R_g^{M*}\varphi=\varphi,$$

$$i(U^M)\varphi=0,$$

for any $g\in G$, $u\in lG$, then there exists a unique k-form ψ on N (with the values in the linear space L) such that

$$\tau^{M*}\psi=\varphi.$$ □

III. LIE SYMPLECTIC TRANSFORMATION GROUPS AND REDUCTION

Let M be a manifold where a closed nondegenerate two-form ω^M is given. In this case M is said to be a symplectic manifold.

Let M be a symplectic manifold and F be a diffeomorphic mapping of the manifold M onto itself. If

$$F^*\omega^M=\omega^M,$$

then F is called a symplectic transformation. Suppose now that M is a right G-space and

$$R_g^{M*}\omega^M=\omega^M,\tag{3.1}$$

for any $g\in G$. In this case M is called a right symplectic G-space. From (3.1) it follows that

$$U^M(\omega^M)=0$$

for any $u\in lG$, i.e., the vector fields U^M are locally Hamiltonian.[1,2] Suppose that for any $u\in lG$ the vector field U^M is Hamiltonian,[1,2] in this case one can construct a linear mapping

$$\varphi^M:\ u\in lG\to\varphi_u^M\in C^\infty(M),$$

such that

$$i(U^M)\omega^M=d\varphi_u^M.$$

In such a situation M is called a right Hamiltonian G-space.

Let lG^* be the dual of lG. We associate with the mapping φ^M the mapping Φ^M: $M\to lG^*$, given by the equality

$$\langle\Phi^M(m)|u\rangle\equiv\varphi_u^M(m),$$

for any $m\in M$, $u\in lG$. Here and henceforth we denote the action of an element $\varkappa\in lG^*$ on an element $u\in lG$ by $\langle\varkappa|u\rangle$.

Suppose that the manifold M is connected, then for any $g\in G$ the following relation is valid:

$$\Phi^M\circ R_g^M=\mathrm{Ad}^*(g^{-1})\circ(\Phi^M+\mu^M(g)),\tag{3.2}$$

where $\mu^M(g)$ is constant on M, and $\mathrm{Ad}^*(g)$ is the operator of the coadjoint representation of the Lie group G, connected with the operator $\mathrm{Ad}(g)$ of the adjoint representation[6,7] by the equality

$$\langle\mathrm{Ad}^*(g)\varkappa|u\rangle\equiv\langle\varkappa|\mathrm{Ad}(g^{-1})u\rangle,$$

for all $\varkappa\in lG^*$, $u\in lG$. For any $g_1,g_2\in G$ from (3.2) we get

$$\mu^M(g_1g_2)=\mathrm{Ad}^*(g_1)\mu^M(g_2)+\mu^M(g_1).\tag{3.3}$$

Considering μ^M as a one-dimensional cochain of the Lie group G with the coefficients in lG^*, we conclude from (3.3) that μ^M is a cocycle.[8]

The mapping Φ^M is called a momentum mapping. A momentum mapping Φ^M is called Ad^*-equivariant provided $\mu^M(g)=0$ for all $g\in G$.

Let now a symplectic manifold M be a left G-space and

$$L_g^*\omega^M=\omega^M,$$

for any $g\in G$. In this case M is called a left Hamiltonian G-space if there exists a linear mapping

$$\bar{\varphi}^M:\ u\in lG\to\bar{\varphi}_u^M\in C^\infty(M),$$

such that

$$i(\bar{U}^M)\omega^M=d\bar{\varphi}_u^M.$$

Here a momentum mapping $\bar{\Phi}^M$: $M\to lG^*$ is given by the equality

$$\langle\bar{\Phi}^M(m)|u\rangle\equiv\bar{\varphi}_u^M(m),$$

for all $m\in M$, $u\in$ lG. If the manifold M is connected, then for any $g\in G$ the following relation is valid:

$$\bar{\Phi}^M\circ L_g^M=\mathrm{Ad}^*(g)\circ\bar{\Phi}^M-\bar{\mu}^M(g),$$

where $\bar{\mu}^M$ is a one-dimensional cocycle of the Lie group G with the coefficients in lG^*. If for any $g\in G$ $\bar{\mu}^M(g)=0$, then the momentum mapping $\bar{\Phi}^M$ is called Ad^*-equivariant.

Let M be a right Hamiltonian G-space and Φ^M be a corresponding momentum mapping. Suppose that $\varkappa$ is a regular value of the mapping Φ^M then the set $M'_\varkappa\equiv\Phi^{M-1}(\varkappa)$ is a submanifold of M. Introduce the notation

$$G_\varkappa\equiv\{g\in G\,|\,\mathrm{Ad}^*(g^{-1})(\varkappa+\mu^M(g))=\varkappa\}.$$

It is clear, that the submanifold $M'_\varkappa$ is invariant with respect to the transformations R_g^M, where $g\in G_\varkappa$. It can be easily shown, that $G_\varkappa$ is a closed subgroup of the Lie group G. Thus, the action R^M of the Lie group G on M generates the action $R^{M'_\varkappa}$ of the Lie group $G_\varkappa$, on $M'_\varkappa$, here for all $g\in G_\varkappa$

$$\iota^{M'_\varkappa}\circ R_g^{M'_\varkappa} = R_g^{M}\circ\iota^{M'_\varkappa},$$

where $\iota^{M'_\varkappa}$ is the inclusion mapping of $M'_\varkappa$ into M. For all $g\in G_\varkappa$ the following relation is valid:

$$R_g^{M'_\varkappa *}\omega^{M'_\varkappa} = \omega^{M'_\varkappa},$$

where

$$\omega^{M'_\varkappa} \equiv \iota^{M'_\varkappa *}\omega^M.$$

It can also be shown, that

$$i(U^{M'_\varkappa})\omega^{M'_\varkappa} = 0,$$

for all $u\in lG_\varkappa$. Suppose that on the quotient set $M_\varkappa \equiv M'_\varkappa/G_\varkappa$ the structure of a manifold can be introduced in such a way that the canonical projection $\tau^{M'_\varkappa}$: $M'_\varkappa \to M_\varkappa$ be a submersion on $M'_\varkappa$. From Proposition 2.1 it then follows that on $M_\varkappa$ there exists a unique two-form $\omega^{M_\varkappa}$, which satisfies the condition

$$\omega^{M'_\varkappa} = \tau^{M'_\varkappa *}\omega^{M_\varkappa}.$$

It is clear that the two-form $\omega^{M_\varkappa}$ is closed. It can be shown that from the condition

$$i(x)\omega^{M'_\varkappa} = 0,$$

where $x\in T_m(M'_\varkappa)$, $m\in M'_\varkappa$, it follows that

$$x = U_m^{M'_\varkappa},$$

for some $u\in lG_\varkappa$. From here we see that the two-form $\omega^{M_\varkappa}$ is nondegenerate.

Thus, the manifold $M_\varkappa$ has a natural structure of a symplectic manifold. This symplectic manifold is called the reduced symplectic manifold of the level $\varkappa$, and the procedure described above is called the reduction of the level $\varkappa$ of the momentum mapping Φ^M.[5] A detailed description of the reduction procedure may be found in Refs. 1, 2, 9, and 10.

IV. STERNBERG CONSTRUCTION

Let $E(M,F,G,\sigma)$ be a fiber bundle, associated with a principal fiber bundle $P(M,G,\pi)$,[6,7] $\{(V_i,\psi_i)\}_{i\in I}$ be an atlas of the fiber bundle P with the transition functions g_{ik}, and $\{(V_i,\varphi_i)\}_{i\in I}$ be an atlas of the fiber bundle E with the same transition functions. Consider a mapping $\chi: P\times F\to E$, given locally by the relation

$$\chi|_{\pi^{-1}(V_i)\times F}(p,f) \equiv \varphi_i^{-1}(\pi(p),\psi_{i\pi(p)}(p)\cdot f). \qquad (4.1)$$

Introduce on $P\times F$ the structure of a right G-space supposing that

$$R_g^{P\times F}(p,f) \equiv (p\cdot g, g^{-1}\cdot f). \qquad (4.2)$$

Note that

$$\chi(p',f') = \chi(p,f),$$

iff there exists an element $g\in G$ such that

$$(p',f') = (p,f)\cdot g,$$

besides, the mapping χ is surjective. Thus, one can identify E with the quotient set $P\times F/G$. It can be also shown that χ is a submersion on $P\times F$.

Suppose now that M is a symplectic manifold, F is a left Hamiltonian G-space with an Ad*-equivariant momentum $\overline{\Phi}^F$. Let γ be a connection form of some connection in the principal fiber bundle P. Construct a two-form

$$\omega^{P\times F} \equiv \mathrm{pr}_P^*\pi^*\omega^M + \mathrm{pr}_F^*\omega^F - d\langle \mathrm{pr}_F^*\overline{\Phi}^F|\mathrm{pr}_P^*\gamma\rangle.$$

Proposition 4.1: For any $g\in G$, $u\in lG$ the following equalities are valid:

$$R_g^{P\times F*}\omega^{P\times F} = \omega^{P\times F},$$

$$i(U^{P\times F})\omega^{P\times F} = 0. \qquad \square$$

From Propositions 2.1 and 4.1 it follows that there exists a unique two-form ω^E on E which satisfies the condition

$$\chi^*\omega^E = \omega^{P\times F}. \qquad (4.3)$$

Theorem 4.1: If for any $\varkappa\in\overline{\Phi}^F(F)$ and any point $p\in P$ from the equality

$$i(x)(\pi^*\omega^M - \langle\varkappa|\Gamma\rangle) = 0,$$

where $x\in T_p(P)$, and Γ is the curvature form of the connection, given by the form γ, it follows that $x = U_p^P$ for some $u\in lG$, then the two-form ω^E defines on E the structure of a symplectic manifold.

Proof: It is clear that the two-form ω^E is closed. Let us prove that it is nondegenerate. Suppose that

$$i(x)\omega^{P\times F} = 0, \qquad (4.4)$$

where $x\in T_{(p,f)}(P\times F)$. Using the explicit expression for the two-form $\omega^{P\times F}$, it can be shown that equality (4.4) is equivalent to the two equalities

$$i(\mathrm{pr}_{F*}(x))\omega^F + \langle d\overline{\Phi}^F|i(\mathrm{pr}_{P*}(x))\gamma\rangle = 0, \qquad (4.5)$$

$$i(\mathrm{pr}_{P*}(x))\pi^*\omega^M - \langle i(\mathrm{pr}_{F*}(x))d\overline{\Phi}^F|\gamma_p\rangle - \langle\overline{\Phi}_f^F|i(\mathrm{pr}_{P*}(x))d\gamma\rangle = 0. \qquad (4.6)$$

From (4.5) for any $u\in lG$ we obtain

$$\langle i(\mathrm{pr}_{F*}(x))d\overline{\Phi}^F|u\rangle = \langle\overline{\Phi}_f^F|[i(\mathrm{pr}_{P*}(x))\gamma,u]\rangle.$$

Wherefrom it follows that

$$\langle i(\mathrm{pr}_{F*}(x))d\overline{\Phi}^F|\gamma_p\rangle = \langle\overline{\Phi}_f^F|[i(\mathrm{pr}_{P*}(x))\gamma,\gamma_p]\rangle. \qquad (4.7)$$

Using (4.7) and the equality[6,7]

$$d\gamma + \tfrac{1}{2}[\gamma,\gamma] = \Gamma,$$

from relation (4.6) we get

$$i(\mathrm{pr}_{P*}(x))\pi^*\omega^M - \langle\overline{\Phi}_f^F|i(\mathrm{pr}_{P*}(x))\Gamma\rangle = 0. \qquad (4.8)$$

If the condition of the Theorem we are proving is valid, then from (4.8) it follows that $\mathrm{pr}_{P*}(x) = U_p^P$ for some $u\in lG$. From equality (4.5) we then have $\mathrm{pr}_{F*}(x) = -\overline{U}_f^F$, hence

$$x = U_{(p,f)}^{P\times F}.$$

Suppose now that

$$i(x')\omega^E = 0,$$

where $x'\in T_e(E)$. Consider a point $(p,f)\in P\times F$ such that $\chi(p,f) = e$, and choose a vector $x\in T_{(p,f)}(P\times F)$ in such a way that $\chi_*(x) = x'$. Then for any vector $y\in T_{(p,f)}(P\times F)$ we have

$$\omega^{P\times F}(x,y) = \omega^E(x',\chi_*(y)) = 0,$$

hence

$$i(x)\omega^{P\times F} = 0,$$

and $x = U^{P\times F}_{(p,f)}$ for some $u\in lG$. As

$$\chi\circ R^{P\times F}_g = \chi,$$

for any $g\in G$, then

$$x' = \chi_*(x) = 0.$$

Thus, the two-form ω^E is nondegenerate. □

Sternberg[3,4] considered the construction described above for the following special choice of the principal fiber bundle P and the connection in it. Let $Q(N,G,\rho)$ be a principal fiber bundle, and δ be a connection form on Q. Put $M\equiv T^*(N)$ and denote by $p_N: T^*(N)\to N$ the canonical projection. We shall consider $T^*(N)$ as a symplectic manifold with a natural symplectic structure.[1,2] Consider the induced fiber bundle $P = p_N'(Q)$ with the connection form $\gamma\equiv p^*_{NQ}\delta$, where (p_N, p_{NQ}) is the canonical homomorphism from the fiber bundle P to the fiber bundle Q. It can be shown that under such a choice of Q and γ the condition of Theorem 4.1 is valid.

An alternative version of the Sternberg construction was given by Weinstein.[11] He used the reduction procedure to build the universal symplectic manifold independent of any connection (see also Ref. 12). Note, that his method can be used only in the case considered by Sternberg.

In the case of an arbitrary symplectic manifold M there arises a question, if we may choose the connection γ so that the condition of Theorem 4.1 be valid. An answer to this question is unknown to us. This question and related topics were discussed in Refs. 13, 14.

V. REDUCTION OF ASSOCIATED FIBER BUNDLES

Let M be a symplectic manifold, $P(M,G,\pi)$ be a principal fiber bundle, F be a left Hamiltonian G-space with an Ad*-equivariant momentum $\overline{\Phi}^F$, and γ be a connection form on P. Consider a fiber bundle $E(M,F,G,\sigma)$, asociated with P. Suppose that the condition of Theorem 4.1 is valid, hence the two-form ω^E, given by relation (4.3), defines on E the structure of a symplectic manifold. Let a right action of a Lie group H is also given on F, so that F is a right Hamiltonian H-space. Denote a corresponding momentum mapping by Ψ^F. We shall consider that the manifold F is connected, then

$$\Psi^F\circ R^F_h = \mathrm{Ad}^*(h^{-1})\circ(\Psi^F + \nu^F(h)), \tag{5.1}$$

where ν^F is a one-dimensional cocycle of the Lie group H with the coefficients in lH^*. Suppose that for any $g\in G$ and $h\in H$

$$L^F_g\circ R^F_h = R^F_h\circ L^F_g, \tag{5.2}$$

and besides

$$\Psi^F\circ L^F_g = \Psi^F,$$

$$\overline{\Phi}^F\circ R^F_h = \overline{\Phi}^F. \tag{5.3}$$

Let $\{(V_i,\psi_i)\}_{i\in I}$ be an atlas of the fiber bundle P with the transition functions g_{ik}, and $\{(V_i,\varphi_i)\}_{i\in I}$ be an atlas of the fiber bundle E with the same transition functions. Introduce on E a right action of the Lie group H in the following way. Let $e\in\sigma^{-1}(V_i)$, then for any $h\in H$ put

$$R^E_h(e)\equiv\varphi^{-1}_{i\sigma(e)}\circ R^F_h\circ\varphi_{i\sigma(e)}(e).$$

Using (5.2), it is easy to show that this definition is independent of the choice of a chart, and

$$\sigma\circ R^E_h = \sigma.$$

Defining the mapping $\chi: P\times F\to E$ by Eq. (4.1), we obtain

$$\chi\circ R^{P\times F}_h = R^E_h\circ\chi,$$

where

$$R^{P\times F}_h(p,f)\equiv(p,f\cdot h).$$

We have denoted the right action of the Lie group H on $P\times F$ in the same manner as the right action of the Lie group G on $P\times F$ given by formula (4.2). In what follows, it is always clear from the context the action of which group we have in mind.

Define a mapping $\Psi^{P\times F}$: $P\times F\to lH^*$, putting

$$\Psi^{P\times F}\equiv\Psi^F\circ\mathrm{pr}_F.$$

From (4.2) and (5.3) we get

$$\Psi^{P\times F}\circ R^{P\times F}_g = \Psi^{P\times F},$$

for all $g\in G$. From Proposition 2.1 it then follows that we can uniquely define the mapping Ψ^E: $E\to lH^*$, which satisfies the condition

$$\Psi^{P\times F} = \Psi^E\circ\chi.$$

Proposition 5.1: The action R^E of the Lie group H defines on E the structure of a right Hamiltonian H-space with a momentum mapping Ψ^E. The cocycle ν^E in the relation

$$\Psi^E\circ R^E_h = \mathrm{Ad}^*(h^{-1})\circ(\Psi^E + \nu^E(h))$$

coincides with the cocycle ν^F from relation (5.1). □

Let $\lambda\in\Psi^F(F)$, suppose that necessary conditions to perform the reduction of a level λ of the momentum mapping Ψ^F are satisfied (see Sec. III). As the result of the reduction we get the symplectic manifold $F_\lambda\equiv F'_\lambda/H_\lambda$, where $F'_\lambda\equiv\Psi^{F-1}(\lambda)$,

$$H_\lambda\equiv\{h\in H\,|\,\mathrm{Ad}^*(h^{-1})(\lambda+\nu^F(h)) = \lambda\}.$$

Denote by $\iota^{F'_\lambda}$ the inclusion mapping of F'_λ into F. The symplectic two-form ω^{F_λ} on F_λ satisfies the condition

$$\omega^{F'_\lambda} = \iota^{F'_\lambda *}\omega^F = \tau^{F'_\lambda *}\omega^{F_\lambda}, \tag{5.4}$$

where $\tau^{F'_\lambda}$: $F'_\lambda\to F_\lambda$ is the canoncial projection.

From (5.3) it follows that the submanifold F'_λ is invariant with respect to the left action L^F of the Lie group G on F. Thus the action L^F generates the left action $L^{F'_\lambda}$ of the Lie group G on F'_λ. Here for any $g\in G$

$$\iota^{F'_\lambda}\circ L^{F'_\lambda}_g = L^F_g\circ\iota^{F'_\lambda}.$$

From (5.2) it now follows that we can uniquely define the left action of L^{F_λ} of the Lie group G on F_λ, satisfying the condition

$$L^{F_\lambda}_g\circ\tau^{F'_\lambda} = \tau^{F'_\lambda}\circ L^{F'_\lambda}_g, \tag{5.5}$$

for any $g\in G$.

Proposition 5.2: The action L^{F_λ} supplies F_λ with the structure of a left Hamiltonian G-space with an Ad*-equivariant momentum mapping $\overline{\Phi}^{F_\lambda}$, given by the condition

$$\overline{\Phi}^{F_\lambda}\circ\tau^{F'_\lambda} = \overline{\Phi}^F\circ\iota^{F'_\lambda}.$$ □

Proceed now to the consideration of the reduction of the level λ of the momentum mapping Ψ^E. It is clear that $\lambda \in \Psi^E(E)$ and $\lambda \in \Psi^{P\times F}(P\times F)$, here

$$\chi(\Psi^{P\times F-1}(\lambda)) = \Psi^{E-1}(\lambda),$$

and, besides,

$$\Psi^{P\times F-1}(\lambda) = P\times \Psi^{F-1}(\lambda).$$

It is easy to see that $E'_\lambda(M,F'_\lambda,G,\sigma'_\lambda)$, where $E'_\lambda \equiv \Psi^{E-1}(\lambda)$, $\sigma'_\lambda \equiv \sigma|_{E_\lambda}$, is the fiber bundle, for which the set $\{(V_i,\varphi'_{\lambda i})\}_{i\in I}$, where

$$\varphi'_{\lambda i} \equiv \varphi_i|_{\sigma^{-1}(V_i)\cap E'_\lambda}, \tag{5.6}$$

is an atlas. It can be shown that the transition functions of the atlas $\{(V_i,\varphi'_{\lambda i})\}_{i\in I}$ coincide with g_{ik}, hence, the fiber bundle E'_λ is associated with the principal fiber bundle P. Introduce a mapping $\chi'_\lambda: P\times F'_\lambda \to E'_\lambda$, locally given by the equality

$$\chi'_\lambda|_{\pi^{-1}(V_i)\times F'_\lambda}(p,f) \equiv \varphi'^{-1}_{\lambda i}(\pi(p),\psi_{i\pi(p)}(p)\cdot f).$$

From the definition of the mapping χ'_λ we get

$$\chi\circ\iota^{P\times F'_\lambda} = \iota^{E'_\lambda}\circ\chi'_\lambda, \tag{5.7}$$

where $\iota^{P\times F'_\lambda}$ is the inclusion mapping of $P\times F'_\lambda$ into $P\times F$, and $\iota^{E'_\lambda}$ is the inclusion mapping of E'_λ into E.

From Proposition 5.1 it follows that the submanifold E'_λ is invariant with respect to the restriction of the action R^E to the subgroup H_λ. Hence, on E'_λ there is given the right action $R^{E'_\lambda}$ of the Lie group H_λ, satisfying the condition

$$\iota^{E'_\lambda}\circ R_h^{E'_\lambda} = R_h^E\circ\iota^{E'_\lambda},$$

for any $h\in H_\lambda$. It is clear that

$$\sigma'_\lambda\circ R_h^{E'_\lambda} = \sigma'_\lambda,$$

hence, we can correctly and uniquely define the mapping $\sigma_\lambda:\ E_\lambda \equiv E'_\lambda/H_\lambda \to M$, satisfying the condition

$$\sigma'_\lambda\circ\tau^{E'_\lambda} = \sigma_\lambda,$$

where $\tau^{E'_\lambda}:\ E'_\lambda \to E_\lambda$ is the canonical projection.

Let $i\in I$, for any $h\in H_\lambda$ the mapping $\varphi'_{\lambda i}$, given by the equality (5.6), satisfies the condition

$$\varphi'_{\lambda i}\circ R_h^{E'_\lambda} = R^{V_i\times F'_\lambda}\circ\varphi'_{\lambda i},$$

where $R_h^{V_i\times F'_\lambda}(m,f) \equiv (m,f\cdot h)$. Hence, we can uniquely define the mapping $\varphi_{\lambda i}:\sigma_\lambda^{-1}(V_i) \to V_i\times F_\lambda$, satisfying the condition

$$\varphi_{\lambda i}\circ\tau^{E'_\lambda} = \tau^{V_i\times F'_\lambda}\circ\varphi'_{\lambda i}, \tag{5.8}$$

where $\tau^{V_i\times F'_\lambda}(m,f) \equiv (m,\tau^{F'_\lambda}(f))$. From here we get the equality

$$\mathrm{pr}_{V_i}\circ\varphi_{\lambda i} = \sigma_\lambda.$$

It can be shown, that the mapping $\varphi_{\lambda i}$ is a bijection. Introduce on E_λ the manifold structure, requiring that for all $i\in I$ the mappings $\varphi_{\lambda i}$ be diffeomorphisms. It is easy to understand that we have got a fiber bundle $E_\lambda(M,F_\lambda,G,\sigma_\lambda)$, for which the set $\{(V_i,\varphi_{\lambda i})\}_{i\in I}$ is an atlas. Using (5.5) and (5.8), one may show that the transition functions of the atlas $\{(V_i,\varphi_{\lambda i})\}_{i\in I}$ coincide with g_{ik}, therefore, the fiber bundle E_λ is associated with the principal fiber bundle P.

So, we see that necessary conditions to perform the reduction of the level λ of the momentum mapping Ψ^E are satisfied. The symplectic two-form ω^{E_λ} on the reduced symplectic manifold E_λ satisfies the condition

$$\tau^{E'_\lambda *}\omega^{E_\lambda} = \omega^{E'_\lambda}, \tag{5.9}$$

where $\omega^{E'_\lambda} = \iota^{E'_\lambda *}\omega^E$. Introducing the mapping $\chi_\lambda:\ P\times F_\lambda \to E_\lambda$ locally given by the condition

$$\chi_\lambda|_{\pi^{-1}(V_i)\times F_\lambda}(p,f) \equiv \varphi^{-1}_{\lambda i}(\pi(p),\psi_{i\pi(p)}(p)\cdot f),$$

the following equality is valid:

$$\chi_\lambda\circ\tau^{P\times F'_\lambda} = \tau^{E'_\lambda}\circ\chi'_\lambda, \tag{5.10}$$

where $\tau^{P\times F'_\lambda}(m,f) \equiv (m,\tau^{F'_\lambda}(f))$.

Theorem 5.1: If we define the two-form $\omega^{P\times F_\lambda}$ by

$$\omega^{P\times F_\lambda} \equiv \mathrm{pr}_P^*\pi^*\omega^M + \mathrm{pr}_{F_\lambda}^*\omega^{F_\lambda} - d\,\langle \mathrm{pr}_{F_\lambda}^*\overline{\Phi}^{F_\lambda}|\mathrm{pr}_P^*\gamma\rangle,$$

then the following equality is valid:

$$\chi_\lambda^*\omega^{E_\lambda} = \omega^{P\times F_\lambda}.$$

Proof: Using Eqs. (5.10), (5.9), and (5.7), we get

$$\tau^{P\times F'_\lambda *}\chi_\lambda^*\omega^{E_\lambda} = \iota^{P\times F'_\lambda *}\omega^{P\times F}. \tag{5.11}$$

It is easy to show that the following equality is valid:

$$\mathrm{pr}_P\circ\iota^{P\times F'_\lambda} = \mathrm{pr}_P\circ\tau^{P\times F'_\lambda}.$$

From here it follows that

$$\iota^{P\times F'_\lambda *}\mathrm{pr}_P^*\pi^*\omega^M = \tau^{P\times F'_\lambda *}\mathrm{pr}_P^*\pi^*\omega^M, \tag{5.12}$$

$$\iota^{P\times F'_\lambda *}\mathrm{pr}_P^*\gamma = \tau^{P\times F'_\lambda *}\mathrm{pr}_P^*\gamma. \tag{5.13}$$

It is also easy to verify the validity of the equalities

$$\mathrm{pr}_F\circ\iota^{P\times F'_\lambda} = \iota^{F'_\lambda}\circ\mathrm{pr}_{F'_\lambda}, \tag{5.14}$$

$$\tau^{F'_\lambda}\circ\mathrm{pr}_{F'_\lambda} = \mathrm{pr}_{F_\lambda}\circ\tau^{P\times F'_\lambda}. \tag{5.15}$$

Using Eqs. (5.14), (5.4), and (5.15), we have

$$\iota^{P\times F'_\lambda *}\mathrm{pr}_F^*\omega^F = \tau^{P\times F'_\lambda *}\mathrm{pr}_{F_\lambda}^*\omega^{F_\lambda}. \tag{5.16}$$

Analogously,

$$\iota^{P\times F'_\lambda *}\mathrm{pr}_F^*\overline{\Phi}^F = \tau^{P\times F'_\lambda *}\mathrm{pr}_{F_\lambda}^*\overline{\Phi}^{F_\lambda}. \tag{5.17}$$

From (5.11)–(5.13), (5.16), and (5.17) it follows that

$$\tau^{P\times F'_\lambda *}\chi_\lambda^*\omega^{E_\lambda} = \tau^{P\times F'_\lambda *}\omega^{P\times F_\lambda}.$$ □

Thus, we see that to construct the reduced symplectic manifold E_λ one may first, perform the reduction of the level λ of the momentum mapping Ψ^F, and then use the obtained symplectic manifold F_λ in the Sternberg construction.

From the point of view of the alternative version of the Sternberg construction given by Weinstein[11] we may say that we have proven the commutativity of the reduction in stages.

In the next paper we shall use Theorem 5.1 to perform the reduction of cotangent bundles.

[1]R. Abraham and J. Marsden, *Foundation of Mechanics* (Benjamin, New York, 1978).
[2]V. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[3]S. Sternberg, Proc. Natl. Acad. Sci. USA **74**, 5253 (1977).
[4]S. Sternberg, "On the Role of Field Theories in our Physical Conception of Geometry," in *Lecture Notes in Mathematics* (Springer, Berlin, 1978), Vol. 676.
[5]J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[6]R. Sulanke and P. Wintgen, *Differentialgeometrie und Faserbundel* (Verlag der Wissenschaften, Berlin, 1972).
[7]S. Kobayashi and K. Nomizu, *Foundation of Dfferential Geometry* (Interscience, New York, 1963), Vol. I.
[8]A. Guichardet, *Cohomologie des Groupes Topologiques et des Algebre de Lie* (Cedic, Paris, 1980).
[9]N. Woodhouse, *Geometric Quantization* (Claredon, Oxford, 1980).
[10]V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U. P., Cambridge, 1984).
[11]A. Weinstein, Lett. Math. Phys. **2**, 417 (1978).
[12]V. Guillemin and S. Sternberg, Lett. Math. Phys. **6**, 231 (1982).
[13]A. Weinstein, Adv. Math. **37**, 239 (1980).
[14]E. Lerman, Lett. Math. Phys. **15**, 335 (1988).

The problem of gravity-gyroscopic waves, which are excited by the oscillations of a curve

V. M. Kharik and U. D. Pletner
Department of Mathematics, Physics Faculty of Moscow State University, 117234 Moscow, USSR

(Received 13 September 1989; accepted for publication 8 November 1989)

The problem of the oscillations of an ideal stratified rotating fluid, which are excited by a curve in the case when the distribution of pressure on both sides of curve is prescribed, is considered. The solution to the problem, as well as results about symmetry properties of the potentials used for the solution of such problems, is obtained. The question of the uniqueness of the solution is also considered.

I. INTRODUCTION

Problems of the dynamics of stratified fluids are now under consideration by many scientists. It is necessary to note that the general theoretical aspects of the mathematical models, which describe dynamics and mechanics of such fluids, have been analyzed due to the success of the modern theory of the differential equations. One may not see the same situation in the case of the concrete initial-boundary value problems, which may have explicit solutions. This type of problem can be useful in applications and can help bring about a deeper understanding of the general mathematical models of the physical phenomena in fluids.

This paper continues research started in Refs. 1–5 and is connected to problems of the excitation of oscillations in stratified rotating fluids by oscillating curves. These works presented solutions to these problems. Such problems are connected with some questions in cryogenic fluid technology and oceanography.

This paper presents the case when the distribution of pressure on both sides of a curve is prescribed. In this case uncommon boundary conditions (which include time derivatives) have appeared due to the structure of the potential theory for the equation to be considered later (see Ref. 6). The solution is obtained by using two potentials, which were developed in Ref. 3.

II. THE DEFINITION OF "PROBLEM W"

We shall consider flat movements of an ideal rotating stratified fluid as in Refs. 1–5. The consideration of such problems leads to the equation of gravity-gyroscopic waves in two-dimensional space (see, for example, Ref. 2):

$$\frac{\partial^2}{\partial t^2}\nabla^2 U + \omega_0^2 U_{x_1x_1} + \alpha^2 U_{x_2x_2} = 0, \tag{2.1}$$

where ∇^2 is Laplace operator in two-dimensional space with variables x_1 and x_2, ω_0^2 is the square of the Waissala–Brunt frequency,[5] and α is the Coriolis parameter. We note that in this paper we do not disucuss physical aspects of Eq. (2.1); we suggest that the reader see Ref. 6. We only remark that the values ω_0 and α are given constants and $\omega_0 \neq \alpha$. The function $U(x,t)$ $[x=(x_1,x_2)]$ is a streamfunction, and the components of the velocity vector $\vec{v}$ of the fluid particles can be represented by this function in the following expressions:

$$v_1 = -U_{x_2}, \quad v_2 = U_{x_1}.$$

Let us consider the curve

$$\Gamma \equiv \{(x_1,x_2)\colon x_1 = x_1(s),\ x_2 = x_2(s),\ s\in[0,l]\},$$

which will be called curve Γ, in a fluid whose dynamics are described by Eq. (2.1), and orient curve Γ by setting its sides Γ^+ and Γ^- in the following way. Notice that we have no requirements for the size of the curve Γ (it can be a finite or infinite plane in the third dimension), because we consider only two-dimensional space. We denote the tangent vector at point $x(s) = (x_1(s),x_2(s))$ of the curve Γ by $\vec{\tau}_s$, and the normal vector at $x(s)\in\Gamma$ of the curve Γ by $\vec{n}_s$. If we rotate $\vec{\tau}_s$ for $\pi/2$ counterclockwise we shall obtain $\vec{n}_s$. We shall call the side of the curve Γ that we see by looking toward the vector $\vec{n}_s$ by Γ^+ and the opposite side of the curve Γ by Γ^-.

We assume that before time $t=0$ there was no movement of fluid and curve Γ. After time $t=0$, the pressure distributions on the two sides of the curve Γ are, in general, different. Mathematically it is equivalent to the prescription, on both sides $\Gamma^\pm$, of boundary conditions of the following uncommon kind (see, for example, Ref. 6) for the function $U(x,t)$:

$$(\mathscr{N}_{tx} U(x,t))\big|_{x=x(s)\in\Gamma^\pm} \equiv \left(\frac{\partial^2}{\partial t^2}\frac{\partial U}{\partial n_s} + \omega_0 \cos(\vec{n}_s,x_1)\frac{\partial U}{\partial x_1} + \alpha^2\cos(\vec{n}_s,x_2)\frac{\partial U}{\partial x_2}\right)\bigg|_{x=x(s)\in\Gamma^\pm} = \varphi_\pm(s,t). \tag{2.2}$$

Our assumptions require that the function $U(x,t)$ must satisfy the following initial conditions:

$$U(x,0) = U_t(x,0) = 0. \tag{2.3}$$

To select the unique solution to the problem we have to set, as in Ref. 5, the following conditions of regularity at infinity for the function $U(x,t)$:

$$|D_t^k U| \leqslant A_k(t)/|x|, \quad |D_t^k D_{x_j} U| \leqslant \widetilde{A}_k(t)/|x|^2, \tag{2.4}$$

for $|x| = (x_1^2+x_2^2)^{1/2} \to +\infty$, where

$$D_t^k U \equiv \frac{\partial^k}{\partial t^k} U, \quad K=1,2;$$

$$D_{x_j} U \equiv \frac{\partial}{\partial x_j} U, \quad j=1,2;$$

and $A_k(t)$ and $\tilde{A}_k(t)$ are continuous non-negative functions of t.

Since the geometry of the field has singular points at the ends of curve Γ, then we naturally assume that the function $U(x,t)$ or its gradient may have singularities in the neighborhoods of the end points of curve Γ. We obtain the following conditions in the neighborhoods of the ends of the curve (exactly as, for instance, in Ref. 1) by considering more closely the possible character of these singularities. The function $U(x,t)$ and its derivative $U_t(x,t)$ are bound in the neighborhoods of the end points of curve Γ. Other kinds of derivatives of this function, $D_{x_j}U(x,t), D_t^2 D_{x_j}U(x,t)$, behave like

$$|D_{x_j}U(x,t)|,\ |D_t^2 D_{x_j}U(x,t)| \sim O(r_{1,2}^{-1/2}), \tag{2.5}$$

where $j=1,2$ and $r_{1,2}$ is the distance to the ends of curve Γ.

Problem W: Find the continuous function $U(x,t)$ in the space for $t\geqslant 0$, which has continuous derivative $U_t(x,t)$ and satisfies Eq. (2.1) in the classical sense, in the space $R^2\setminus\Gamma$ with initial conditions (2.3), boundary conditions (2.2) on the sides of curve Γ, and conditions of regularity at infinity (2.4). Moreover, the function $U(x,t)$ must satisfy conditions (2.5) in the neighborhoods of the end points of curve Γ.

III. THE CLASSICAL SOLUTION TO PROBLEM W

To find the classical solution to problem W we need several important results, which will help us to obtain this solution and show some interesting properties of the potentials used to find explicit solutions to such problems.[1–5]

Let us give the following definitions. We shall say that a function $v(s)$ belongs to the $C_{1/2}^{(0,h)}(\Gamma)$, which is given on the curve Γ, if the function $d(s)v(s)\in C^{(0,h)}(\Gamma)$, where $d(s) \equiv |x(s)-x(0)|^{1/2}|x(s)-x(l)|^{1/2}$, $x(0)=(x_1(0),x_2(0))$, and $x(l)=(x_1(l),x_2(l))$. The points $x(0)$, $x(l)$ are end points of the curve Γ and $|x(s)-x(0)|$, $|x(s)-x(l)|$ are distances from point $x(s)=(x_1(s),x_2(s))$ to the end points of the curve Γ. We denote the sets of functions

$$C^{(0)}[0,T;C^{(0,h)}(\Gamma)] \equiv \{\mu(s,t)\in C^{(0)}[0,t;C^{(0,h)}(\Gamma)]:\ \mu(s,0)=\mu_t(s,0)=0\},$$

$$\bar{C}_0^{(2)}[0,T;C_{1/2}^{(0,h)}(\Gamma)] \equiv \left\{\mu(s,t)\in C^{(2)}[0,T;C_{1/2}^{(0,h)}(\Gamma)]:\ \int_{-1}^{1}\mu(\sigma,t)d\sigma=0,\ t\in[0,T]\right\}.$$

Let us consider the dynamic logarithmic potential and the angle potential for Eq. (2.1) (see Ref. 3):

$$V[\mu](x,t)=\int_\Gamma \mu(s,t)\ln|x-y(s)|ds+\int_0^t\int_\Gamma \mu(s,t-\tau)\times\frac{1}{\tau}\left[1-\cos\left[\frac{|x-y(s)|_n}{|x-y(s)|}\tau\right]\right]ds\,d\tau, \tag{3.1}$$

$$T[\nu](x,t)=\int_\Gamma \nu(s,t)\Psi(x,s)ds-\int_0^t\int_\Gamma \nu(s,t-\tau)\Phi[\Psi(x,s);\tau]ds\,d\tau, \tag{3.2}$$

where

$$|x|=(x_1^2+x_2^2)^{1/2},\quad |x|_*=(\alpha^2x_1^2+\omega_0^2x_2^2)^{1/2},$$

$$y=(y_1(s),y_2(s))\in\Gamma,$$

$$\Phi(\xi,t)=\int_0^\xi(\omega_0^2\sin^2\theta+\alpha^2\cos^2\theta)^{1/2}\times\sin\{\omega_0^2\sin^2\theta+\alpha^2\cos^2\theta)^{1/2}t\}d\theta,$$

and $\Psi(x,s)$ is the kernel of the angle potential, which is defined in the following way[7]:

$$\cos\Psi(x,s)=\frac{x_1-y_1(s)}{|x-y(s)|},\quad \sin\Psi(x,s)=\frac{x_2-y_2(s)}{|x-y(s)|}.$$

To assure single-valuedness of the function (3.2) we shall require, as in Ref. 3,

$$(1\cdot\nu)_{L_2(\Gamma)}=\int_\Gamma \nu(s,t)ds=0.$$

We assume that the curve $\Gamma\in A^{(1,\lambda)}$, $0<\lambda<1$.[8] We may easily prove the following lemma by using the results of the theory of dynamic potentials for Eq. (2.1) developed in Ref. 3.

Lemma 1: If

$$\Gamma\in A^{(1,\lambda)},\quad \nu(s,t),\mu(s,t)\in\bar{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)],$$

then the following hold.

(1) The potentials $T[\bar{\nu}](x,t)$, $V[\bar{\mu}](x,t)$ satisfy Eq. (2.1) in the field $R^2\setminus\Gamma$, the initial conditions (2.3), the conditions of regularity at infinity (2.4), and the conditions (2.5) in the neighborhood of the end points of the curve Γ, and are continuous in $R^2\setminus\Gamma$ (the potential $V[\bar{\mu}](x,t)$ is continuous in R^2).

(2) If the point $x(s)$ is not the end point of the curve Γ, then

$$\lim_{x\to x(s)\in\Gamma^\pm}\mathcal{N}_{tx}T[\bar{\mu}](x,t)=\frac{\partial}{\partial s}\bar{V}[\mu](x,t)=\int_\Gamma\mu(\sigma,t)\frac{\sin\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma+(\alpha^2-\omega_0^2)\int_0^t\int_\Gamma\mu(\sigma,t-\tau)\cos\Psi(s,\sigma)\times\sin\Psi(s,\sigma)\cdot\sin\{\tau[\omega_0^2\sin^2\Psi(s,\sigma)+\alpha^2\cos^2\Psi(s,\sigma)]^{1/2}\}\frac{\cos\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma\,d\tau,$$

where $\overline{V}[\mu](x,t)$ is the value of the potential $V[\mu]$ on the curve Γ, $\theta(s,\sigma)$ is the angle measured counterclockwise between the vectors $\vec{n}_s$ and $\overrightarrow{x(s)y(\sigma)}$, to the point

$$|x(s)-y(\sigma)|\cdot\sin\theta(s,\sigma) = -\tau_s\overrightarrow{x(s)y(\sigma)}$$

and

$$\bar{\mu}(s,t)=\int_0^t (t-\tau)\mu(s,\tau)d\tau,$$

$$\bar{\nu}(s,t)=\int_0^t (t-\tau)\cdot\nu(s,\tau)d\tau.$$

Later we shall need one more result, which we formulate in a kind of lemma.

Lemma 2: If

$$\Gamma\in A^{(1,\lambda)},\quad \nu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)],$$

and a point $x(s)$ is not the end point of the curve Γ, then

$$\lim_{x\to x(s)\in\Gamma^\pm}\mathcal{N}_{tx}V[\bar{\nu}](x,t)$$
$$=-\lim_{x\to x(s)\in\Gamma^\pm}\frac{\partial}{\partial\tau_s}T[\tilde{\nu}](x,t)$$
$$=\pm\pi(E-\omega_0S_{\omega_0t^*})(E-\alpha S_{\alpha t^*})\nu(s,t)$$
$$+D[\tilde{\nu}](s,t),$$

where

$$D[\tilde{\nu}](s,t)=\int_\Gamma\tilde{\nu}(\sigma,t)\frac{\cos\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma$$
$$-\int_0^t\int_\Gamma\tilde{\nu}(\sigma,t-\tau)\frac{|x(s)-y(\sigma)|_*}{|x(s)-y(\sigma)|}$$
$$\times\sin\left(\tau\frac{|x(s)-y(\sigma)|_*}{|x(s)-y(\sigma)|}\right)$$
$$\times\frac{\cos\theta(s,\sigma)}{|x(s)-y(\sigma)|}d\sigma\,d\tau,$$

$$\tilde{\nu}(s,t)=\left(\frac{\partial^2}{\partial t^2}+\omega_0^2\right)\left(\frac{\partial^2}{\partial t^2}+\alpha^2\right)\int_0^t(t-\tau)\bar{\bar{\nu}}(s,\tau)d\tau,$$

and $(E-\beta S_{\beta t^*})$ is an operator defined by the expressions

$$(E-\beta S_{\beta t^*})\nu(t)=\nu(t)-\beta\int_0^t S(\beta\cdot(t-\tau))\cdot\nu(\tau)d\tau,$$

$$S(\beta t)=-\int_0^{\beta t}\frac{J_1(\xi)}{\xi}d\xi, \tag{3.3}$$

where $J_1(\xi)$ is the Bessel function of the first order.

Proof: Let us consider the system of equations

$$\left(\frac{\partial^2}{\partial t^2}+\omega_0^2\right)U_{x_1}=-v_{x_2},\quad \left(\frac{\partial^2}{\partial t^2}+\alpha^2\right)U_{x_2}=v_{x_1}. \tag{3.4}$$

This system was used in Ref. 3 for the construction of the dynamic angle potential and plays the same role for Eq. (2.1) as the Cauchy–Riemann system for the Laplace equation. One can show by direct calculation that the functions

$$v=V[\bar{\nu}](x,t),\quad U=T[\nu_0](x,t),$$

$$\left\{\nu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)],\right.$$
$$\left.\nu_0(s,t)=\int_0^t(t-\tau)\bar{\bar{\nu}}(s,\tau)d\tau\right\},$$

satisfy the system of equations (3.3) in $R^2\setminus\Gamma$. It can be shown by using this fact that, for arbitrary $x\in R^2\setminus\Gamma$,

$$\mathcal{N}_{tx}V[\bar{\nu}](x,t)=-\frac{\partial}{\partial\tau_s}T[\tilde{\nu}](x,t).$$

It is important for later consideration that

$$\tilde{\nu}(s,t)\in C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)].$$

We use that and the results of Ref. 3 to obtain the formula

$$\lim_{x\to x(s)\in\Gamma^\pm}\mathcal{N}_{tx}V[\bar{\nu}](x,t)$$
$$=-\lim_{x\to x(s)\in\Gamma^\pm}\frac{\partial}{\partial\tau_s}T[\tilde{\nu}](x,t)$$
$$=\pm\pi(E-\omega_0J_{\omega_0t^*})(E-\alpha J_{\alpha t^*})\tilde{\nu}(s,t)+D[\tilde{\nu}].$$

One can show by using the Laplace transformation with respect to t that

$$\pm\pi(E-\omega_0J_{\omega_0t^*})(E-\alpha J_{\alpha t^*})\tilde{\nu}(s,t)$$
$$=\pm(E-\omega_0S_{\omega_0t^*})(E-\alpha S_{\alpha t^*})\nu(s,t),$$

where the operator $(E-\beta S_{\beta t^*})$ is the defined by formula (3.3) and

$$(E-\beta J_{\beta t^*})\nu(t)=\nu(t)-\beta\int_0^t J_1(\beta(t-\tau))\nu(\tau)d\tau.$$

The lemma has thus been proved.

Lemmas 1 and 2 show some symmetry properties of the operators $V(x,t)$ and $T(x,t)$.

Let us make several remarks. Everywhere later we shall assume that $\Gamma\in A^{(2,\lambda)}$ and the functions $\varphi_\pm(s,t)$ in the boundary conditions (2.2) belong to $C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)]$ and satisfy the following condition of the correspondence:

$$\int_\Gamma[\varphi_+(s,t)-\varphi_-(s,t)]ds=0,\quad t\in[0,\infty). \tag{3.5}$$

We shall look for a solution to problem W of the kind

$$U(x,t)=V[\bar{\nu}](x,t)+T[\bar{\mu}](x,t), \tag{3.6}$$

where $\nu(s,t),\mu(s,t)\in\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$. According to Lemma 1 the function $U(x,t)$ satisfies all the conditions of problem W except the boundary conditions (2.2). We obtain the following system of integral equations for functions μ and ν by using Lemmas 1 and 2:

$$\begin{aligned}&\frac{\partial}{\partial s}\overline{V}[\mu](s,t)+\pi(E-\omega_0S_{\omega_0t^*})(E-\alpha S_{\alpha t^*})\nu(s,t)\\&\quad+D[\tilde{\nu}](s,t)=\varphi_+(s,t),\\&\frac{\partial}{\partial s}\overline{V}[\mu](s,t)-\pi(E-\omega_0S_{\omega_0t^*})(E-\alpha S_{\alpha t^*})\nu(s,t)\\&\quad+D[\tilde{\nu}](s,t)=\varphi_-(s,t).\end{aligned} \tag{3.7}$$

By adding and subtracting Eqs. (3.7) we may obtain

$$(E-\omega_0S_{\omega_0t^*})(E-\alpha S_{\alpha t^*})\nu(s,t)$$
$$=(1/2\pi)(\varphi_+(s,t)-\varphi_-(s,t)), \tag{3.8}$$

$$\frac{\partial}{\partial s}\overline{V}[\mu](s,t)=\frac{1}{2}[\varphi_+(s,t)+\varphi_-(s,t)]-D[\tilde{\nu}](s,t). \tag{3.9}$$

We can find the explicit solution to Eq. (3.8). It is

$$v(s,t) = (1/2\pi)(E - \omega_0 J_{\omega_0 t^*})(E - \alpha J_{\alpha t^*}) \times (\varphi_+(s,t) - \varphi_-(s,t)).$$

This solution belongs to the $\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,\alpha)}(\Gamma)]$, because the functions $\varphi_\pm(s,t)\in C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)]$, therefore the function

$$\tilde{v}(s,t)\in C^{(0)}[0,\infty;C^{(0,\alpha)}(\Gamma)].$$

Notice that the operators $(E - \omega_0 J_{\omega_0 t^*})\cdot(E - \alpha J_{\alpha t^*})$ and $(E - \omega_0 S_{\omega_0 t^*})(E - \alpha S_{\alpha t^*})$ appear often in the solutions to problems of gravity-gyroscopic waves.

One can show by using the earlier representation of the operator $D[\tilde{v}]$ and $\Gamma\in A^{(2,\lambda)}$ that, for arbitrary functions $\eta(s,t)\in C^{(0)}[0,\infty;C^{(0,h)}(\Gamma)]$,

$$D[\eta](s,t)\in C^{(0)}[0,\infty;C^{(0,\lambda)}(\Gamma)].$$

Thus we reduce the problem of classical solvability to the problem of the solvability of Eq. (3.8), which has the right side from $C^{(0)}[0,\infty;C^{(0,\gamma)}(\Gamma)]$, $\gamma = \min\{\alpha,\lambda\}$ in the set of functions $\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$. This equation was carefully considered in Ref. 4. We therefore shall not repeat this work, but we shall formulate the final result.

Lemma 3: Equation (3.9) has a unique solution from the set of functions $\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$ for an arbitrary right side chosen from the set of functions $C^{(0)}[0,\infty;C^{(0,\gamma)}(\Gamma)]$.

In summary, as a result of all the lemmas we obtain the following theorem.

Theorem 1: Problem W has the classic solution (3.6), where

$$v(s,t) = (1/2\pi)(E - \omega_0 J_{\omega_0 t^*})(E - \alpha J_{\alpha t^*}) \times [\varphi_+(s,t) - \varphi_-(s,t)]$$

and $\mu(s,t)$ is the solution of Eq. (3.9) from the set of functions $\overline{C}_0^{(2)}[0,\infty;C_{1/2}^{(0,h)}(\Gamma)]$ for arbitrary $v_\pm(s,t)\in C^{(0)}[0,\infty;C^{(0,\alpha)}(\Gamma)]$, which satisfy the conditions of correspondence (3.5).

Let us consider the question of uniqueness of the solution (3.6). We may obtain the energetic relation for the equation (2.1) by the product of Eq. (2.1) and U_t and by carrying out the integration of some compact field D in R^2, which has the smooth boundary ∂D:

$$\frac{\partial}{\partial t}\left\{\frac{1}{2}\|\nabla U_t\|^2_{L_2(D)} + \frac{\omega_0^2}{2}\|U_{x_1}\|^2_{L_2(D)} + \frac{\alpha^2}{2}\|U_{x_2}\|^2_{L_2(D)}\right\} = \int_{\partial D}(\mathcal{N}_{tx}\cdot U)U_t\,d(\partial D),$$

where $\vec{n}$ in the expression for the $\mathcal{N}_{tx}$ is the external normal vector to the boundary of the field D.

Following Ref. 4 we obtain the next theorem.

Theorem 2: Solution (3.6) to problem W is the unique solution.

IV. ANALYSIS OF THE RESULTS

We should note that we have considered the general form of the curve Γ. If some scientists or engineers use the results of this paper and consider the specific forms of the curve, then problem W can be solved more easily.

For example, if we replace curve Γ by the line

$$\Gamma_0 \equiv \{(x_1,x_2):\ x_1 = s\cdot\cos\varphi,\ x_2 = s\cdot\sin\varphi,\ -1\leqslant s\leqslant 1\},\quad \varphi\in[0,\pi/2],$$

where φ is angle between the axis Ox_1 and the line Γ_0 then the second equation of the system (3.8) and (3.9) may be written in the form

$$\int_{-1}^{1}\frac{\mu(\sigma,t)}{\sigma - s}\,d\sigma = -\frac{1}{2}[\varphi_+(s,t) + \varphi_-(s,t)]. \quad (4.1)$$

This equation was studied in Refs. 1 and 2 and has the explicit unique solution, in the set of the functions $\overline{C}_0^{(2)}[0,\infty; C_{1/2}^{(0,h)}(\Gamma)]$,

$$\mu(s,t) = \frac{1}{2\pi}(1 - s^2)^{-1/2}\int_{-1}^{1}\frac{(1-\xi^2)^{1/2}}{\xi - s} \times [\varphi_+(s,t) + \varphi_-(s,t)]d\xi. \quad (4.2)$$

One can prove the following theorem by using our remark, Theorem 1, and Theorem 2.

Theorem 3: In the case where curve Γ is replaced by the line Γ_0, problem W has the explicit unique solution defined by the expression

$$U(x,t) = V[\bar{v}](x,t) + T[\bar{\mu}](x,t), \quad (4.3)$$

where

$$v(s,t) = (1/2\pi)(E - \omega_0 J_{\omega_0 t^*})(E - \alpha J_{\alpha t^*}) \times [\varphi_+(s,t) - \varphi_-(s,t)]$$

and $\mu(s,t)$ is defined by the expression (4.2), for arbitrary

$$\varphi_\pm(s,t)\in C^{(0)}[0,\infty;C^{(0,k)}(\Gamma)].$$

ACKNOWLEDGMENTS

The authors are very grateful to S. A. Gabov, Professor of Moscow State University, and to H. T. Williams, Professor of Washington and Lee University, for their attention to this work.

[1]S. A. Gabov "A solution to the problem of dynamics of the stratified fluid," Akad. Nauk SSSR. Zh. Vych. Mat. Mat. Fiz. **25**, 718 (1985).
[2]S. A. Gabov and U. D. Pletner, "An initial-value problem for the equation of the gravity-gyroscopic waves," Akad. Nauk SSSR. Zh. Vych. Mat. Mat. Fiz. **25**, 1689 (1985).
[3]S. A. Gabov and U. D. Pletner, "An equation of the gravity-gyroscopic waves," Akad. Nauk SSSR. Zh. Vych. Mat. Mat. Fiz. **27**, 102 (1987).
[4]S. A. Gabov and U. D. Pletner, "Solvability of the initial-value problem for the equation of the gravity-gyroscopic waves." Akad. Nauk SSSR Zh. Vych. Mat. Mat. Fiz. **27**, 711 (1987).
[5]U. D. Pletner, "A problem of the gravity-gyroscopic waves excited by the oscillations of the line segment," Akad. Nauk SSSR. Zh. Vych. Mat. Mat. Fiz. **28**, 524 (1988).
[6]S. A. Gabov and A. G. Sveshnikov, *Problems of the Dynamics of Stratified Fluids* (Nauka, Moscow, 1986).
[7]S. A. Gabov, "The angular potential," Dokl. Akad. Nauk SSSR **278**, 527 (1984).
[8]C. Miranda, *Partial Differential Equations of Elliptic Type*, translated by Z. C. Motteler (Springer, New York, 1970).

Conserved quantities and symmetries of KP hierarchy

J. Matsukidaira and J. Satsuma
Department of Applied Physics, Faculty of Engineering, University of Tokyo, Tokyo 113, Japan
W. Strampp
Fachbereich 17-Mathematik, Gesamthochschule Kassel, Nora-Platiel-Strasse 1, 3500 Kassel, Federal Republic of Germany

(Received 23 May 1989; accepted for publication 17 January 1990)

Conserved quantities and symmetries of the KP equation from the point of view of the Sato theory that provides a unifying approach to soliton equations is studied. Conserved quantities are derived from the generalized Lax equations. Some reductions of the KP hierarchy such as KdV, Boussinesq, a coupled KdV, and Sawada–Kotera equation are also considered. By expansion of the squared eigenfunctions of the Lax equations in terms of the τ function, symmetries of the KP equations are obtained. The relationship of this procedure to the two-dimensional recursion operator newly found by Fokas and Santini is discussed.

I. INTRODUCTION

The discovery of the inverse scattering transform (IST) for the Korteweg–deVries (KdV) equation was a big breakthrough in the analysis of nonlinear evolution equations.[1] Since then, many soliton equations have been revealed to be exactly solvable by the method. A key step of the finding was in the calculation of the conserved quantities of the equation. Miura succeeded in proving the existence of an infinite number of conserved quantities by using the transformation between the KdV and the modified KdV equations, which is now called the Miura transformation.[2] The transformation actually gave a hint to derive the eigenvalue problem of the IST for the KdV equation.

Conserved quantities are closely related to symmetries of equations. The existence of an infinite number of conserved quantities or symmetries is a widely accepted definition of complete integrability of equations.[3] We now know that most of the soliton equations possess such properties. Extension of the concept to equations in the higher-dimensional case has also been done. Fokas and Santini proved the existence of the recursion operator which generates infinitely many symmetries for the Kadomtsev–Petviashvili (KP) equation.[4–6]

Besides the IST, there are several analytical methods for obtaining solutions of soliton equations. In Hirota's method, we transform an equation into a bilinear form, from which we can get soliton solutions successively by means of a kind of perturbational technique. The Bäcklund transformation is also employed to obtain solutions from a known solution of the concerned equation. The existence of such analytical methods reflects a rich algebraic structure of soliton equations. In 1981, based on algebraic analysis, Sato presented a theory that provides a unified description of the soliton equations.[7–9] We call it the Sato theory hereafter. The origin of the Sato theory was the discovery that there is a bijection between a class of microdifferential operators and the solution space of soliton equations that is seen as a Grassmann manifold. Time development of the coefficients of microdifferential operators are governed either by the Sato equation or by the generalized Lax equations. The Sato equation is solved by means of the τ function, which closely relates to the representation theory of groups. It satisfies a certain class of bilinear equations obtained from Plücker's relations. The equations are called the KP hierarchy, in which the KP equation is the simplest one.

The Sato theory also clarifies the relationship among the IST, Hirota's method, and the Bäcklund transformation. It may be expected that the theory is also a powerful tool to understand the algebraic structure of soliton equations. Motivated by this expectation, we investigate the symmetry properties of soliton equations based on the Sato theory.

In Sec. II, we give a brief introduction of the Sato theory. We first present the generalized Lax equation, which is written by a microdifferential operator. We see that the eigenfunction of the linear system can be expressed by the τ function, which satisfies all of the KP hierarchy. We then briefly mention the reduction procedure of the hierarchy. The introduction of infinitely many time variables in the Sato theory helps to understand the existence of an infinite number of conserved quantities of soliton equations. In Sec. III, we give the definition of symmetries, recursion operators, and conserved quantities of nonlinear evolution equations in one-spatial and one-temporal dimensions. We also refer to the extension of these quantities to the higher-dimensional case. In Sec. IV, we show that the conservation laws for the KP hierarchy are naturally derived from the generalized Lax equations. If we perform suitable reductions of the KP hierarchy, we obtain a series of equations such as the KdV, the Boussinesq, a coupled KdV, and the Sawada–Kotera equations. The conservation laws for these equations are also obtained by the reduction procedure.

As mentioned before, the existence of the conserved quantities is closely related to that of the symmetries. For the KdV equation, the squared eigenfunction and their corresponding linear operator play an important role to obtain symmetries. We show in Sec. V that the same situation holds for the KP equation. The squared eigenfunction can be expressed as a series expansion whose coefficients are differential polynomials obtained from the microdifferential operator. From this fact we can get a series of commuting symmetries of the KP equation. The existence of an infinite

 0022-2488/90/061426-09$03.00

number of symmetries indicates that of the recursion operator. Finally in Sec. VI, we discuss the relationship between our result and the recursion scheme for symmetries of the KP equation which has been obtained by Fokas and Santini.

II. SATO THEORY

Let us introduce a microdifferential operator,

$$L = \partial + u_2 \partial^{-1} + u_3 \partial^{-2} + u_4 \partial^{-3} + \cdots, \tag{2.1}$$

where ∂ denotes ∂/∂_x, and u_n $n = 2,3,\ldots$, are functions in x and infinitely many time variables $t = (t_1,t_2,t_3,t_4,\ldots)$. It is noted that t_1 is identified with x. We define B_n as the differential part of L^n. For example,

$$B_1 = \partial, \tag{2.2a}$$

$$B_2 = \partial^2 + 2u_2, \tag{2.2b}$$

$$B_3 = \partial^3 + 3u_2\partial + 3u_3 + 3u_{2,x}, \tag{2.2c}$$

$$B_4 = \partial^4 + 4u_2\partial^2 + (4u_3 + 6u_{2,x})\partial + 4u_4 + 6u_{3,x} + 4u_{2,xx} + 6u_2^2, \tag{2.2d}$$

where subscript x denotes partial differentiation in x.

Consider a system of linear equations for an eigenfunction ψ,

$$L\psi = \lambda\psi, \tag{2.3}$$

$$\frac{\partial}{\partial t_n}\psi = B_n\psi, \quad n = 1,2,\ldots . \tag{2.4}$$

From the compatibility condition of Eqs. (2.3) and (2.4), we have

$$\frac{\partial L}{\partial t_n} = [B_n,L] = B_nL - LB_n, \tag{2.5}$$

or equivalently

$$\frac{\partial B_m}{\partial t_n} - \frac{\partial B_n}{\partial t_m} = [B_n,B_m]. \tag{2.6}$$

The KP hierarchy are obtained from Eqs. (2.5) or (2.6). Especially, if $n = 2$ and $m = 3$ are taken, Eq. (2.6) gives the KP equation itself,

$$\frac{\partial}{\partial x}\left(\frac{\partial u_2}{\partial t_3} - \frac{1}{4}\frac{\partial^3 u_2}{\partial x^3} - 3u_2\frac{\partial u_2}{\partial x}\right) - \frac{3}{4}\frac{\partial^2 u_2}{\partial t_2^2} = 0. \tag{2.7}$$

The linear system, Eqs. (2.3) and (2.4), has a formal solution of the form,

$$\psi = \left(1 + \sum_{j=1}^{\infty} w_j\lambda^{-j}\right)\exp\xi(t,\lambda), \tag{2.8}$$

where

$$\xi(t,\lambda) \equiv \sum_{n=1}^{\infty} t_n\lambda^n, \tag{2.9}$$

and the w_j's are related to the u_j's as

$$u_2 = -w_{1,x}, \tag{2.10a}$$

$$u_3 = -w_{2,x} + w_1w_{1,x}, \tag{2.10b}$$

$$u_4 = -w_{3,x} + w_1w_{2,x} + w_{1,x}w_2 - w_1^2w_{1,x} - w_{1,x}^2. \tag{2.10c}$$

From the theory of solution space of the KP hierarchy, i.e., the theory of the Grassmann manifold and τ function, it is shown that w_j is expressed as

$$w_j = [1/\tau(t)]p_j(-\tilde{\partial})\tau(t), \tag{2.11}$$

where $p_j, j = 1,2,\ldots$, are polynomials defined by

$$\exp\left(\sum_{n=1}^{\infty} t_n\lambda^n\right) = \sum_{j=0}^{\infty} p_j(t)\lambda^j, \tag{2.12}$$

and $\tilde{\partial}$ is a differential operator given by

$$\tilde{\partial} = \left(\frac{\partial}{\partial t_1}, \frac{1}{2}\frac{\partial}{\partial t_2}, \frac{1}{3}\frac{\partial}{\partial t_3}, \ldots\right). \tag{2.13}$$

Then we find that the eigenfunction can be written in terms of the τ function as

$$\psi = \frac{\tau(t_1 - 1/\lambda, t_2 - 1/2\lambda^2,\ldots)}{\tau(t_1,t_2,\ldots)}\exp\xi(t,\lambda). \tag{2.14}$$

From Eqs. (2.4), (2.10), and (2.14), we see that the u_n are also written in τ and its derivatives; for example,

$$u_2 = \frac{\partial^2}{\partial x^2}\log\tau, \tag{2.15a}$$

$$u_3 = \frac{1}{2}\left(\frac{\partial^2}{\partial x\,\partial t_2} - \frac{\partial^3}{\partial x^3}\right)\log\tau. \tag{2.15b}$$

Thus the KP hierarchy can be rewritten into a set of nonlinear differential equations for a single function τ. According to Ref. 10, the τ function satisfies

$$\sum_{n=0}^{\infty} p_n(-2y)p_{n+1}(\tilde{D})\exp\left(\sum_{i=0}^{\infty} y_iD_i\right)\tau\cdot\tau = 0, \tag{2.16}$$

for any $y = (y_1,y_2,y_3,\ldots)$, where we have used Hirota's operators defined by

$$D_j^m a\cdot b = \left(\frac{\partial}{\partial t_j} - \frac{\partial}{\partial t_j'}\right)^m a(t)b(t')|_{t=t'} = \frac{\partial^m}{\partial s_j^m}a(t_j + s_j)b(t_j - s_j)|_{s_j=0}, \tag{2.17}$$

and

$$\tilde{D} = \left(D_1, \frac{1}{2}D_2, \ldots, \frac{1}{n}D_n, \ldots\right). \tag{2.18}$$

Equation (2.16) includes

$$(D_1^4 + 3D_2^2 - 4D_1D_3)\tau\cdot\tau = 0, \tag{2.19}$$

which is the bilinear form of the KP equation.

Instead of the linear problem for ψ, we can consider that for the adjoint wave function ψ^*, which is given by

$$\psi^* = \left(1 + \sum_{j=1}^{\infty} w_j^*\lambda^{-j}\right)\exp\{-\xi(t,\lambda)\}, \tag{2.20}$$

with

$$w_j^* = (1/\tau)p_j(\tilde{\partial})\tau, \tag{2.21}$$

or

$$\psi^* = \frac{\tau(t_1 + 1/\lambda, t_2 + 1/2\lambda^2,\ldots)}{\tau(t_1,t_2,\ldots)}\exp\{-\xi(t,\lambda)\}. \tag{2.22}$$

The linear problem for ψ^* is written as

$$L^*\psi^* = \lambda\psi^*, \tag{2.23}$$

$$\frac{\partial}{\partial t_n}\psi^* = -B_n^*\psi^*, \tag{2.24}$$

where L^* is adjoint of L and B_n^* is the differential part of $(L^*)^n$. If we demand that τ does not depend on $t_\ell, t_{2\ell}, t_{3\ell},\ldots$ for a positive integer ℓ, the system of nonlinear equations given by the compatibility condition of the linear problem (2.3) and (2.4) is called the ℓreduction of the KP hierarchy. In this case, ψ satisfies

$$\frac{\partial}{\partial t_\ell}\psi = \lambda^\ell\psi, \tag{2.25}$$

$$B_\ell\psi = \lambda^\ell\psi. \tag{2.26}$$

The two-reduction includes the KdV equation

$$\frac{\partial u_2}{\partial t_3} - \frac{1}{4}\frac{\partial^3 u_2}{\partial x^3} - 3u_2\frac{\partial u_2}{\partial x} = 0, \tag{2.27}$$

and the three-reduction does the Boussinesq-like equation

$$3\frac{\partial^2 u_2}{\partial t_2^2} + \frac{\partial^4 u_2}{\partial x^4} + 6\frac{\partial^2 u_2}{\partial x^2} = 0, \tag{2.28}$$

which reduces to the Boussinesq equation by means of a suitable variable transformation. Moreover, the four-reduction includes

$$\frac{\partial u_2}{\partial t_3} - \frac{\partial^3 u_2}{\partial x^3} - 3\frac{\partial^2 u_3}{\partial x^2} - 3\frac{\partial u_4}{\partial x} - 6u_2\frac{\partial u_2}{\partial x} = 0, \tag{2.29a}$$

$$\frac{\partial u_3}{\partial t_3} + \frac{3}{4}\frac{\partial^4 u_2}{\partial x^4} + 2\frac{\partial^3 u_3}{\partial x^3} + \frac{3}{2}\frac{\partial^2 u_4}{\partial x^2} + \frac{9}{2}u_2\frac{\partial^2 u_2}{\partial x^2} + \frac{9}{2}\left(\frac{\partial u_2}{\partial x}\right)^2 + 3u_2\frac{\partial u_3}{\partial x} + 3\frac{\partial u_2}{\partial x}u_3 = 0, \tag{2.29b}$$

$$\frac{\partial u_4}{\partial t_3} - \frac{3}{8}\frac{\partial^5 u_2}{\partial x^5} - \frac{3}{4}\frac{\partial^4 u_3}{\partial x^4} - \frac{1}{4}\frac{\partial^3 u_4}{\partial x^3} - \frac{3}{4}u_2\frac{\partial^3 u_2}{\partial x^3} - \frac{27}{4}\frac{\partial u_2}{\partial x}\frac{\partial^2 u_2}{\partial x^2} + 3u_2\frac{\partial^2 u_3}{\partial x^2} - \frac{9}{2}\frac{\partial u_2}{\partial x}\frac{\partial u_3}{\partial x} - \frac{3}{2}\times\frac{\partial^2 u_2}{\partial x^2}u_3 + 6u_2\frac{\partial u_4}{\partial x} + 3u_3\frac{\partial u_3}{\partial x} + 9u_2^2\frac{\partial u_2}{\partial x} = 0, \tag{2.29c}$$

which reduce to a coupled KdV equation[11]

$$\frac{\partial u}{\partial t_3} = \frac{1}{4}\frac{\partial^3 u}{\partial x^3} + 3u\frac{\partial u}{\partial x} + 3\frac{\partial}{\partial x}(-\phi^2 + \omega), \tag{2.30a}$$

$$\frac{\partial \phi}{\partial t_3} = -\frac{1}{2}\frac{\partial^3 \phi}{\partial x^3} - 3u\frac{\partial \phi}{\partial x}, \tag{2.30b}$$

$$\frac{\partial \omega}{\partial t_3} = -\frac{1}{2}\frac{\partial^3 \omega}{\partial x^3} - 3u\frac{\partial \omega}{\partial x}, \tag{2.30c}$$

by rewriting

$$u_2 = u, \tag{2.31a}$$

$$u_3 = -\frac{1}{2}\frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x}, \tag{2.31b}$$

$$u_4 = \frac{1}{4}\frac{\partial^2 u}{\partial x^2} - \frac{1}{2}u^2 - \frac{\partial^2 \phi}{\partial x^2} - \phi^2 + \omega. \tag{2.31c}$$

There exists another hierarchy which is called the BKP hierarchy.[10] The BKP hierarchy is the system of nonlinear differential equations obtained from the linear problem which has the same form as Eqs. (2.3) and (2.4), but includes only the odd time variables t_1, t_3, t_5,... and is imposed by the constraint that the constant terms of B_n for $n = 1,3,5,\ldots$ should vanish. The three-reduction of the BKP hierarchy includes the Sawada–Kotera equation,

$$\frac{\partial u_2}{\partial x_5} = -\frac{1}{9}u_{2,xxxxx} - \frac{5}{3}u_2u_{2,xxx} - \frac{5}{3}u_{2,x}u_{2,xx} - 5u_2^2u_{2,x}. \tag{2.32}$$

For further details of the Sato theory, a reader may refer to Ref. 9.

III. CONSERVED QUANTITIES AND SYMMETRIES

We here give a brief survey of conserved quantities and symmetries of the 1 + 1 (say x and t)-dimensional nonlinear evolution equations. Consider an evolution equation,

$$u_t = K(u), \tag{3.1}$$

where K is a functional of u. The following equation is called the linearized equation of Eq. (3.1):

$$S_t = K'(u)[S], \tag{3.2}$$

where $K'(u)[S]$ means the Fréchet derivative of K at the point u in the direction of S, i.e.,

$$K'(u)[S] = \frac{\partial}{\partial\epsilon}K(u+\epsilon S)\Big|_{\epsilon=0}. \tag{3.3}$$

A functional $S(u)$ satisfying Eq. (3.2) is called a symmetry. From the equation

$$S_t = S'[u_t], \tag{3.4}$$

it follows that a symmetry S must satisfy

$$[S,K] \equiv S'[K] - K'[S] = 0. \tag{3.5}$$

An operator satisfying

$$R'[K] + [R,K'] = 0 \tag{3.6}$$

is called a recursion operator. It is easily shown that recursion operators maps symmetries into symmetries.

A functional I is a conserved quantity, iff

$$I_t = 0 \tag{3.7}$$

or

$$I'[u_t] = I'[K] = \langle \operatorname{grad} I, K\rangle = 0, \tag{3.8}$$

where

$$\langle f,g\rangle \equiv \int fg\,dx. \tag{3.9}$$

Differentiating this equation in the arbitrary direction v, it follows that γ is the gradient of a conserved quantity I iff

$$\gamma[K] + K'^*[\gamma] = 0 \tag{3.10}$$

and

$$\gamma' = \gamma'^*, \tag{3.11}$$

where the asterisk denotes the adjoint. The functional γ is called a conserved covariant. The adjoint R^* of R, which is often called the squared eigenfunction operator, maps conserved covariants into conserved covariants. The quantity ρ satisfying

$$\rho_t = \operatorname{div} J, \tag{3.12}$$

is called a conserved density. Obviously, $\int \rho\, dx$ is a conserved quantity. For further details of symmetries and conserved quantities, a reader may refer to Ref. 3. The recursion operator plays an important role in the theory for equations in one time and one spatial variable $(1+1)$. This motivated the research for recursion operators for equations in one time and two spatial variables $(2+1)$. Fokas and Santini succeeded in obtaining it for the KP equation.[4–6] They considered the equation as a reduction of a $3+1$ system, i.e., a system in the variables x, t, y_1, y_2. The notions of symmetries, conserved covariants, and recursion operator are generalized to the extended $3+1$ case by introducing a new suitable bilinear form and a directional derivative for $3+1$ quantities. Then they discovered an extended recursion operator mapping symmetries to symmetries. The adjoint of the recursion operator in the extended sense maps conserved covariants into conserved covariants. Finally, by taking the limit $y_2 \to y_1$ of extended symmetries and conserved covariants, they obtained symmetries and conserved covariants of the KP equation in the usual sense, which satisfy Eqs. (3.2) and (3.10), respectively.

IV. CONSERVED DENSITIES OF KP HIERARCHY

Let us expand $\partial(=\partial/\partial x)$ in powers of the microdifferential operator L,

$$\partial = L + \sigma_1^{(1)} L^{-1} + \sigma_2^{(1)} L^{-2} + \sigma_3^{(1)} L^{-3} + \cdots . \tag{4.1}$$

The coefficients $\sigma_j^{(1)}$ are determined by comparing Eq. (4.1) with Eq. (2.1). In the Appendix, the list of $\sigma_j^{(1)}$ is given for $1 \leqslant j \leqslant 7$. Applying Eq. (4.1) on the eigenfunction ψ and using Eq. (2.3), we have

$$\frac{\partial \psi}{\partial x} = \lambda \psi + \sum_{j=1}^{\infty} \frac{\sigma_j^{(1)} \psi}{\lambda^j}, \tag{4.2}$$

which gives

$$\frac{\partial}{\partial x} \log \psi = \lambda + \sum_{j=1}^{\infty} \frac{\sigma_j^{(1)}}{\lambda^j}. \tag{4.3}$$

We now define $\sigma^{(1)}$ by

$$\sigma^{(1)} \equiv \sum_{j=1}^{\infty} \frac{\sigma_j^{(1)}}{\lambda^j} = \frac{\partial}{\partial x} \log \psi - \lambda. \tag{4.4}$$

Differentiating Eq. (4.4) with respect to one of the time variables, say t_n, we obtain the formula of conservation law,

$$\frac{\partial \sigma^{(1)}}{\partial t_n} = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial t_n} \log \psi \right). \tag{4.5}$$

Each $\sigma_j^{(1)}$ gives a conserved density of the KP hierarchy. We shall show later that the conserved densities of several soliton equations are derived from $\sigma_j^{(1)}$ through the reduction procedure.

We may consider conservation laws in the other directions. Let us expand B_m in powers of L,

$$B_m = L^m + \sigma_1^{(m)} L^{-1} + \sigma_2^{(m)} L^{-2} + \sigma_3^{(m)} L^{-3} + \cdots . \tag{4.6}$$

Applying Eq. (4.6) on ψ, we obtain

$$\frac{\partial \psi}{\partial t_m} = \lambda^m \psi + \sum_{j=1}^{\infty} \frac{\sigma_j^{(m)} \psi}{\lambda^j}, \tag{4.7}$$

from which we find that

$$\sigma^{(m)} = \sum_{j=1}^{\infty} \frac{\sigma_j^{(m)}}{\lambda^j} = \frac{\partial}{\partial t_m} \log \psi - \lambda^m \tag{4.8}$$

satisfies the conservation law

$$\frac{\partial \sigma^{(m)}}{\partial t_n} = \frac{\partial}{\partial t_m} \left(\frac{\partial}{\partial t_n} \log \psi \right). \tag{4.9}$$

In the Appendix, we give the list of $\sigma_j^{(m)}$ for $2 \leqslant m \leqslant 5$ and $1 \leqslant j \leqslant 8-m$.

It is possible to express $\sigma_j^{(m)}$ in a compact form by means of the τ function. Substituting Eq. (2.14) into Eq. (4.8), we obtain

$$\begin{aligned} \sigma^{(m)} &= \frac{\partial}{\partial t_m} \left\{ \log \tau \left(t_1 - \frac{1}{\lambda}, t_2 - \frac{1}{2\lambda^2}, \dots \right) \right. \\ &\quad \left. - \log \tau(t_1, t_2, \dots) + \sum_{\ell=1}^{\infty} t_\ell \lambda^\ell \right\} - \lambda^m \\ &= \sum_{j=1}^{\infty} \frac{\partial}{\partial t_m} \frac{\{ p_j(-\tilde{\partial}) \log \tau \}}{\lambda^j}, \end{aligned} \tag{4.10}$$

which gives

$$\sigma_j^{(m)} = \frac{\partial}{\partial t_m} p_j(-\tilde{\partial}) \log \tau. \tag{4.11}$$

By using a property of the polynomial $p_j(t)$, we can show that

$$\frac{\partial}{\partial t_\ell \, \partial t_m} \log \tau = -\ell \sigma_\ell^{(m)} - \sum_{j=1}^{\ell-1} \frac{\partial}{\partial t_j} \sigma_{\ell-j}^{(m)}. \tag{4.12}$$

We now derive the conserved densities of the KP Eq. (2.7) from $\sigma_j^{(1)}$ obtained in the above. Since Eq. (2.7) is written only in terms of u_2, we have to eliminate $u_3, u_4, \dots$ from $\sigma_j^{(1)}$. For the purpose, we employ Eq. (2.5) with $n=2$. Equating the coefficients of $\partial^{-j}, j = 1,2,3,\dots$, we obtain

$$u_{2,y} = 2u_{3,x} + u_{2,xx}, \tag{4.13a}$$

$$u_{3,y} = 2u_{4,x} + u_{3,xx} + 2u_2 u_{2,x}, \tag{4.13b}$$

$$u_{4,y} = 2u_{5,x} + u_{4,xx} - 2u_2 u_{2,xx} + 4u_3 u_{2,x}, \tag{4.13c}$$

$$u_{5,y} = 2u_{6,x} + u_{5,xx} + 2u_2 u_{2,xxx} - 6u_3 u_{2,xx} + 6u_4 u_{2,x}, \tag{4.13d}$$

where we have changed the variable t_2 into y for convenience. From Eqs. (4.13), we have

$$u_3 = \tfrac{1}{2} \partial_x^{-1} u_{2,y} - \tfrac{1}{2} u_{2,x}, \tag{4.14a}$$

$$u_4 = -\tfrac{1}{2} u_2^2 + \tfrac{1}{4} u_{2,xx} + \tfrac{1}{4} \partial_x^{-2} u_{2,yy} - \tfrac{1}{2} u_{2,y}, \tag{4.14b}$$

$$\begin{aligned} u_5 = {} & -u_2 \partial_x^{-1} u_{2,y} + \tfrac{3}{2} u_2 u_{2,x} - \tfrac{1}{8} u_{2,xxx} + \tfrac{3}{8} u_{2,xy} \\ & + \tfrac{1}{8} \partial_x^{-3} u_{2,yyy} - \tfrac{3}{8} \partial_x^{-1} u_{2,yy} + \tfrac{1}{4} \partial_x^{-1} (u_2^2)_y, \end{aligned} \tag{4.14c}$$

$$\begin{aligned} u_6 = {} & \tfrac{1}{2} u_2^3 + \tfrac{3}{8} (\partial_x u_2)^2 - \tfrac{3}{4} u_2 \, \partial_x^{-2} u_{2,yy} - \tfrac{1}{2} (\partial_x^{-1} u_{2,y})^2 \\ & + \tfrac{3}{4} \partial_x^{-1} (u_2 \, \partial_x^{-1} u_{2,yy}) + \{ 2u_2 \, \partial_x^{-1} u_{2,y} - \tfrac{7}{8} (u_2^2)_x \\ & + \tfrac{1}{16} u_{2,xxx} - \tfrac{3}{16} u_{2,xy} \}_x + \{ \tfrac{1}{16} \partial_x^{-4} u_{2,yyy} - \tfrac{1}{8} \partial_x^{-2} u_{2,yy} \end{aligned}$$

$$-\tfrac{1}{2}u_{2,y}+\tfrac{1}{8}\partial_x^{-2}(u_2^2)_y-\tfrac{1}{8}\partial_x^{-2}u_{2,yy}+\tfrac{5}{8}u_2^2+\tfrac{1}{8}u_{2,yy}\}_y, \tag{4.14d}$$

$\dots$,

where ∂_x^{-1} denotes the integration with respect to x. Substitution of Eqs. (4.14) into $\sigma_j^{(1)}$ gives

$$\sigma_1^{(1)}=-u_2, \tag{4.15a}$$

$$\sigma_2^{(1)}=-\tfrac{1}{2}\partial_x^{-1}u_{2,y}+\tfrac{1}{2}u_{2,x}, \tag{4.15b}$$

$$\sigma_3^{(1)}=-\tfrac{1}{2}u_2^2-\tfrac{1}{2}u_{2,xx}+(\tfrac{1}{2}u_2-\tfrac{1}{4}\partial_x^{-2}u_{2,y})_y, \tag{4.15c}$$

$$\begin{aligned}\sigma_4^{(1)}=&-\tfrac{1}{2}u_2\partial_x^{-1}u_{2,y}-2\partial_x^{-1}(\tfrac{1}{4}u_2^2)_y+\tfrac{3}{4}(u_2^2)_x\\&+\tfrac{1}{8}\partial_x^{-3}u_{2,yyy}-\tfrac{1}{4}\partial_x^{-1}u_{2,yy}+\tfrac{3}{8}u_{2,xy}-\tfrac{1}{4}\partial_x^{-1}(u_2^2)_y\\&-\tfrac{1}{8}\partial_x^{-1}u_{2,yy}-\tfrac{1}{8}u_{2,xxx}+\tfrac{1}{4}(u_2^2)_x,\end{aligned} \tag{4.15d}$$

$\dots$,

which are the conserved densities of the KP equation.

We next derive the conserved densities of the equations obtained from the KP hierarchy through the reduction procedure. The first example is the KdV Eq. (2.27). Since Eq. (2.27) is included in the two-reduction of the KP hierarchy, it holds that $L^2=B_2$. This condition demands that all the coefficients of $\partial^{-1}, j>0$ in L^2 should be zero. Hence we have

$$u_3=-\tfrac{1}{2}u_{2,x}, \tag{4.16a}$$

$$u_4=\tfrac{1}{4}u_{2,xx}-\tfrac{1}{2}u_2^2, \tag{4.16b}$$

$$u_5=-\tfrac{1}{8}u_{2,xxx}+\tfrac{3}{2}u_2u_{2,x}, \tag{4.16c}$$

$$u_6=\tfrac{1}{16}u_{2,xxxx}-\tfrac{7}{4}u_2u_{2,xx}-\tfrac{11}{8}u_{2,x}^2+\tfrac{1}{2}u_2^3, \tag{4.16d}$$

$\dots$.

Substituting Eqs. (4.16) into $\sigma_j^{(1)}$, we obtain

$$\sigma_1^{(1)}=-u_2, \tag{4.17a}$$

$$\sigma_2^{(1)}=-\tfrac{1}{2}u_{2,x}, \tag{4.17b}$$

$$\sigma_3^{(1)}=-(u_2^2/2+u_{2,xx}/4), \tag{4.17c}$$

$$\sigma_4^{(1)}=(u_{2,xx}/8+u_2^2/2)_x, \tag{4.17d}$$

$$\sigma_5^{(1)}=\frac{u_2^3}{2}-\frac{u_{2,x}^2}{8}+\left(\frac{3u_2^2}{8}+\frac{u_{2,xx}}{16}\right)_{xx}, \tag{4.17e}$$

$\dots$,

which are conserved densities of the KdV equation. We note that trivial conserved densities appear at two intervals. This fact can be explained by Eq. (4.12). In the case of ℓ reduction, τ does not depend on $\ell, 2\ell,\dots,n\ell,\dots$. Therefore we have

$$\sigma_{n\ell}^{(1)}=\frac{1}{n\ell}\sum_{j=1}^{n\ell-1}\frac{\partial\sigma_{n\ell-j}^{(1)}}{\partial t_j},\quad n=1,2,\dots. \tag{4.18}$$

By means of Eq. (4.7), the above equation reduces to

$$\sigma_{n\ell}^{(1)}=\frac{1}{n\ell}\frac{\partial}{\partial x}\left(\sum_{j=1}^{n\ell-1}\sigma_{n\ell-j}^{(j)}\right), \tag{4.19}$$

which shows that the trivial conserved densities of the equations in the ℓ reduction of the KP hierarchy appear at ℓ intervals.

The second example is the Boussinesq-like Eq. (2.28). For this equation, we have the condition $L^3=B_3$, which yields

$$u_4=-u_{3,x}-\tfrac{1}{3}u_{2,xx}-u_2^2, \tag{4.20a}$$

$$u_5=\tfrac{2}{3}u_{3,xx}+\tfrac{1}{3}u_{2,xxx}+2u_2u_{2,x}-2u_2u_3, \tag{4.20b}$$

$$\begin{aligned}u_6=&-\tfrac{1}{3}u_{3,xxx}+4u_{3,x}u_2-\tfrac{2}{9}u_{2,xxxx}-u_2u_{2,xx}\\&-u_{2,x}^2+3u_{2,x}u_3-u_3^2+\tfrac{5}{3}u_2^3,\end{aligned} \tag{4.20c}$$

$\dots$.

Substituting Eq. (4.20) into $\sigma_j^{(1)}$, noticing $u_3=\tfrac{1}{2}\partial_x^{-1}u_{2,t_2}-\tfrac{1}{2}u_{2,x}$ and introducing $v=\partial_x^{-1}u_{2,t_2}$, we obtain the conserved densities of Eq. (2.28);

$$\sigma_1^{(1)}=-u_2, \tag{4.21a}$$

$$\sigma_2^{(1)}=\tfrac{1}{2}u_{2,x}-\tfrac{1}{2}v, \tag{4.21b}$$

$$\sigma_3^{(1)}=\tfrac{1}{6}(-u_{2,x}+3v)_x, \tag{4.21c}$$

$$\sigma_4^{(1)}=-\tfrac{1}{2}u_2v+(-\tfrac{1}{3}v_x-\tfrac{1}{4}u_2^2)_x, \tag{4.21d}$$

$$\begin{aligned}\sigma_5^{(1)}=&\tfrac{1}{2}u_2^3-\tfrac{1}{4}v^2-\tfrac{1}{12}u_{2,x}^2+(-\tfrac{1}{18}u_{2,xxx}\\&+\tfrac{5}{6}u_2u_{2,x}+\tfrac{1}{2}u_2v+\tfrac{1}{6}v_{xx})_x,\end{aligned} \tag{4.21e}$$

$\dots$.

The third example is the coupled KdV Eq. (2.30), which belongs to the four-reduction of the KP hierarchy. From the condition $L^4=B_4$, we have

$$\sigma_1^{(1)}=-u_2, \tag{4.22a}$$

$$\sigma_2^{(1)}=-u_3, \tag{4.22b}$$

$$\sigma_3^{(1)}=-u_4-u_2^2, \tag{4.22c}$$

$$\sigma_4^{(1)}=\tfrac{1}{4}(6u_4+4u_{3,x}+u_{2,xx}+5u_2^2)_x, \tag{4.22d}$$

$$\begin{aligned}\sigma_5^{(1)}=&-u_2^3-\tfrac{1}{2}u_3^2-u_4u_2+\tfrac{1}{4}u_{2,x}^2-\tfrac{1}{2}u_2u_{3,x}\\&+\tfrac{1}{8}(-10u_{4,x}-10u_{3,xx}-12u_2u_3\\&-3u_{2,xxx}-22u_2u_{2,x})_x,\end{aligned} \tag{4.22e}$$

$\dots$.

Substituting Eqs. (4.22) into $\sigma_j^{(1)}$ and transforming the variables as Eqs. (2.31), we obtain the conserved densities of Eq. (2.30);

$$\sigma_1^{(1)}=-u, \tag{4.23a}$$

$$\sigma_2^{(1)}=\tfrac{1}{2}(u-2\phi)_x, \tag{4.23b}$$

$$\sigma_3^{(1)}=-\tfrac{1}{2}(u^2-2\phi^2+2\omega)+\tfrac{1}{4}(-u_x+4\phi_x)_x, \tag{4.23c}$$

$$\sigma_4^{(1)}=\tfrac{1}{8}(u_{xx}+4u^2-4\phi_{xx}-12\phi^2+12\omega)_x, \tag{4.23d}$$

$$\begin{aligned}\sigma_5^{(1)}=&-\tfrac{1}{2}(u^3-\tfrac{1}{4}u_x^2-2u\phi^2+\phi_x^2+2u\omega)\\&+\tfrac{1}{16}(-u_{xxx}-12uu_x-16u\phi_x\\&+40\phi_x\phi-20\omega_x)_x,\end{aligned} \tag{4.23e}$$

$\dots$.

Finally, we discuss the conserved densities of the BKP hierarchy. From the constraint that the constant terms of B_n for odd n vanish, we have the relations

$$u_3=-u_{2,x},$$

$$u_5=-2u_{4,x}+u_{2,xxx},$$

$$u_7=-3u_{6,x}+5u_{4,xxx}-3u_{2,xxxxx},$$

$\dots$.

Substituting the above into $\sigma_j^{(1)}$, we obtain the conserved densities of the BKP hierarchy,

$$\sigma_1^{(1)} = -u_2, \quad (4.24a)$$

$$\sigma_2^{(1)} = u_{2,x}, \quad (4.24b)$$

$$\sigma_3^{(1)} = -u_4 - u_2^2, \quad (4.24c)$$

$$\sigma_4^{(1)} = (2u_2^2 + 2u_4 - u_{2,xx})_x, \quad (4.24d)$$

$$\sigma_5^{(1)} = -u_6 - 4u_4u_2 - 2u_2^3 - 2u_2u_{2,xx} + 5u_{2,x}^2, \quad (4.24e)$$

$$\sigma_6^{(1)} = (11u_4u_2 + \tfrac{16}{3}u^3 + 3u_6 - 5u_{4,xx} + 3u_{2,xxxx} + 5u_{2,x}^2 - 3u_2u_{2,xx})_x, \quad (4.24f)$$

$$\cdots.$$

Let us derive the conserved densities of the Sawada–Kotera Eq. (2.32). Since the equation is included in the three-reduction of the BKP hierarchy, we have the conditon $L^3 = B_3$, which yields

$$u_4 = -u_2^2 + \tfrac{2}{3}u_{2,xx}, \quad (4.25a)$$

$$u_6 = \tfrac{5}{3}u_2^3 - 5u_{2,x}^2 - 5u_2u_{2,xx} + \tfrac{1}{9}u_{2,xxxx}, \quad (4.25b)$$

$$\cdots.$$

Substituting Eqs. (4.25) into Eqs. (4.24), we obtain the conserved densities of Eq. (2.32);

$$\sigma_1^{(1)} = -u_2, \quad (4.26a)$$

$$\sigma_2^{(1)} = u_{2,x}, \quad (4.26b)$$

$$\sigma_3^{(1)} = -\tfrac{2}{3}u_{2,xx}, \quad (4.26c)$$

$$\sigma_4^{(1)} = \tfrac{1}{3}u_{2,xxx}, \quad (4.26d)$$

$$\sigma_5^{(1)} = \tfrac{1}{3}(u_2^3 - u_{2,x}^2) + \tfrac{1}{18}(3u^2 - 2u_{2,xx})_{xx}, \quad (4.26e)$$

$$\sigma_6^{(1)} = -\tfrac{2}{3}(u_2^3 + u_2u_{2,xx})_x, \quad (4.26f)$$

$$\cdots.$$

We note that the $\sigma_j^{(1)}$'s for $j = 2,4,6,\ldots$ and $j = 3,6,9,\ldots$ are the trivial conserved densities. The former is due to the property of the BKP hierarchy and the latter to that of the three-reduction.

V. SYMMETRIES OF THE KP EQUATION

In this section, we consider symmetries of the KP equation. As mentioned in Sec. II, the KP Eq. (2.7) is obtained from the compatibility condition of

$$\frac{\partial \psi}{\partial t_2} = B_2\psi = \frac{\partial^2 \psi}{\partial x^2} + 2u\psi \quad (5.1)$$

and

$$\frac{\partial \psi}{\partial t_3} = B_3\psi = \frac{\partial^3 \psi}{\partial x^3} + 3u\frac{\partial \psi}{\partial x} + \frac{3}{2}\frac{\partial u}{\partial x}\psi + \frac{3}{2}\left(\partial_x^{-1}\frac{\partial u}{\partial t_2}\right)\psi, \quad (5.2)$$

where we have used Eq. (4.13a) to eliminate u_3 and rewritten u_2 as u for simplicity. The adjoint linear problem is given by

$$\frac{\partial \psi^*}{\partial t_2} = -B_2^*\psi^* = -\frac{\partial^2 \psi^*}{\partial x^2} - 2u\psi^* \quad (5.3)$$

and

$$\frac{\partial \psi^*}{\partial t_3} = -B_3^*\psi^* = \frac{\partial^3 \psi^*}{\partial x^3} + 3u\frac{\partial \psi^*}{\partial x} + \frac{3}{2}\frac{\partial u}{\partial x}\psi^* - \frac{3}{2}\left(\partial_x^{-1}\frac{\partial u}{\partial t_2}\right)\psi^*. \quad (5.4)$$

If we integrate Eq. (2.7) with respect to x, we obtain

$$\frac{\partial u}{\partial t_3} - \frac{1}{4}\frac{\partial^3 u}{\partial x^3} - 3u\frac{\partial u}{\partial x} - \frac{3}{4}\partial_x^{-1}\frac{\partial^2 u}{\partial t_2^2} = 0. \quad (5.5)$$

Hence the linearized KP equation may be written by

$$\frac{\partial S}{\partial t_3} - \frac{1}{4}\frac{\partial^3 S}{\partial x^3} - 3\frac{\partial}{\partial x}(uS) - \frac{3}{4}\partial_x^{-1}\frac{\partial^2 S}{\partial t_2^2} = 0. \quad (5.6)$$

Using Eqs. (5.1)–(5.4), we find that $\psi\psi^*$ satisfies

$$\frac{\partial s}{\partial t_3} - \frac{1}{4}\frac{\partial^3 s}{\partial x^3} - 3u\frac{\partial s}{\partial x} - \frac{3}{4}\partial_x^{-1}\frac{\partial^2 s}{\partial t_2^2} = 0, \quad (5.7)$$

which means that $(\partial/\partial x)(\psi\psi^*)$ is a solution of Eq. (5.6). By the definition in Sec. III, $(\partial/\partial x)(\psi\psi^*)$ gives a symmetry of the KP equation. We show that it also generates an infinite number of symmetries. From Eqs. (2.8) and (2.20), we obtain

$$\psi\psi^* = \sum_{n=0}^{\infty} s_n\lambda^{-n}, \quad (5.8)$$

where

$$s_n = \sum_{j=0}^{n} w_j w_{n-j}^*, \quad (5.9)$$

and where w_0 is defined to be 1. Therefore,

$$S_n = \frac{\partial}{\partial x}s_n, \quad n = 1,2,3,\ldots \quad (5.10)$$

give a series of symmetries. By expressing w_j and w_j^* in terms of u, we find

$$S_0 = \frac{\partial}{\partial x}(1), \quad (5.11a)$$

$$S_1 = \frac{\partial}{\partial x}(0), \quad (5.11b)$$

$$S_2 = \frac{\partial}{\partial x}(u), \quad (5.11c)$$

$$S_3 = \frac{\partial}{\partial t_2}(u), \quad (5.11d)$$

$$S_4 = \frac{1}{4}\frac{\partial^3 u}{\partial x^3} + 3u\frac{\partial u}{\partial x} + \frac{3}{4}\partial_x^{-1}\left(\frac{\partial^2 u}{\partial t_2^2}\right), \quad (5.11e)$$

$$S_5 = \frac{1}{2}\frac{\partial^3 u}{\partial x^2\partial t_2^2} + 12u\frac{\partial u}{\partial t_2} + 2\frac{\partial u}{\partial x}\partial_x^{-1}\left(\frac{\partial u}{\partial t_2}\right) + \frac{1}{2}\partial_x^{-2}\left(\frac{\partial^3 u}{\partial t_2^3}\right), \quad (5.11f)$$

$$\cdots.$$

We now express S_n by the τ function. From Eqs. (2.14) and (2.22), we have

$$\psi\psi^* = \frac{\tau(t_1 - 1/\lambda, t_2 - 1/2\lambda^2,...)\tau(t_1 + 1/\lambda, t_2 + 1/2\lambda^2,...)}{\tau(t_1,t_2,...)^2},$$

$$= \frac{1}{\tau(t)^2}\exp\left(\sum_{n=1}^{\infty}\frac{\partial_{y_n}}{n\lambda^n}\right)\tau(t+y)\tau(t-y)\Big|_{y=0},$$

$$= \frac{1}{\tau^2}\sum_{n=0}^{\infty}\frac{p_n(\widetilde{D})\tau\cdot\tau}{\lambda^n}. \qquad (5.12)$$

Comparing Eq. (5.12) with Eq. (5.8), we obtain

$$s_n = (1/\tau^2)p_n(\widetilde{D})\tau\cdot\tau. \qquad (5.13)$$

If we expand Eq. (2.16) in powers of y_j, we have

$$\sum_{n=0}^{\infty} p_n(-2y)p_{n+1}(\widetilde{D})\Big\{1 + \sum_{i=0}^{\infty} y_i D_i + \frac{1}{2}\Big(\sum_{i=0}^{\infty} y_i D_i\Big)^2 + \cdots\Big\}\tau\cdot\tau = 0. \qquad (5.14)$$

The coefficient of the linear term in y_n gives

$$\tfrac{1}{2}D_1 D_n \tau\cdot\tau = p_{n+1}(\widetilde{D})\tau\cdot\tau. \qquad (5.15)$$

Therefore, s_n may be written by

$$s_n = (2\tau^2)^{-1}D_1 D_{n-1}\tau\cdot\tau, \qquad (5.16)$$

which reduces to

$$s_n = \partial_x^{-1}\frac{\partial u}{\partial t_{n-1}} \qquad (5.17)$$

by means of Eq. (2.15a). The bilinear Eq. (5.15) is then written by

$$\frac{\partial u}{\partial t_n} = \frac{\partial}{\partial x}s_{n+1} = S_{n+1}, \qquad (5.18)$$

which forms a subset of Eq. (2.16) and is considered as a higher-order KP equation for $n \geqslant 4$. The symmetries S_m and S_n satisfy

$$S'_n[S_m] = \frac{\partial}{\partial t_{n-1}}S_m,$$
$$= \frac{\partial^2 u}{\partial t_{n-1}\,\partial t_{m-1}},$$
$$= \frac{\partial}{\partial t_{m-1}}S_n,$$
$$= S'_m[S_n],$$

where the prime is the Fréchet derivative introduced in Sec. III. Therefore, the higher-order KP equations constitute a hierarchy of commutative equations.

VI. RECURSION OPERATOR

As mentioned in Sec. III, Fokas and Santini have presented a recursion operator mapping symmetries of the KP equation into symmetries. In this section, we discuss the relationship between the recursion operator and the results obtained in Sec. V. Before considering the KP equation, we briefly study the KdV equation in order to demonstrate the basic idea.

Since the KdV equation belongs to the two-reduction of the KP hierarchy, Eqs. (5.2)–(5.5) hold for the KdV case by taking $\partial u/\partial t_2 = 0$, $\partial\psi/\partial t_2 = \lambda^2\psi$, and $\partial\psi^*/\partial t_2 = -\lambda^2\psi^*$. Then it is straightforward to show that $\psi\psi^*$ satisfies

$$\left(\frac{1}{4}\frac{\partial^3}{\partial x^3} + 2u\frac{\partial}{\partial x} + u_x\right)s = \lambda^2\frac{\partial}{\partial x}s \qquad (6.1)$$

or

$$R^* s = \lambda^2 s, \qquad (6.2)$$

where

$$R^* = \frac{1}{4}\frac{\partial^2}{\partial x^2} + 2u - \partial_x^{-1}u_x \qquad (6.3)$$

is known as the squared eigenfunction operator. This operator maps conserved covariants of the KdV equation into conserved covariants. The adjoint operator,

$$R = \frac{1}{4}\frac{\partial^2}{\partial x^2} + 2u + u_x\partial_x^{-1}, \qquad (6.4)$$

is the recursion operator which maps symmetries into symmetries. In fact, substituting Eq. (5.8) into Eq. (6.2) and using Eq. (5.10), we can show that

$$RS_n = S_{n+2}, \qquad (6.5)$$

where the S_n's are the symmetries of the KdV equation obtained by applying the two-reduction on Eqs. (5.11).

The two-dimensional version of the recursion operator presented by Fokas and Santini[5] is written as follows:

$$\Phi^{(12)} = \frac{\partial^2}{\partial x^2} + q^+ + \frac{\partial}{\partial x}q^+\partial_x^{-1} + q^-\partial_x^{-1}q^-\partial_x^{-1}, \qquad (6.6)$$

where

$$q^\pm = q^{(1)} \pm q^{(2)} + \alpha\left(\frac{\partial}{\partial t_2^{(1)}} \mp \frac{\partial}{\partial t_2^{(2)}}\right), \qquad (6.7)$$

and $q^{(1)}(x,t_2^{(1)})$ and $q^{(2)}(x,t_2^{(2)})$ are the solutions of the KP equation possessing the prescribed arguments, respectively. The extended symmetries are given by

$$S_n^{(12)} = (\Phi^{(12)})^n S_0^{(12)}, \quad n = 0,1,2,..., \qquad (6.8)$$

where

$$S_0^{(12)} = \left(\Phi^{(12)}\frac{\partial}{\partial x}\right)1$$
$$= \frac{\partial q^{(1)}}{\partial x} + \frac{\partial q^{(2)}}{\partial x} + (q^{(1)} - q^{(2)})\partial_x^{-1}(q^{(1)} - q^{(2)})$$
$$+ \alpha\,\partial_x^{-1}\left(\frac{\partial}{\partial t_2^{(1)}}q^{(1)} - \frac{\partial}{\partial t_2^{(2)}}q^{(2)}\right). \qquad (6.9)$$

The two-dimensional version of the squared eigenfunction operator is the adjoint $\Phi^{(12)*}$ of (6.6). The associated linear problem is written by

$$\alpha \frac{\partial}{\partial t_2^{(j)}} \psi^{(j)} = \frac{\partial^2}{\partial x^2} \psi^{(j)} + q^{(j)} \psi^{(j)}, \quad j = 1,2, \tag{6.10}$$

where $\psi^{(1)}$ and $\psi^{(2)}$ have the arguments $(x,t_2^{(1)})$ and $(x,t_2^{(2)})$, respectively, and α is an arbitrary parameter. Fokas and Santini have shown that

$$\Phi^{(12)*} \psi^{(1)} \psi^{(2)*} = 0, \tag{6.11}$$

and $\Phi^{(12)}$ gives the recursion operator for the KP equation, if the limit of $t_2^{(2)} \to t_2^{(1)}$ is taken.

We now compare these results with those obtained in Sec. V. If $\alpha = 1$ and $q^{(j)} = 2u^{(j)}$ for $j = 1,2$ are taken, Eq. (6.7) reduces to Eq. (5.1). Then the eigenfunction in Sec. V satisfies

$$\Phi^{(12)*} \psi^{(1)} \psi^{(2)*} = 0, \tag{6.12}$$

where $\psi^{(1)}$ and $\psi^{(2)}$ correspond to the potentials $q^{(j)} = 2u^{(j)}$ for $j = 1$ and 2, respectively. From Eqs. (2.8) and (2.20), we obtain

$$\psi^{(1)} \psi^{(2)*} = s(t^{(1)},t^{(2)}) \exp\{(t_2^{(1)} - t_2^{(2)})\lambda^2\}, \tag{6.13}$$

$$s(t^{(1)},t^{(2)}) = \sum_{n=0}^{\infty} s_n(t^{(1)},t^{(2)})\lambda^{-n}, \tag{6.14}$$

where the two sets of infinitely many time variables,

$$t^{(j)} = (t_1, t_2^{(j)}, t_3, \ldots), \quad j = 1,2,$$

are introduced and $s_n(t^{(1)}, t^{(2)})$ are given by

$$s_n(t^{(1)},t^{(2)}) = \sum_{j=0}^{\infty} w_j(t^{(1)}) w_{n-j}^*(t^{(2)}). \tag{6.15}$$

Equation (6.13) gives

$$\left(\frac{\partial}{\partial t_2^{(1)}} - \frac{\partial}{\partial t_2^{(2)}}\right) \psi^{(1)} \psi^{(2)*} = \left\{\left(\frac{\partial}{\partial t_2^{(1)}} - \frac{\partial}{\partial t_2^{(2)}}\right) s + 2\lambda^2\right\} \exp\{(t_2^{(1)} - t_2^{2})\lambda^2\} \tag{6.16a}$$

and

$$\left(\frac{\partial}{\partial t_2^{(1)}} + \frac{\partial}{\partial t_2^{(2)}}\right)^n \psi^{(1)} \psi^{(2)*} = \left\{\left(\frac{\partial}{\partial t_2^{(1)}} + \frac{\partial}{\partial t_2^{(2)}}\right)^n s\right\} \exp\{(t_2^{(1)} - t_2^{(2)})\lambda^2\}. \tag{6.16b}$$

Substituting Eqs. (6.16) into Eq. (6.12), we obtain

$$\Phi^{(12)*} s(t^{(1)},t^{(2)}) = 4\lambda^2 s(t^{(1)},t^{(2)}). \tag{6.17}$$

Consequently, we have from Eq. (6.14) that

$$\tfrac{1}{4}\Phi^{(12)*} s_n(t^{(1)},t^{(2)}) = s_{n+2}(t^{(1)},t^{(2)}), \tag{6.18}$$

which is the two-dimensional version of Eq. (6.5). Thus Eq. (6.18) is the link between the Sato theory and the two-dimensional recursion operator of Fokas and Santini. The symmetries obtained by both schemes are given by Eqs. (5.11).

Fokas and Santini also presented a theorem that

$$S_n^{(12)} = 0 \tag{6.19}$$

gives an auto-Bäcklund transformation for the KP equation and its higher-order equations. We here consider this result from the view point of the Sato theory. For $n = 0$, Eq. (6.19) is written by

$$\frac{\partial q^{(1)}}{\partial x} + \frac{\partial q^{(2)}}{\partial x} + (q^{(1)} - q^{(2)})\partial_x^{-1}(q^{(1)} - q^{(2)}) + \alpha\, \partial_x^{-1}\left(\frac{\partial}{\partial t_2^{(1)}} q^{(1)} - \frac{\partial}{\partial t_2^{(2)}} q^{(2)}\right) = 0. \tag{6.20}$$

By chosing $\alpha = 1$, changing the dependent variables as $q^{(1)} = 2(\partial^2/\partial x^2)\log \tau^{(1)}(x,t_2^{(1)})$ and $q^{(2)} = 2(\partial^2/\partial x^2) \times \log \tau^{(2)}(x,t_2^{(2)})$, and taking $t_2^{(1)} = t_2^{(2)} (= t_2)$, Eq. (6.20) is reduced to

$$(D_x^2 + D_{t_2})\tau^{(1)} \cdot \tau^{(2)} = 0, \tag{6.21}$$

which is nothing but the lowest order of the equations in the first modified KP hierarchy, or in other words, an auto-Bäcklund transformation for the KP equation.

From Eqs. (6.18), we see that $(\partial/\partial x) s_n(t^{(1)},t^{(2)})$ corresponds to the extended symmetry $S_n^{(12)}$. By following the same procedure to get Eq. (5.13), we find that

$$S_n^{(12)} = [\underline{P}_n(\widetilde{D})\tau^{(1)} \cdot \tau^{(2)}]/\tau^{(1)}\tau^{(2)}, \tag{6.22}$$

where we again take $t_2^{(1)} = t_2^{(2)} (= t_2)$. Hence, Eq. (6.19) gives

$$\underline{P}_n(\widetilde{D})\tau^{(1)} \cdot \tau^{(2)} = 0. \tag{6.23}$$

It has been shown by Date *et al.*[10] that the mth modified KP hierarchy is given by

$$\sum_{n=0}^{\infty} \underline{P}_n(-2y)\underline{P}_{n+m+1}(\widetilde{D}) \exp\left(\sum_{i=0}^{\infty} y_i D_i\right) \tau \cdot \tau' = 0. \tag{6.24}$$

The terms which do not include y in Eq. (6.24) yields

$$\underline{P}_{m+1}(\widetilde{D})\tau \cdot \tau' = 0. \tag{6.25}$$

Therefore, Eq. (6.23) is considered to be a part of the modified KP hierarchy.

APPENDIX

We here give a list of $\sigma_j^{(m)}$ for $1<m<5$ and $1<j<8-m$:

$$\sigma_1^{(1)} = -u_2,$$

$$\sigma_2^{(1)} = u_3,$$

$$\sigma_3^{(1)} = -u_4 - u_2^2,$$

$$\sigma_4^{(1)} = -u_5 - 3u_3u_2 + u_2u_{2,x},$$

$$\sigma_5^{(1)} = -u_6 - 4u_4u_2 - 2u_3^2 - 2u_2^3 + u_{3,x}u_2 - u_{2,xx}u_{2,x} + 3u_3u_{2,x},$$

$$\sigma_6^{(1)} = -u_7 - 5u_5u_2 - 5u_4u_3 - 10u_3u_2^2 + 6u_2^2u_{2,x} + 6u_4u_{2,x} - 4u_3u_{2,xx} + u_2u_{2,xxx} + 3u_3u_{3,x} - u_2u_{3,xx} + u_{4,x}u_2,$$

$$\sigma_7^{(1)} = -u_8 - 6u_6u_2 - 6u_5u_3 - 3u_4^2 - 15u_4u_2^2 - 15u_3^2u_2 - 5u_2^4 + 29u_3u_2u_{2,x} + 10u_5u_{2,x} - 6u_2u_{2,x}^2 - 8u_2^2u_{2,xx}$$
$$- 10u_4u_{2,xx} + 5u_3u_{2,xxx} - u_2u_{2,xxxx} + 7u_3u_2^2 + 6u_4u_{3,x} - 4u_3u_{3,xx} + u_2u_{3,xxx} + 3u_3u_{4,x} - u_2u_{4,xx} + u_2u_{5,x},$$
$$\sigma_1^{(2)} = u_{2,x} + 2u_3,$$
$$\sigma_2^{(2)} = u_2^2 + u_{3,x} + 2u_4,$$
$$\sigma_3^{(2)} = 4u_2u_3 + u_{4,x} + 2u_5,$$
$$\sigma_4^{(2)} = u_{2,xx}u_2 - u_{2,x}^2 - 3u_{2,x}u_3 + 2u_2^3 + u_2u_{3,x} + 6u_2u_4 + 3u_3^2 + u_{5,x} + 2u_6,$$
$$\sigma_5^{(2)} = -u_{2,xxx}u_2 + u_{2,xx}u_{2,x} + 5u_{2,xx}^2u_3 - 4u_{2,x}u_2^2 - 4u_{2,x}u_{3,x} - 8u_{2,x}u_4 + 12u_2^2u_3 + u_2u_{3,xx} + 2u_2u_{4,x}$$
$$+ 8u_2u_5 - 2u_{3,x}u_3 + 8u_3u_4 + u_{6,x} + 2u_7,$$
$$\sigma_6^{(2)} = u_{2,xxxx}u_2 - u_{2,xxx}u_{2,x} - 6u_{2,xxx}u_3 + 8u_{2,xx}u_2^2 + 4u_{2,xx}u_{3,x} + 14u_{2,xx}u_4 - 30u_{2,x}u_2u_3 + u_{2,x}u_{3,xx}$$
$$- 7u_{2,x}u_{4,x} - 15u_{2,x}u_5 + 5u_2^4 - 2u_2^2u_{3,x} + 20u_2^2u_4 - u_2u_{3,xxx} + 20u_2u_3^2 + u_2u_{4,xx} + 3u_2u_{5,x} + 10u_2u_6$$
$$+ 5u_{3,xx}u_3 - 3u_{3,x}^2 - 7u_{3,x}u_4 - u_3u_{4,x} + 10u_3u_5 + 5u_4^2 + u_{7,x} + 2u_8,$$
$$\sigma_1^{(3)} = u_{2,xx} + 3u_2^2 + 3u_{3,x} + 3u_4,$$
$$\sigma_2^{(3)} = 6u_2u_3 + u_{3,xx} + 3u_{4,x} + 3u_5,$$
$$\sigma_3^{(3)} = 2u_{2,xx}u_2 - u_{2,x}^2 - 3u_{2,x}u_3 + 4u_2^3 + 3u_2u_{3,x} + 9u_2u_4 + 3u_3^2 + u_{4,xx} + 3u_{5,x} + 3u_6,$$
$$\sigma_4^{(3)} = -u_{2,xxx}u_2 + 6u_{2,xx}u_3 - 6u_{2,x}u_2^2 - 6u_{2,x}u_{3,x} - 9u_{2,x}u_4 + 18u_2^2u_3 + 3u_2u_{3,xx} + 6u_2u_{4,x} + 12u_2u_5 + 9u_3u_4$$
$$+ u_{5,xx} + 3u_{6,x} + 3u_7,$$
$$\sigma_5^{(3)} = u_{2,xxxx}u_2 - u_{2,xxx}u_{2,x} - 6u_{2,xxx}u_3 + u_{2,xx}^2 + 12u_{2,xx}u_2^2 + 5u_{2,xx}u_{3,x} + 16u_{2,xx}u_4 - 36u_{2,x}u_2u_3 - 2u_{2,x}u_{3,xx}$$
$$- 13u_{2,x}u_{4,x} - 18u_{2,x}u_5 + 9u_2^4 + 30u_2^2u_4 - u_2u_{3,xxx} + 24u_2u_3^2 + 4u_2u_{4,xx} + 9u_2u_{5,x} + 15u_2u_6 + 7u_{3,xx}u_3 - 5u_{3,x}^2$$
$$- 6u_{3,x}u_4 + 3u_3u_{4,x} + 12u_3u_5 + 6u_4^2 + u_{6,xx} + 3u_{7,x} + 3u_8,$$
$$\sigma_1^{(4)} = u_{2,xxx} + 6u_{2,x}u_2 + 12u_2u_3 + 4u_{3,xx} + 6u_{4,x} + 4u_5,$$
$$\sigma_2^{(4)} = 2u_{2,xx}u_2 - u_{2,x}^2 + 4u_2^3 + 6u_2u_{3,x} + 12u_2u_4 + u_{3,xxx} + 6u_3^2 + 4u_{4,xx} + 6u_{5,x} + 4u_6,$$
$$\sigma_3^{(4)} = 6u_{2,xx}u_3 - 6u_{2,x}u_{3,x} - 6u_{2,x}u_4 + 24u_2^2u_3 + 6u_2u_{3,xx} + 12u_2u_{4,x} + 16u_2u_5 + 12u_3u_4 + u_{4,xxx}$$
$$+ 4u_{5,xx} + 6u_{6,x} + 4u_7,$$
$$\sigma_4^{(4)} = u_{2,xxxx}u_2 - 2u_{2,xxx}u_{2,x} - 5u_{2,xxx}u_3 + u_{2,xx}^2 + 12u_{2,xx}u_2^2 + 5u_{2,xx}u_{3,x} + 16u_{2,xx}u_4 - 6u_{2,x}^2u_2 - 30u_{2,x}u_2u_3$$
$$- 5u_{2,x}u_{3,xx} - 16u_{2,x}u_{4,x} - 16u_{2,x}u_5 + 9u_2^4 + 6u_2^2u_{3,x} + 36u_2^2u_4$$
$$+ u_2u_{3,xxx} + 36u_2u_3^2 + 10u_2u_{4,xx} + 18u_2u_{5,x} + 20u_2u_6$$
$$+ 10u_{3,xx}u_3 - 5u_{3,x}^2 - 6u_{3,x}u_4 + 6u_3u_{4,x} + 16u_3u_5 + 6u_4^2 + u_{5,xxx} + 4u_{6,xx} + 6u_{7,x} + 4u_8,$$
$$\sigma_1^{(5)} = u_{2,xxxx} + 10u_{2,xx}u_2 + 5u_{2,x}^2 + 10u_{2,x} + 10u_2^3 + 20u_2u_{3,x} + 20u_2u_4 + 5u_{3,xxx} + 10u_3^2 + 10u_{4,xx}$$
$$+ 10u_{5,x} + 5u_6,$$
$$\sigma_2^{(5)} = 10u_{2,xx}u_3 + 30u_2^2u_3 + 10u_2u_{3,xx} + 20u_2u_{4,x} + 20u_2u_5 + u_{3,xxxx}$$
$$+ 10u_{3,x}u_3 + 20u_3u_4 + 5u_{4,xxx} + 10u_{5,xx} + 10u_{6,x} + 5u_7,$$
$$\sigma_3^{(5)} = 2u_{2,xxxx}u_2 - 2u_{2,xxx}u_{2,x} - 5u_{2,xxx}u_3 + u_{2,xx}^2 + 20u_{2,xx}u_2^2 + 5u_{2,xx}u_{3,x} + 20u_{2,xx}u_4 - 20u_{2,x}u_2u_3 - 5u_{2,x}u_{3,xx}$$
$$- 10u_{2,x}u_{4,x} - 10u_{2,x}u_5 + 15u_2^4 + 20u_2^2u_{3,x} + 50u_2^2u_4 + 5u_2u_{3,xxx} + 40u_2u_3^2 + 20u_2u_{4,xx} + 30u_2u_{5,x} + 25u_2u_6$$
$$+ 10u_{3,xx}u_3 - 5u_{3,x}^2 + 10u_3u_{4,x} + 20u_3u_5 + u_{4,xxxx} + 10u_4^2 + 5u_{5,xxx} + 10u_{6,xx} + 10u_{7,x} + 5u_8,$$

[1]C. S. Gardner, J. M. Greene, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[2]R. M. Miura, J. Math. Phys. **9**, 1202 (1968).
[3]A. S. Fokas, Stud. Appl. Math. **77**, 253 (1987).
[4]A. S. Fokas and P. M. Santini, Stud. Appl. Math. **75**, 179 (1986).
[5]P. M. Santini and A. S. Fokas, Commun. Math. Phys. **115**, 375 (1988).
[6]A. S. Fokas and P. M. Santini, Commun. Math. Phys. **116**, 449 (1988).
[7]M. Sato, RIMS Kokyuroku **439**, 30 (1981).
[8]M. Sato and Y. Sato, in *Nonlinear Partial Differential Equations in Applied Science*, edited by H. Fujita, P. D. Lax, and G. Strang (Kinokuniya/North-Holland, Tokyo, 1983), p. 259.
[9]Y. Ohta, J. Satsuma, D. Takahashi, and T. Tokihiro, Prog. Theor. Phys. Suppl. **94**, 210 (1988).
[10]E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, in *Nonlinear Integrable Systems—Classical Theory and Quantum Theory*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).
[11]J. Satsuma and R. Hirota, J. Phys. Soc. Jpn. **51**, 3390 (1982).

The asymptotic localization of position probability in quantum mechanics

J. E. G. Farina and S. N. Mosley
Department of Mathematics, University of Nottingham, Nottingham NG7 2RD, England

(Received 22 September 1989; accepted for publication 31 January 1990)

The method of stationary phase is discussed in a mathematically rigorous way. A resulting lemma is used to derive the asymptotic localization of position probability under the evolution operator $U_t = F^{-1}\exp(-i\omega t)F$, where ω is a continuous function of the wave vector $\mathbf{k} = (k_1,...,k_N)$ $(N = 1,2,...)$ with continuous first and second derivatives. The measurement of velocity is discussed, and the interpretation of the self-adjoint operator $\hat{v}_i = F^{-1}\nabla_i\omega F$ as representing the ith component of velocity is also discussed.

I. INTRODUCTION

The method of stationary phase has long been known to physicists as a heuristic method of determining where a wave function is concentrated. It rests on the observation that the integral

$$\int_{k_1}^{k_2} M(k)\exp[iA(k)]dk, \qquad (1.1)$$

where M and A are real functions and M is non-negative, is "small" if the rate of oscillation of $A(k)$ as k increases from k_1 to k_2 is sufficiently great.

Suppose a wave packet ψ at time zero, moving in the space $\mathbb{R}^N$ $(N = 1,2,...)$, evolves into the wave packet $U_t\psi$ at time t, where

$$U_t\psi(\mathbf{x}) = (2\pi)^{-N/2}\int F\psi(\mathbf{k})\exp[i(\mathbf{k}\cdot\mathbf{x} - \omega t)]d^N k. \qquad (1.2)$$

Here $\mathbf{x} = (x_1,...,x_N)\in\mathbb{R}^N$, $\mathbf{k} = (k_1,...,k_N)$, F is the Fourier–Plancherel operator, $\omega = \omega(\mathbf{k})$ is real, $d^N k = dk_1\cdots dk_N$, and the integral is over all $\mathbb{R}^N$. If $\omega = \hbar k^2/2m$, U_t is the evolution operator for the free motion of a nonrelativistic particle in N dimensions. The method of stationary phase suggests that when t is large, $U_t\psi$ should be concentrated in the "classically allowed" region where $\mathbf{x} - (\nabla\omega)t$ vanishes for at least one value of $\mathbf{k}$ in the support of $F\psi$.

A rigorous discussion of this is given in Appendix 1 to Sec. XI.3 of the book by Reed and Simon,[1] based on work by Hörmander (see the note on p. 348 of Ref. 1). There it is shown that if ψ is a continuous function with continuous derivatives of all orders, and $F\psi$ has compact support, $U_t\psi$ does indeed fall off rapidly outside the classically allowed region when $t\to\infty$. They also point out that the result in the case $\omega = (k^2 + m^2)^{1/2}$ is important in the case of Haag–Ruelle scattering theory, and discuss the application of the result to proofs of asymptotic completeness.

Wan and McLean have also studied this problem.[2,3] In Ref. 3 they prove that

$$\lim_{t\to\pm\infty}\|E(\mathbf{x}\in[\mathbf{v}t,\mathbf{w}t])V_t\psi\| = \lim_{t\to\pm\infty}\|E(\hbar\mathbf{k}\in[m\mathbf{v},m\mathbf{w}])V_t\psi\|, \qquad (1.3)$$

for all $\psi\in\mathscr{H} = \mathscr{L}_2(\mathbb{R}^N)$, where $E(p)$ is the projection operator associated with the proposition p, $\hbar\mathbf{k} = \hbar(k_1,...,k_N)$ is the momentum, $\mathbf{v}$ and $\mathbf{w}$ are in $\mathbb{R}^N$, and $[\mathbf{a},\mathbf{b}]$ is the rectangular parallelepiped $[a_1,b_1]\times\cdots\times[a_N,b_N]$. In (1.3), V_t is the evolution operator for the nonrelativistic Hamiltonian $H_0 + V$, where H_0 is the kinetic energy operator $-\hbar^2\nabla^2/2m$. To prove this they assumed asymptotic completeness, when it follows immediately from their previously proved result that

$$\lim_{t\to\pm\infty}\|E(x\in[\mathbf{v}t,\mathbf{w}t])U_t\psi\|^2 = \|E(\hbar\mathbf{k}\in[m\mathbf{v},m\mathbf{w}])\psi\|^2, \qquad (1.4)$$

in the case where U_t is the evolution operator for a nonrelativistic particle moving freely in N dimensions.

The result expressed by the statement, "For a free quantum mechanical particle, the probability that a position measurement will find the particle in a region $t\Delta$ at time t, tends as $t\to\infty$ to the probability that the velocity is in Δ," has also been obtained in the nonrelativistic case by Strichartz,[4] as a consequence of work on the asymptotic behavior of waves.

In this paper we shall consider evolution under an evolution operator U_t of the form

$$U_t = F^{-1}\exp(-i\omega t)F, \qquad (1.5)$$

where ω is a continuous function of $\mathbf{k} = (k_1,...,k_N)$ with continuous first and second derivatives. According to the principles of quantum mechanics, if

$$\hat{\mathbf{v}} = F^{-1}\nabla\omega F, \qquad (1.6a)$$

or, equivalently,

$$\hat{v}_i = F^{-1}\nabla_i\omega F \quad (i = 1,...,N) \qquad (1.6b)$$

where

$$\nabla_i\omega = \frac{\partial\omega}{\partial k_i}, \qquad (1.7)$$

then $\hat{v}_i$ is a self-adjoint operator and so can, in principle, represent an observable. Any state μ will therefore assign to the proposition $\mathbf{v}\in\mathscr{B}$ (where $\mathscr{B}$ is a Borel subset of $\mathbf{R}^N$) a probability $\mathrm{Prob}(\mathbf{v}\in\mathscr{B}|\mu)$. Suppose $\mathscr{B}$ is the region between the planes $v_i = u$, $v_i = w$ $(u < w)$, so $\mathbf{v}\in\mathscr{B}$ is equivalent to $v_i\in[u,w]$. We shall show that if the state evolves according to (1.5) then

$$\lim_{t\to\infty}\mathrm{Prob}(x_i\in[ut,wt]|\hat{U}_t\mu) = \mathrm{Prob}(v_i\in[u,w]|\mu),$$

 0022-2488/90/061435-06$03.00

where $\hat{U}_t$ is the evolution operator in state space induced by U_t.

Our result is attractive in several ways. First, it requires of ω only that it be continuous with continuous first and second derivatives. Second, it is true for any quantum state. Finally, it leads to the interpretation of $\mathbf{v}$ as the velocity by the time-of-flight method. Although the time-of-flight definition of momentum is not new—see, for example, Refs. 5 and 6 in the nonrelativistic case—our results are, we believe, more general.

The plan of the paper is as follows. In Sec. II we derive a simple lemma on stationary phase. This is used in Sec. III to derive the asymptotic localization of wave packets, and in Sec. IV to derive the asymptotic localization of position probability. In Sec. V the measurement of velocity is discussed in the light of these results, and we summarize our conclusions in Sec. VI.

II. A BASIC LEMMA ON THE OSCILLATION OF CERTAIN INTEGRALS

Let $\theta(\cdot)$ be a continuous function with continuous first and second derivatives on the real finite interval $[k_1,k_2]$ such that

$$|\theta'(k)|\geqslant m', \quad |\theta''(k)|\leqslant M'' \quad (k_1\leqslant k\leqslant k_2), \tag{2.1}$$

where m' and M'' are positive constants, and let ψ be the complex number

$$\psi=\int_{k_1}^{k_2}\exp[i\theta(k)]dk. \tag{2.2}$$

Since $|\theta'(k)|\geqslant m'>0$, θ is either strictly increasing or strictly decreasing. For definiteness we shall suppose that θ is strictly increasing, since the other case is similar.

Let n be the largest non-negative integer such that

$$\theta(k_1)+2n\pi\leqslant\theta(k_2), \quad \theta(k_1)+2(n+1)\pi>\theta(k_2). \tag{2.3}$$

Furthermore, by the intermediate value theorem we can define a finite, strictly increasing sequence of points $\kappa_0,\kappa_1,...,\kappa_{n+1}$ by

$$\theta(\kappa_r)=\theta(k_1)+2r\pi \quad (r=0,...,n), \quad \kappa_{n+1}=k_2. \tag{2.4}$$

We now define complex numbers $\psi_1,...,\psi_{n+1}$ by

$$\psi_r=\int_{\kappa_{r-1}}^{\kappa_r}\exp[i\theta(k)]dk \quad (r=1,...,n+1); \tag{2.5}$$

then, by (2.2),

$$\psi=\sum_{r=1}^{n+1}\psi_r. \tag{2.6}$$

Suppose $1\leqslant r\leqslant n$. By (2.4) and (2.5),

$$\psi_r=\int_{\theta_r}^{\theta_r+2\pi}\exp[i\theta(k)]\frac{dk}{d\theta}d\theta, \quad \theta_r=\theta(\kappa_{r-1}).$$

Therefore

$$\psi_r=\int_{\theta_r}^{\theta_r+2\pi}\left[\frac{1}{\theta'(k)}-\frac{1}{\theta'(\kappa_r)}\right]\exp[i\theta(k)]d\theta,$$

since the second integral vanishes. It follows by the mean value theorem that

$$\psi_r=\int_{\theta_r}^{\theta_r+2\pi}\frac{(\kappa_r-k)\theta''(\kappa)}{[\theta'(\kappa)]^2}\exp[i\theta(k)]d\theta,$$

where $\kappa_{r-1}<\kappa<\kappa_r$. Since $\kappa_{r-1}\leqslant k\leqslant\kappa_r$, we easily deduce, using (2.1) that

$$|\psi_r|\leqslant 2\pi(\kappa_r-\kappa_{r-1})M''/(m')^2 \quad (r=1,...,n). \tag{2.7}$$

Note that $\kappa_0=k_1$ and $\kappa_n\leqslant k_2$, so it follows from (2.7) that

$$\left|\sum_{r=1}^{n}\psi_r\right|\leqslant\frac{2\pi(k_2-k_1)M''}{(m')^2}. \tag{2.8}$$

By the mean value theorem,

$$(k_2-\kappa_n)\theta'(\kappa)=\theta(k_2)-\theta(\kappa_n),$$

where $\kappa_n<\kappa<k_2$. Since $\theta(k_2)-\theta(\kappa_n)<2\pi$ and $\theta'(\kappa)\geqslant m'$, we obtain $k_2-\kappa_n\leqslant 2\pi/m'$. Now, $k_2=\kappa_{n+1}$, so (2.5) gives

$$|\psi_{n+1}|\leqslant 2\pi/m'. \tag{2.9}$$

Combining (2.6), (2.8), and (2.9), we obtain

$$|\psi|\leqslant 2\pi(k_2-k_1)M''/(m')^2+2\pi/m'.$$

A similar argument is applicable in the case when $\theta'(k)\leqslant -m'$ $(k_1\leqslant k\leqslant k_2)$. We have thus proved the following lemma.

Lemma 2.1: Let $\theta(\cdot)$ be a continuous function with continuous first and second derivatives on the finite interval $[k_1,k_2]$ such that

$$|\theta'(k)|\geqslant m'>0, \quad |\theta''(k)|\leqslant M'' \quad (k_1\leqslant k\leqslant k_2). \tag{2.10}$$

Then if

$$\psi=\int_{k_1}^{k_2}\exp[i\theta(k)]dk, \tag{2.11}$$

$$|\psi|\leqslant 2\pi(k_2-k_1)M''/(m')^2+2\pi/m'. \tag{2.12}$$

III. THE ASYMPTOTIC LOCALIZATION OF WAVE PACKETS

Let $\mathscr{K}$ be the finite N-dimensional parallelopiped $[\mathbf{k}_1,\mathbf{k}_2]$, $c_{\mathscr{K}}$ be the characteristic function of $\mathscr{K}$, and $\psi=(2\pi)^{N/2}F^{-1}c_{\mathscr{K}}$, where F is the Fourier–Plancherel operator. Define the evolution operator U_t by

$$U_t=F^{-1}\exp(-i\omega t)F. \tag{3.1}$$

This U_t is the evolution operator corresponding to a Hamiltonian $H_0=\hbar F^{-1}\omega F$; $\omega=\omega(\mathbf{k})$ will be assumed to be continuous, and to have continuous first and second derivatives, everywhere in $\mathbf{k}$-space. Since $\psi=(2\pi)^{N/2}F^{-1}c_{\mathscr{K}}$,

$$U_t\psi(\mathbf{x})=\int c_{\mathscr{K}}(\mathbf{k})\exp[i(\mathbf{k}\cdot\mathbf{x}-\omega t)]d^N k. \tag{3.2}$$

Suppose $N\geqslant 2$, and let $\mathbf{k}_i=(0,...,0,k_i,0,...,0)$, $\mathbf{k}_i^\perp=(k_1,...,k_{i-1},0,k_{i+1},...,k_N)$, with $\mathbf{x}_i$ and $\mathbf{x}_i^\perp$ defined similarly. Then $\mathbf{k}=\mathbf{k}_i+\mathbf{k}_i^\perp$ and $\mathbf{x}=\mathbf{x}_i+\mathbf{x}_i^\perp$.

Denote by $d^{N-1}x_i$ the volume element $dx_1\cdots dx_{i-1}dx_{i+1}\cdots dx_N$ and by $\int\cdots d^{N-1}x_i^\perp$ an integral over all values of the $N-1$ coordinates $x_1,...,x_{i-1}$, $x_{i+1},...,x_N$, with a similar notation with k replacing x. A straightforward calculation using (3.2) and Fourier's intergral theorem yields, if $N\geqslant 2$,

$$\int |U_t\psi(\mathbf{x})|^2 \, d^{N-1}x_i^\perp$$
$$= (2\pi)^{N-1}\int \left| \int c_{\mathscr{K}}(\mathbf{k})\exp[i(k_i x_i - \omega t)]dk_i \right|^2 \times d^{N-1}k_i^\perp. \tag{3.3}$$

If $N=1$, (3.3) remains valid if the integrals over $x_i^\perp$ and $k_i^\perp$ are omitted and $i=1$.

When nonzero, the integral between the modulus signs on the right-hand side of (3.3) is of the form (2.11) with

$$\theta(k_i) = k_i x_i - \omega t,$$

the other components of **k** being fixed. Recalling (1.6),

$$\frac{\partial \theta}{\partial k_i} = x_i - v_i t. \tag{3.4}$$

Suppose that v_L and v_U are real numbers satisfying

$$v_L < v_{i\,\min} \equiv \min_{\mathbf{k}\in\mathscr{K}} v_i, \quad v_U > v_{i\,\max} \equiv \max_{\mathbf{k}\in\mathscr{K}} v_i; \tag{3.5}$$

such numbers exist since $v_i = \partial\omega/\partial k_i$ is continuous everywhere by hypothesis. Suppose further that $t>0$; then, by (3.4) and (3.5),

$$x_i - v_U t < x_i - v_{i\,\max} t \leqslant \frac{\partial\theta}{\partial k_i} \leqslant x_i - v_{i\,\min} t < x_i - v_L t. \tag{3.6}$$

Finally suppose $x_i \leqslant v_L t$. Then, by (3.6), $x_i - v_{i\,\min} t < 0$, so we can apply Lemma 2.1 with $m' = |x_i - v_{i\,\min} t| = v_{i\,\min} t - x_i$. Also, from (3.4), since $v_i = \partial\omega/\partial k_i$,

$$\frac{\partial^2\theta}{\partial k_i^2} = -\frac{\partial^2\omega}{\partial k_i^2} t;$$

hence, if

$$\Omega_{ii} = \max_{\mathbf{k}\in\mathscr{K}} \left| \frac{\partial^2\omega}{\partial k_i^2} \right|, \tag{3.7}$$

we can take $M'' = \Omega_{ii} t$. Thus by (2.11) and (2.12) (Lemma 2.1),

$$\left| \int c_{\mathscr{K}}(\mathbf{k})\exp[i(k_i x_i - \omega t)]dk_i \right|$$
$$\leqslant \frac{2\pi}{v_{i\,\min} t - x_i}\left\{ \frac{(k_{2i}-k_{1i})\Omega_{ii} t}{v_{i\,\min} t - x_i} + 1 \right\}$$
$$\leqslant \frac{2\pi}{v_{i\,\min} t - x_i}\left[\frac{(k_{2i}-k_{1i})\Omega_{ii}}{v_{i\,\min} - v_L} + 1 \right], \tag{3.8}$$

where the last inequality follows from the fact that $x_i \leqslant v_L t$. It now follows immediately from (3.3) and (3.8) that, if $N \geqslant 2$,

$$\int |U_t\psi(\mathbf{x})|^2 \, d^{N-1}x_i^\perp \leqslant \frac{c}{(x_i - v_{i\,\min} t)^2}, \tag{3.9}$$

where

$$c = (2\pi)^{N+1}\left[\frac{(k_{2i}-k_{1i})\Omega_{ii}}{v_{i\,\min} - v_L} + 1 \right]^2 \prod_{\substack{j=1 \\ (j\neq i)}}^{N} (k_{2j} - k_{1j}). \tag{3.10}$$

If $N=1$, (3.9) and (3.10) remain valid if, in (3.9), the integral over $x_i^\perp$ is omitted while in (3.10) the product over j is omitted, and $i=1$.

From (3.9), we obtain

$$\int_{-\infty}^{v_L t} dx_i \int d^{N-1}x_i^\perp |U_t\psi(\mathbf{x})|^2 \leqslant \frac{c}{(v_{i\,\min} - v_L)t} \to 0$$

when $t\to\infty$.

If $N=1$, this is valid if the integral over $x_i^\perp$ is omitted and $i=1$, and so

$$\lim_{t\to\infty} \|E(x_i \in (-\infty, v_L t])U_t\psi\|^2 = 0 \quad (N\in\mathbb{N}). \tag{3.11a}$$

Similarly

$$\lim_{t\to\infty} \|E(x_i \in [v_U t, \infty))U_t\psi\|^2 = 0 \quad (N\in\mathbb{N}). \tag{3.11b}$$

We have proved (3.11) on the assumption that $\psi = (2\pi)^{N/2}F^{-1}c_{\mathscr{K}}$ and $\mathscr{K}$ is the parallelopiped $[\mathbf{k}_1, \mathbf{k}_2]$. It follows from the linearity of the operators involved that it is also valid if ψ is a finite linear combination of such functions, v_L and v_U still being given by (3.5), but now $\mathscr{K} = \operatorname{supp} F\psi$, the support of $F\psi$. Such linear combinations are dense in the space of wave functions whose Fourier transforms have support contained in some compact subset $\mathscr{K}$ of $\mathbb{R}^N$; finite v_L and v_U then exist that satisfy (3.5) since v_i is everywhere continuous by assumption. Since the operators involved are bounded it is straightforward to show that (3.11) are valid if $\operatorname{supp} F\psi \subseteq \mathscr{K}$. Finally suppose of ψ only that v_i is bounded on $\operatorname{supp} F\psi$. Then $v_{i\,\min}$ and $v_{i\,\max}$ may not exist, but the greatest lower bound $\underline{v}_i$ and least upper bound $\bar{v}_i$ do. Thus ψ can be approximated arbitrarily closely by functions whose Fourier transforms have compact support [for example, by $E(\mathbf{k}\in\mathscr{S})\psi$, where $\mathscr{S}$ is a sphere of arbitrarily large radius and center the origin], and so (3.11) remains valid in this case also if $v_L < \underline{v}_i$, $v_U > \bar{v}_i$. We have therefore proved the following proposition.

Proposition 3.1: Suppose $\psi \in \mathscr{L}_2(\mathbb{R}^N)$ has the property that $\nabla_i\omega$ is bounded on $\mathscr{K} = \operatorname{supp} F\psi$. Let v_L, v_U be numbers satisfying $v_L < \underline{v}_i$, $v_U > \bar{v}_i$, where $\underline{v}_i$ and $\bar{v}_i$ are the greatest lower bound and least upper bound, respectively, of $v_i = \nabla_i\omega$ on $\operatorname{supp} F\psi$. Then

$$\lim_{t\to\infty} E(x_i \in (-\infty, v_L t])U_t\psi = 0, \tag{3.12a}$$

$$\lim_{t\to\infty} E(x_i \in [v_U t, \infty))U_t\psi = 0. \tag{3.12b}$$

It will be convenient from now on to abbreviate our notation. Let I be an interval of $\mathbb{R}$; then we write

$$E_x I \equiv E(x\in I), \quad E_v I \equiv E(v_i \in I). \tag{3.13}$$

For example, $E_v(u,w]$ means $E(v_i\in(u,w])$, etc. In this notation Eqs. (3.12) may be written

$$\lim_{t\to\infty} E_x(-\infty, v_L t]U_t\psi = 0, \tag{3.14a}$$

$$\lim_{t\to\infty} E_x[v_U t, \infty)U_t\psi = 0. \tag{3.14b}$$

Let u, w, and δ be real numbers such that $u < u+\delta < w-\delta < w$, and consider

$$E_x(-\infty, ut]U_t E_v[u,w]\psi.$$

This may be written

$$E_x(-\infty,ut]U_tE_v[u,u+\delta]\psi$$
$$+E_x(-\infty,ut]U_tE_v(u+\delta,w]\psi.$$

Suppose

$$E_v[u,u+]\psi\equiv\lim_{\delta\to 0+}E_v[u,u+\delta]\psi=0.$$

Then given $\varepsilon>0$ we can choose δ (depending on ψ) so that

$$\|E_v[u,u+\delta]\psi\|<\tfrac{1}{2}\varepsilon,$$

hence

$$\|E_x(-\infty,ut)U_tE_v[u,u+\delta]\psi\|<\tfrac{1}{2}\varepsilon.$$

By Proposition 3.1, we can choose t_0, depending on δ and ψ, so that, for $t\geqslant t_0$,

$$\|E_x(-\infty,ut)U_tE_v(u+\delta,w]\psi\|<\tfrac{1}{2}\varepsilon.$$

It follows that, for $t\geqslant t_0$,

$$\|E_x(-\infty,ut]U_tE_v[u,w]\psi\|<\varepsilon$$

and so $E_x(-\infty,ut]U_tE_v[u,w]\psi\to 0$ when $t\to\infty$. Similarly if $E_v[w-,w]\psi=0$, then $E_x[wt,\infty)U_tE_v[u,w]\psi\to 0$ when $t\to\infty$.

Our results are summarized by the following proposition.

Proposition 3.2: Suppose v_i is bounded on Supp $F\psi$ while u and w are real numbers such that $u<w$; suppose further that

$$E_v[u,u+]\psi=E_v[w-,w]\psi=0. \tag{3.15}$$

Then

$$\lim_{t\to\infty}E_x(-\infty,ut]U_tE_v[u,w]\psi=0, \tag{3.16a}$$

$$\lim_{t\to\infty}E_x[wt,\infty)U_tE_v[u,w]\psi=0. \tag{3.16b}$$

Equation (3.15) may not be satisfied. For example, suppose $N=1$ and $\omega=ck$, where c is a positive constant, while $u=c$. Then $E_v[u,u+\delta]=E_v[c,c+\delta]=I$, since the velocity $d\omega/dk=c\in[c,c+\delta]$, for all positive values of δ. On the other hand, if v_i is a strictly increasing or decreasing function of k_i when the other components of $\mathbf{k}$ are fixed (3.15) is valid for all ψ.

Proposition 3.3: Suppose

$$E_v[u,u+]\psi=E_v[u-,u]\psi=E_v[w-,w]\psi$$
$$=E_v[w,w+]\psi=0; \tag{3.17}$$

then

$$\lim_{t\to\infty}\{E_x[ut,wt]U_t\psi-U_tE_v[u,w]\psi\}=0, \tag{3.18}$$

$$\lim_{t\to\infty}\|E_x[ut,wt]U_t\psi\|=\|E_v[u,w]\psi\|. \tag{3.19}$$

(This result asserts the asymptotic localization of a wave packet with velocities in the interval $[u,w]$ in the interval $[ut,wt]$ [cf. (1.4)].)

Proof:

$$E_x[ut,wt]U_t\psi-E_x[ut,wt]U_tE_v[u,w]\psi$$
$$=E_x[ut,wt]U_tE_v(-\infty,u)\psi+E_x[ut,wt]$$
$$\times U_tE_v(w,\infty)\psi. \tag{3.20}$$

Now

$$E_x[ut,wt]U_tE_v(-\infty,u)\psi$$
$$=E_x[ut,wt]U_tE_v(-\infty,u')\psi+E_x[ut,wt]$$
$$\times U_tE_v[u',u)\psi, \tag{3.21}$$

where $u'<u$. Since $E_x[ut,wt]$ and U_t are bounded and $E_v(-\infty,u')\psi\to 0$ when $u'\to-\infty$ we can choose u' so that the norm of the first term on the right-hand side of (3.21) is less than $\tfrac{1}{2}\varepsilon$. Also

$$\|E_x[ut,wt]U_tE_v[u',u)\psi\|\leqslant\|E_x[ut,\infty)U_tE_v[u',u]\psi\|,$$

which tends to zero when $t\to\infty$ by (3.16b), since $E_v[u-,u]\psi=0$. We can therefore choose t_0 so that, if $t\geqslant t_0$, the norm of the second term on the right-hand side of (3.21) is less than $\tfrac{1}{2}\varepsilon$. It follows that, for $t\geqslant t_0$, the norm of the left-hand side of (3.21) is less than ε; that is, the first term on the right-hand side of (3.20) tends to zero when t tends to infinity. Similarly the second term on the right-hand side of (3.20) tends to zero when t tends to infinity, so

$$\lim_{t\to\infty}\{E_x[ut,wt]U_t\psi-E_x[ut,wt]U_tE_v[u,w]\psi\}=0. \tag{3.22}$$

Note now that

$$E_x[ut,wt]U_tE_v[u,w]\psi-U_tE_v[u,w]\psi$$
$$=-E_x(-\infty,ut)U_tE_v[u,w]\psi$$
$$-E_x(wt,\infty)U_tE_v[u,w]\psi.$$

By Proposition 3.2 each term on the right-hand side tends to zero when t tends to infinity since $E_v[u,u+]\psi=E_v[w-,w]\psi=0$, so

$$\lim_{t\to\infty}\{E_x[ut,wt]U_tE_v[u,w]\psi-U_tE_v[u,w]\psi\}=0. \tag{3.23}$$

Equation (3.18) now follows from (3.22) and (3.23). Equation (3.19) is an immediate consequence of (3.18), since U_t is unitary. □

IV. THE ASYMPTOTIC LOCALIZATION OF POSITION PROBABILITY

A general state μ in quantum mechanics assigns to every proposition p a probability $\text{Prob}(p|\mu)$. By Gleason's theorem it is given by

$$\text{Prob}(p|\mu)=\sum_j w_j\|E(p)\psi_j\|^2, \tag{4.1}$$

where $\{w_j\}$ is a countable set of positive numbers with unit sum, and $\{\psi_j\}$ is a corresponding set of unit vectors in $\mathscr{H}$. The pair

$$(\{w_j\},\{\psi_j\}) \tag{4.2}$$

determines the state, but the converse is false. Given any state μ there is usually more than one pair of the form (4.2) (Ref. 7, p. 9). We shall call the pair (4.2) a *representation* of the state.

If μ is the state at time zero it becomes a state $\hat{U}_t\mu$ at time t represented by the pair $(\{w_j\},\{U_t\psi_j\})$ [$\hat{U}_t$ is a (nonlinear) mapping of states into states]. If u and w are real and $u\leqslant w$,

$$\text{Prob}(x_i \in [ut,wt\,] \,|\, \hat{U}_t\mu) = \sum_j w_j \|E(x_i \in [ut,wt\,])\, U_t \psi_j\|^2.$$

In the abbreviated notation (3.13) this is written

$$\text{Prob}(x_i \in [ut,wt\,] \,|\, \hat{U}_t\mu) = \sum_j w_j \|E_x[ut,wt\,]\, U_t \psi_j\|^2. \tag{4.3}$$

Suppose that, for all values of j,

$$E_v[u-,u]\psi_j = E_v[u,u+\,]\psi_j = E_v[w-,w]\psi_j = E_v[w,w+\,]\psi_j = 0. \tag{4.4}$$

Since the series on the right-hand side of (4.3) is uniformly convergent we can use (3.19) (Proposition 3.3) to obtain

$$\lim_{t\to\infty} \text{Prob}(x_i \in [ut,wt\,] \,|\, \hat{U}_t\mu) = \sum_j w_j \|E_v[u,w]\psi_j\|^2. \tag{4.5}$$

The right-hand side of (4.5) is the probability in the state μ that the value v_i of the observable represented by the self-adjoint operator $\hat{v}_i = F^{-1}\nabla_i \omega F$ is in the closed interval $[u,w]$; that is,

$$\lim_{t\to\infty} \text{Prob}(x_i \in [ut,wt\,] \,|\, \hat{U}_t\mu) = \text{Prob}(v_i \in [u,w] \,|\, \mu).$$

Since

$$\text{Prob}(v_i \in [u,w] \,|\, \mu) = \sum_j w_j \|E_v[u,w]\psi_j\|^2, \tag{4.6}$$

it is easy to see that the condition (4.4) is logically equivalent to

$$\text{Prob}(v_i \in [u-,u] \,|\, \mu) = \text{Prob}(v_i \in [u,u+\,] \,|\, \mu) = \text{Prob}(v_i \in [w-,w] \,|\, \mu) = \text{Prob}(v_i \in [w,w+\,] \,|\, \mu) = 0. \tag{4.7}$$

It is straightforward to show that (4.7) is equivalent to saying that the probability distribution of $\hat{v}_i$ is not concentrated at either u or w. We have therefore proved the following theorem.

Theorem 4.1: Let u and w be real numbers such that $u < w$. If the probability distribution of the observable $\hat{v}_i = F^{-1}\nabla_i \omega F$ is not concentrated at either of the points u and w of the real line, then

$$\lim_{t\to\infty} \text{Prob}(x_i \in [ut,wt\,] \,|\, \hat{U}_t\mu) = \text{Prob}(v_i \in [u,w] \,|\, \mu). \tag{4.8}$$

Theorem 4.1 has a simple consequence. Suppose the state μ is such that all values of $\hat{v}_i$ must lie in $[u,w]$, so that the right-hand side of (4.8) is unity. If this is the case the left-hand side is also unity; this means that asymptotically the probability distribution of the ith position coordinate is localized in $[ut,wt]$.

V. THE MEASUREMENT OF VELOCITY

In this section we shall discuss the measurement of velocity in quantum mechanics in the light of Theorem 4.1.

First, we note that if $t > 0$ the proposition $u \leqslant x_i/t \leqslant w$ is logically equivalent to the proposition $x_i \in [ut,wt\,]$; hence, by (4.8),

$$\lim_{t\to\infty} \text{Prob}(u \leqslant x_i/t \leqslant w \,|\, \hat{U}_t\mu) = \text{Prob}(v_i \in [u,w] \,|\, \mu). \tag{5.1}$$

Now to measure the velocity given the initial state μ we should (i) take a large time t so that initial and final uncertainties in the position are small, and (ii) ensure that the particle is moving "freely" throughout. The proposition $u \leqslant x_i/t \leqslant w$ becomes, in the limit as $t \to \infty$, the proposition "$u \leqslant$ the ith component of velocity $\leqslant w$," so (5.1) gives

$$\text{Prob}(u \leqslant i\text{th component of velocity} \leqslant w \,|\, \mu) = \text{Prob}(v_i \in [u,w] \,|\, \mu). \tag{5.2}$$

Equation (5.2) shows that $\hat{v}_i$ is the self-adjoint operator representing velocity provided $\hat{U}_t$ represents "free" motion.

We need to discuss the notion of "free motion." This is easy to answer in the case of the motion of a particle in three dimensions under nonrelativistic quantum mechanics with a local potential V, since, by Ehrenfest's theorem,

$$\frac{d^2}{dt^2}\langle x_i \rangle = \langle -\nabla_i V \rangle. \tag{5.3}$$

This means that if $\langle x_i \rangle$ is a linear function of time, $\langle -\nabla_i V \rangle$ vanishes for all states. From this it is easy to see that $\nabla V = \mathbf{0}$ and so V is a constant. Conversely, if V is constant, $\nabla V = \mathbf{0}$, and so, by (5.3), $\langle x_i \rangle$ is a linear function of time. If free motion means the absence of a force it is therefore equivalent to saying that $\langle x_i \rangle$ is a linear function of time.

In other cases "free motion" is a question of definition. However, it can be shown that if $U_t = F^{-1}\exp(-i\omega t)F$ the expectation value of position, if it exists, is a linear function of time (the proof is given in the Appendix).

An interesting special case is given by

$$\omega = (\hbar^2 \mathbf{k}^2 c^2 + m^2 c^4)^{1/2}/\hbar \quad (m > 0), \tag{5.4}$$

which is important in the case of the motion of a particle in relativistic quantum mechanics, where c is the speed of light. In this case

$$\nabla\omega = \hbar \mathbf{k} c^2 (\hbar^2 \mathbf{k}^2 c^2 + m^2 c^4)^{-1/2}. \tag{5.5}$$

From (5.5) we see immediately that $|\mathbf{v}| = |\nabla\omega| < c$, as required by special relativity.

This result has also been derived by Ruijsenaars, who termed it "asymptotic causality."[8] This does not mean that superluminal velocities cannot be observed over finite distances, but Ruijsenaars has argued that such superluminal velocities can never be observed in practice.[8]

VI. CONCLUSIONS

If a particle moves in N dimensions under the evolution operator $U_t = F^{-1}\exp(-i\omega t)F$, where ω is a continuous function of $\mathbf{k}$ with continuous first and second derivatives, and $\hat{v}_i = F^{-1}\nabla_i \omega F$, then the probability distribution of $\hat{v}_i$ satisfies the asymptotic condition (4.8) provided it is continuous at u and w. Further, if the expectation value $\langle x_i \rangle$ of the ith position coordinate exists then it is a linear function of time, its time rate of change being $\langle \hat{v}_i \rangle$, consistent with the interpretation of $\hat{v}_i$ as the velocity. However, in relativistic quantum mechanics there are deep problems involved with the position operator which are outside the scope of this paper (see, for example, Ref. 8, and references therein).

APPENDIX

Note: In this appendix x, v, and $\hat{v}$ will be used as abbreviations for x_i, v_i, and $\hat{v}_i$, respectively.

The assertion in Sec. VI that the time rate of change of $\langle x\rangle$ is constant and equal to $\langle\hat{v}\rangle$ is made precise by the following theorem.

Theorem A.1: Let $\langle\hat{v}\rangle$ be the expectation value of $\hat{v}$ in the state μ and $\langle x\rangle_t$ be the expectation value of x in the state $\hat{U}_t\mu$. If $\langle\hat{v}\rangle$ exists, and $\langle x\rangle_t$ exists for all real t, then

$$\frac{d}{dt}\langle x\rangle_t = \langle\hat{v}\rangle. \qquad (A1)$$

The purpose of this appendix is to prove Theorem A.1. In order to do this we shall need two preliminary lemmas.

Lemma A.1: Let ψ be an arbitrary function in the Hilbert space of square-integrable functions on $\mathbb{R}^N$, and define the bounded operator $U_t(\theta)$ by

$$U_t(\theta) = F^{-1}\exp[-i\omega(\mathbf{k}-\theta\mathbf{e})t]F, \qquad (A2)$$

where θ is a real number and $\mathbf{e}$ is the unit vector in the direction of the ith coordinate axis. If ω is everywhere continuous with continuous first and second derivatives,

$$\lim_{\theta\to 0}[U_t(\theta)\psi - U_t\psi] = 0.$$

Proof: Let $\mathscr{S}$ be the sphere $\{\mathbf{k}: |\mathbf{k}|\leqslant K\}$, where K is some positive number, and $\mathscr{S}'$ be the rest of $\mathbf{k}$-space; then

$$\|[U_t(\theta) - U_t]\psi\|^2 = \left(\int_{\mathscr{S}} + \int_{\mathscr{S}'}\right)|\{\exp[-i\omega(\mathbf{k}-\theta\mathbf{e})t] - \exp[-i\omega(\mathbf{k})t]\}F\psi(\mathbf{k})|^2\, d^Nk.$$

If ε is a positive number, the integral over $\mathscr{S}'$ can be made smaller than $\frac{1}{2}\varepsilon$ by taking K sufficiently large. Since $\nabla_i\omega$ is continuous and therefore bounded on $\mathscr{S}$, the mean value theorem may be used to show that the integral over $\mathscr{S}$ is less than $\frac{1}{2}\varepsilon$ for all sufficiently small θ. This establishes the lemma. □

Lemma A.2: If ψ is in the domains of x and $\hat{v}$ then so is $U_t\psi$; moreover,

$$xU_t\psi = U_t x\psi + \hat{v}tU_t\psi. \qquad (A3)$$

Proof: First note that

$$\begin{aligned} F\exp(ix\theta)U_t\psi(\mathbf{k}) &= FU_t\psi(\mathbf{k}-\theta\mathbf{e}) \\ &= \exp[-i\omega(\mathbf{k}-\theta\mathbf{e})t]F\psi(\mathbf{k}-\theta\mathbf{e}) \\ &= \exp[-i\omega(\mathbf{k}-\theta\mathbf{e})t]F\exp(ix\theta)\psi; \end{aligned}$$

so, by (A2),

$$\exp(ix\theta)U_t\psi = U_t(\theta)\exp(ix\theta)\psi.$$

If we subtract $U_t\psi$ from both sides, divide by $\theta\neq 0$, and then rearrange the right-hand side, we obtain

$$\begin{aligned} \frac{\exp(ix\theta)-I}{\theta}U_t\psi &= U_t\frac{\exp(ix\theta)-I}{\theta}\psi + \frac{U_t(\theta)-U_t}{\theta}\psi \\ &\quad + [U_t(\theta)-U_t]A_\theta\psi, \end{aligned} \qquad (A4)$$

where $A_\theta = [\exp(ix\theta) - I]/\theta$.

Let $\theta\to 0$. Since ψ is in the domain of x, Stone's theorem shows that the first term on the right-hand side of (A4) tends to $iU_t x\psi$. Since ψ is in the domain of $\hat{v}$ it is easy to see that $U_t\psi$ is also in this domain; hence by Stone's theorem the second term tends to $i\hat{v}tU_t\psi$. Stone's theorem also shows that $A_\theta\psi\to ix\psi = \phi$, say. Now

$$[U_t(\theta)-U_t]A_\theta\psi = [U_t(\theta)-U_t]\phi + [U_t(\theta)-U_t](A_\theta\psi - \phi).$$

The first term on the right-hand side tends to zero when $\theta\to 0$ by Lemma A.1, while

$$\|[U_t(\theta)-U_t][A_\theta\psi - \phi]\|\leqslant 2\|A_\theta\psi - \phi\|\to 0$$

when $\theta\to 0$, so the third term on the right-hand side of (A4) tends to zero when $\theta\to 0$. It follows that the left-hand side of (A4) has a limit as $\theta\to 0$; hence, by Stone's theorem, $U_t\psi$ is in the domain of x, and the limit of the left-hand side is $ixU_t\psi$. Equation (A3) now follows from (A4). □

The premises of Theorem A.1 shows that each ψ_j is in the domain of x and $\hat{v}$, so it follows from Lemma A.2 that, for each value of j,

$$\langle U_t\psi_j|x|U_t\psi_j\rangle = \langle\psi_j|x|\psi_j\rangle + \langle\psi_j|\hat{v}|\psi_j\rangle t. \qquad (A5)$$

We can multiply (A5) by w_j and sum over j to obtain

$$\langle x\rangle_t = \langle x\rangle + \langle v\rangle t, \qquad (A6)$$

the convergence of the series being assured by the premises of Theorem A.1. Equation (A6) implies (A1), so Theorem A.1 follows. □

[1] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, III. Scattering Theory* (Academic, New York, 1979).
[2] K. K. Wan and R. G. D. McLean, Phys. Lett. A **94**, 198 (1983).
[3] K. K. Wan and R. G. D. McLean, Phys. Lett. A **95**, 76 (1983).
[4] R. S. Strichartz, J. Funct. Anal. **40**, 341 (1981).
[5] J. L. Park and H. Margenau, Int. J. Theor. Phys. **1**, 211 (1968).
[6] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), pp. 96–100.
[7] E. G. Beltrametti and G. Cassinelli, *The Logic of Quantum Mechanics* (Addison-Wesley, Reading, MA, 1981).
[8] S. N. Ruijsenaars, Ann. Phys. (NY) **137**, 33 (1981).

The one-dimensional Coulomb potential as a generalized function and the hidden $O(2)$ symmetry

T. A. Weber and C. L. Hammer
Department of Physics, Iowa State University, Ames, Iowa 50011

(Received 14 April 1989; accepted for publication 17 January 1990)

The one-dimensional hydrogen atom problem is solved by treating the potential, $-\lambda/|x|$, as a generalized function. The solutions (although nondegenerate) are nonunique unless fixed by some physical constraint. It is also shown that the hidden $O(2)$ symmetry is a consequence of using solutions that are eigenfunctions of the operator $\operatorname{sgn} x \equiv x/|x|$.

I. INTRODUCTION

Over the years there has been much discussion[1-7] concerning the bound state solutions of the one-dimensional Schrödinger equation ($m = \hbar = 1$)

$$H\psi = E\psi, \tag{1}$$

with

$$H = -\frac{1}{2}\frac{d^2}{dx^2} - \frac{\lambda}{|x|}; \quad E = -\frac{k^2}{2}.$$

Because of the singularity at the origin, solutions must be obtained separately for regions $x > 0$ and $x < 0$ and then matched appropriately at $x = 0$. Since H is symmetric in x, a solution for $x > 0$ can be extended to $x < 0$ to obtain even and odd wavefunctions. We obtain two candidates for the even wavefunction

$$\psi_1(x,k) = A_1|x|e^{-k|x|}M(1-\lambda/k,\ 2,\ 2k|x|). \tag{2a}$$

and

$$\psi_2(x,k) = A_2 e^{-k|x|}U(-\lambda/k,\ 0,\ 2k|x|), \tag{2b}$$

where A_1 and A_2 are normalization constants and M and U are the regular and irregular confluent hypergeometric functions.[8] These wavefunctions are linearly dependent for the special cases where λ/k is a positive integer. The odd extensions are given by

$$\psi_{3,4} = \operatorname{sgn} x\, \psi_{1,2}. \tag{3}$$

The wavefunction ψ_1 is bounded as $|x| \to \infty$ only if M is a polynomial, i.e., only if $\lambda/k = n$; $n = 1,2,...$, the usual spectrum of the three-dimensional hydrogen atom in the case of zero angular momentum. This same energy spectrum applies to ψ_3, implying a double degeneracy. On the other hand, the wavefunction ψ_2 is continuous at $x = 0$ and is Lebesque square integrable for all values of $k > 0$; a result that implies a negative energy continuum.[2]

These strange results arise because $\psi_{1,2,4}$ are not solutions of Eq. (1). In particular, the matrix elements of $(H - E)$ with either ψ_2 or ψ_4 diverge.[7] Although both ψ_1 and ψ_3 satisfy

$$\int \psi^\dagger (H-E)\psi\, dx = 0 \tag{4}$$

for the same eigenvalues E (i.e., doubly degenerate), ψ_1 does not satisfy the original Schrödinger equation at the origin but rather the equation[7]

$$(H-E)\psi_1 = -\psi_1'(0_+)\delta(x), \tag{5}$$

where the prime denotes differentiation.

Loudon[1] attempts to regularize the potential by replacing it with

$$V = -\lim_{\alpha\to 0}[\lambda/(|x|+\alpha)] \tag{6}$$

and finds a ground state at $E = -\infty$ with the corresponding eigenfunction

$$\psi = \lim_{\alpha\to 0}(\alpha)^{1/2}e^{-|x|/\alpha}. \tag{7}$$

This ground state is not required for completeness in the expansion of square integrable functions.[3] The regularization (6) is inappropriate because the conditions placed on the wavefunction at the origin (i.e., ψ and ψ' continuous) are not satisfied in the limit of $\alpha = 0$. This is reminiscent of the Klauder phenomena.[9]

In Sec. II, we treat the potential as a generalized function and obtain regularized solutions that are even and odd but exhibit no degeneracy. Since $|x|^{-1}$ is not uniquely defined as a generalized function, the discrete energy spectrum of the even wavefunctions is arbitrary unless fixed by some physical constraint.

Davtyan *et al.*[6] attempt to solve the one-dimensional hydrogen atom by multiplying the Hamiltonian by $|x|$, thereby removing the singularity in the potential. They start their analysis by transforming the equation,

$$|x|H = |x|E\psi, \tag{8}$$

to the momentum representation and obtain the interesting result that the double degeneracy mentioned above is a manifestation of a hidden $O(2)$ symmetry. In Sec. III, we show that this result is due to the symmetry,

$$\operatorname{sgn} x\,|x|H\operatorname{sgn} x = |x|H, \tag{9}$$

valid when applied to wavefunctions zero at the origin. The double degeneracy then follows from the nonzero commutator of $\operatorname{sgn} x$ with the parity operator. The condition $\psi(0) = 0$ on all wavefunctions effectively divides the space into two disjointed half-spaces. This rather unsatisfactory condition is removed when the potential is treated as a generalized function.

 0022-2488/90/061441-04$03.00

II. THE POTENTIAL AS A GENERALIZED FUNCTION

Since the Hamiltonian of the one-dimensional hydrogen atom is symmetric under parity, we seek even and odd wavefunctions continuous at the origin. The odd wavefunctions are constructed from the regular solutions since these are zero at the origin. On the other hand, the even wavefunctions, if they exist, must be constructed from the irregular solutions that are nonzero at the origin. However, the derivative of these even wavefunctions will then have an infinite discontinuity at the origin. Any regularization of the potential must take this into account. The appropriate condition on the derivative is obtained by integrating the Schrödinger equation across the origin. Thus

$$\psi'(x) - \psi'(-x) = -2\int_{-x}^{x} (E+\lambda|x'|^{-1})\psi(x')dx'. \tag{10}$$

The derivative $\psi'(x)$ for the odd wavefunctions must be continuous since these wavefunctions are zero at the origin thus making the integrand finite throughout the range of integration. For these wavefunctions we therefore choose the odd extensions of the regular solution, $\psi_3(x,k)$, where the eigenvalue k is given by

$$\lambda/k=n;\quad n=1,2,3,\dots\ . \tag{11}$$

For the even wavefunctions, the integral is undefined if $|x|^{-1}$ is treated as an ordinary function. But $|x|^{-1}$ treated as a generalized function regularizes the integral while satisfying the equation

$$xf(x)=\operatorname{sgn} x. \tag{12}$$

A representation of this generalized function given by Lighthill[10] is

$$|x|^{-1}=\frac{d}{dx}(\operatorname{sgn} x \ln|x|). \tag{13}$$

It is apparent that this representation is not unique since another solution of (12) can be obtained by adding to $f(x)$ an arbitrary constant times the Dirac δ function. In the following we use

$$|x|^{-1}=\lim_{\alpha\to 0}\frac{1}{|x|+\alpha}[1+2\alpha\ln(q\alpha)\delta(x)] \tag{14}$$

$$=\lim_{\alpha\to 0}\frac{d}{dx}\operatorname{sgn} x[\ln(|x|+\alpha)+\ln q], \tag{15}$$

which is clearly equivalent to (13) but with the added term $2\ln q\delta(x)$, where q is a real positive constant.

Substituting the representation (14) into Eq. (10) gives for small positive x,

$$\psi'(x)=-2\lambda\psi(0)\ln qx, \tag{16}$$

where $\psi'(x)$ is odd since $\psi(x)$ is even. Note that this condition on the derivative is independent of the limiting parameter α unlike the result obtained from the prescription given in (6). The actual derivative of the irregular solution[8,11] for small positive x is

$$\psi'(x)=-2\lambda\psi(0)\times\left\{\frac{k}{2\lambda}+\ln 2kx+\frac{\Gamma'(1-\lambda/k)}{\Gamma(1-\lambda/k)}+2\gamma\right\}+O(x), \tag{17}$$

where γ is Euler's constant. Comparing (16) and (17) yields,

$$\frac{k}{2\lambda}+\ln\frac{2k}{\lambda}+\frac{\Gamma'(1-\lambda/k)}{\Gamma(1-\lambda/k)}+2\gamma=\ln\frac{q}{\lambda}, \tag{18}$$

the equation for the eigenvalues k.

The even wavefunctions that are continuous at the origin and satisfy the derivative condition (16) are the even extensions of the irregular solutions $\psi_2(x,k)$, where the spectrum is given by (18) for a specific q. As already mentioned, the odd wavefunctions are $\psi_3(x,k)$, where k is given by the usual S wave spectrum of the hydrogen atom, Eq. (11). We take these even and odd wavefunctions to form the complete set of solutions of the Schrödinger equation for the one-dimensional hydrogen atom.

Note that the bound state spectrum exhibits no degeneracy. In fact, for $q=\lambda$ we find that the ground state is described by the even wavefunction with

$$\lambda/k=0.656. \tag{19}$$

As expected the energy levels alternate between even and odd with the even levels approaching the odd from below as $k\to 0$. Let us repeat that unless q is determined by some physical constraint, the potential $-\lambda|x|^{-1}$ is not uniquely defined.

It can be verified that the Hamiltonian is Hermitian when acting on the wavefunctions $\psi_2(x,k)$ and $\psi_3(x,k)$, but it is no longer so if the domain is extended to include even wavefunctions with a different constant q or if it is extended to include $\psi_1(x,k)$.

III. SOLUTIONS IN MOMENTUM SPACE

Following Davtyan *et al.*[6] we seek a solution to

$$|x|\psi''(x)+2(E|x|+\lambda)\psi(x)=0 \tag{20}$$

by transforming to momentum space. The usual procedure is to treat the Fourier transform of $|x|\psi''(x)$ as the convolution of the individual transforms. However, $\psi''(x)$ goes as $|x|^{-1}$ at the origin if $\psi(0)\neq 0$. In this case both the transform of $|x|$ and of $\psi''(x)$ exist only as generalized functions. Unfortunately, the convolution of two generalized functions can only be interpreted under very special circumstances.[12] In order to avoid this difficulty, we take the Fourier transform of $|x|\psi''(x)$ and make two successive integrations by parts,

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|x|\psi''(x)e^{-ipx}\,dx$$
$$=\frac{1}{2\pi}\left\{-p^2\int_{-\infty}^{\infty}|x|\psi(x)e^{-ipx}\,dx\right.$$
$$\left.-2ip\int_{-\infty}^{\infty}\operatorname{sgn}x\psi(x)e^{-ipx}\,dx+4\psi(0)\right\}, \quad (21)$$

and then replace the Fourier transform of each product by the corresponding convolution of transforms. These transforms are

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\psi(x)e^{-ipx}\,dx=a(p), \quad (22)$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|x|e^{-ipx}\,dx=-\frac{1}{\pi p^2}, \quad (23)$$

and

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\operatorname{sgn}xe^{-ipx}\,dx=\frac{1}{\pi ip}. \quad (24)$$

where $|x|$ and sgn x are treated as generalized functions.[10]

Doing the prescribed manipulations, we obtain

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|x|\psi''(x)e^{-ipx}\,dx$$
$$=\frac{\psi(0)}{\pi}+\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{p'^2a(p')dp'}{(p-p')^2}, \quad (25)$$

which when applied to (20) along with the transform of the remaining terms, gives the momentum representation

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{(p'^2+k^2)a(p')dp'}{(p-p')^2}+\frac{\psi(0)}{\pi}=-2\lambda a(p), \quad (26)$$

where $E=-k^2/2$. This integral equation can be cast into the form

$$\frac{k}{\pi}\int_{-\pi}^{\pi}\frac{\chi(\phi')d\phi'}{1-\cos(\phi-\phi')}+\frac{k^2}{\cos^2(\phi/2)}\frac{\psi(0)}{\pi}=-2\lambda\chi(\phi) \quad (27)$$

by changing to the variable ϕ defined by

$$p=k\tan(\phi/2), \quad (28)$$

along with the definition

$$\chi(\phi)=(p^2+k^2)a(p). \quad (29)$$

Our purpose in deriving Eq. (27) is to demonstrate the presence of the term containing $\psi(0)$ which is missing from the equation obtained by Davtyan *et al.* Thus the condition $\psi(0)=0$ is implicit in their work. We make no attempt to solve Eq. (27) for $\psi(0)\neq 0$ since the corresponding configuration space wavefunctions are presumably the even functions $\psi_2(x)$, which we have thoroughly discussed in Sec. II. In the following we set $\psi(0)=0$ to discuss the results of Davtyan *et al.*[6]

With $\psi(0)=0$, the normalized solutions of (27) are

$$\chi_n^{\pm}(\phi)=(2/\pi)^{1/2}(n)^{-3/2}e^{\pm in\phi}, \quad (30)$$

where n is a positive integer. These functions obviously exhibit an $O(2)$ symmetry since they are eigenfunctions of the operator $-i(\partial/\partial\phi)$. From the definitions in (28) and (29) the corresponding solution in momentum space is found to be

$$a_n^{\pm}(p)=\left(\frac{2}{\pi}\right)^{1/2}(n)^{-3/2}(p^2+k^2)^{-1}\left(\frac{p+ik}{p-ik}\right)^{\pm n}, \quad (31)$$

where $k=\lambda/n$. It is easily demonstrated that $a_n^+(p)$ corresponds to configuration wavefunctions zero for $x<0$ while those of $a_n^-(p)$ are zero for $x>0$. The degeneracy arises since there are two distinct wavefunctions for a given integer n. These two configuration wavefunctions are eigenfunctions of the operator sgn x with eigenvalues ± 1. We can demonstrate that this follows from the $O(2)$ symmetry mentioned earlier by writing the operator sgn x in terms of the variable ϕ. Thus

$$\operatorname{sgn}x=\frac{x}{|x|}=\frac{-i(\partial/\partial p)}{|i(\partial/\partial p)|}=\frac{-i(\partial/\partial\phi)}{|i(\partial/\partial\phi)|}. \quad (32)$$

The solutions $\chi_n(\phi)$, since they are eigenfunctions of $-i\partial/\partial\phi$, must give configuration wavefunctions that are eigenfunctions of sgn x.

We emphasize here that the $O(2)$ symmetry occurs only if the origin is excluded and the condition $\psi(0)=0$ is imposed. Since there can be no current across the origin if $\psi(0)=0$, the space is effectively divided into two half-spaces entirely separated from each other.[4,7] The regular solution can thus be extended to give either an odd or even wavefunction with the same value of k. On the other hand, if one includes the origin, the even extension of the regular solution does not satisfy the Schrödinger equation but rather Eq. (5), one is led to consider the irregular solutions as was done in Sec. II.

For the continuous spectrum $E>0$, similar results hold. For $\psi(0)=0$, Davtyan *et al.*[6] again uncover a hidden symmetry that, in configuration space, corresponds to symmetry under the operator sgn x. Their solutions are eigenfunctions of sgn x and just as for the case of the bound states, the space is divided into two disjointed half-spaces.

IV. CONCLUSION

Unlike earlier investigations, we have found the spectrum of the one-dimensional hydrogen atom to be nondegenerate. The fault in the earlier work is the acceptance of wavefunctions that are not solutions to the Schrödinger equation but rather to Eq. (5). Our work here may be criticized on the grounds that we have not used the Coulomb potential but have modified it with the addition of a δ-function term. We make the following remarks.

(1) The strong singularity of the Coulomb potential at the origin completely dominates the additional δ-function term.

(2) In contrast to other regularizations, the treatment of the Coulomb potential as a generalized function that

requires the additional δ-function term preserves the connection formula needed to match ψ and ψ' at $x=0$ in the limit as $\alpha \to 0$ [see Eqs. (14) and (16)]. Otherwise the point $x=0$ divides the space into two disjointed halves.
(3) The earlier results of a degenerate spectrum can be obtained from our results by taking the limit $q \to \infty$ in Eq. (18). Note that in this limit, the δ-function term is part of the potential.

ACKNOWLEDGMENT

This work was supported by the U. S. Department of Energy, Contract No. W-7405-Eng-82, Office of Energy Research (KA-01-01), Division of High Energy and Nuclear Physics.

[1] R. Loudon, Am. J. Phys. **27**, 649 (1959).
[2] L. K. Haines and D. H. Roberts, Am. J. Phys. **37**, 1145 (1969).
[3] M. Andrews, Am. J. Phys. **34**, 1194 (1966).
[4] M. Andrews, Am. J. Phys. **44**, 1064 (1976).
[5] H. N. Spectos and J. Lee, Am. J. Phys. **53**, 248 (1985).
[6] L. S. Davtyan, G. S. Pogosyan, A. N. Sissakian, and V. M. Ter-Antonyan, J. Phys. A: Math. Gen. **20**, 2765 (1987).
[7] C. L. Hammer and T. A. Weber, Am. J. Phys. **56**, 281 (1988).
[8] *Handbook of Mathematical Functions,* National Bureau of Standards, Appl. Math. Series 55, edited by M. Abramowitz and F. A. Stegun (U.S. Govt. Printing Office, Washington, DC, 1966), pp. 504–510.
[9] J. R. Klauder, Acta. Phys. Austr. Supp. **XI**, 341 (1973); J. R. Klauder, Phys. Lett. B **47**, 523 (1973); B. DeFacio and C. L. Hammer, J. Math. Phys. **15**, 1071 (1974).
[10] M. J. Lighthill, *An Introduction to Fourier Analysis and Generalized Functions* (Cambridge U. P., London, 1958), pp. 38–43.
[11] Note that Euler's constant was inadvertently left out of the derivative in Ref. 7.
[12] I. M. Gel'fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), p. 103.

Casimir force on a spherical shell when $\epsilon(\omega)\mu(\omega)=1$

Iver Brevik
Division of Applied Mechanics, University of Trondheim-NTH, N-7034 Trondheim-NTH, Norway
Roger Sollie[a]
Institute of Physics, University of Trondheim-NTH, N-7034 Trondheim-NTH, Norway

(Received 12 October 1989; accepted for publication 7 January 1990)

The Casimir surface force on a spherical shell is calculated, assuming the material to be satisfying the condition $\epsilon(\omega)\mu(\omega) = 1$, $\epsilon(\omega)$ being the spectral permittivity and $\mu(\omega)$ the spectral permeability. The basic formula for the force is given under general conditions, without any restrictive assumption on the thickness of the shell or on the specific dispersion relation. When it comes to numerical evaluations, it is assumed that the shell is of *small* thickness, and also that the simple form $\mu(\omega) = \mu_s\ (\omega \leqslant \omega_0)$, $\mu(\omega) = 1\ (\omega > \omega_0)$, for the dispersion relation. The special case when $\mu_s \to \infty$ or 0 is given particular attention, since this case appears to be of main physical interest and also since it implies mathematical simplifications. The force $\mathscr{F}$ may then be written as the sum of two terms: one "normal" term $\mathscr{F}^{(0)}$ containing an attractive dispersion-induced part as well as a repulsive, nondispersive finite part, and one "abnormal" term $\mathscr{F}^{(1)}$ that becomes divergent when summed over all angular momenta. This particular behavior of $\mathscr{F}^{(1)}$ is a consequence of the assumed small magnitude of the shell thickness. A similar analysis of the opposite extreme case of dilute media is also made, and analogous angular moment divergent results are found. The extraction of physically meaningful information from the divergent expressions is discussed. In general, numerical methods are necessary to handle the Riccati–Bessel functions, although in the special cases mentioned, useful analytic results are obtained using the Debye expansion. The numerical calculation of the Casimir force on shells of *finite* thickness is also commented upon, and in the Appendix the generalization of the theory to the case of finite temperatures is discussed.

I. INTRODUCTION

An important progress in the theory of the electromagnetic Casimir effect[1] in dielectric media in recent years is the realization of the importance of the *dispersive* effect. It was Candelas,[2] in particular, who stressed the need of taking this particular effect into account. The presence of dispersion implies, according to Candelas, that there is a strong, negative, cutoff-dependent contribution to the Casimir energy of a perfectly conducting spherical shell. Candelas' general arguments were in essence supported by the specific model calculation carried out in Ref. 3 for a dispersive, compact spherical ball. We obtained a strong, attractive, contribution to the Casimir *surface force*.

Also from a mathematical point of view, the inclusion of dispersion is welcome since under usual physical conditions (i.e., in the absence of singularities) one avoids the "infinity plus small remainder" expressions that made the extraction of physical results so difficult in earlier investigations. In particular, one no longer has to worry about the legitimacy of interchanging an infinite sum with an integral: In the conventional nondispersive calculations one is confronted with a frequency integral of an infinite sum over all angular momentum variables ℓ. The series, usually calculated by means of the Debye expansion, is an asymptotic high-ℓ expansion and thus not uniformly convergent. If one simply interchanges the sum with an integral one runs the risk of loosing a constant, which is infinite in the case of a nondispersive medium. Problems of this kind are avoided when the medium is taken to be dispersive from the outset. The compact sphere calculation of Ref. 3 is typical in this respect.

For reference purposes we mention that the first calculation of Boyer[4] on the Casimir force on a nondispersive perfectly conducting spherical shell of vanishing thickness gave a repulsive result. More refined calculations—Refs. 5, 6, and 7 for instance—similarly found the force to be repulsive. As already mentioned, the development took a new turn when Candelas[2] and others discovered the importance of dispersion in the present problem. In the model calculation of Ref. 3 at zero temperature, the permittivity $\epsilon(\omega)$ and the permeability $\mu(\omega)$ of the medium were assumed to satisfy the condition

$$\epsilon(\omega)\mu(\omega) = 1. \qquad (1.1)$$

Moreover, a one-absorption-frequency Sellmeir dispersion relation was adopted. In a recent paper[8] we have extended these considerations on a compact sphere to a case of finite temperatures.

The purpose of the present paper is to calculate the Casimir surface force on a spherical *shell*. The medium is still required to satisfy (1.1). When it comes to numerical evaluation of the force, we adopt as the dispersion relation the simplest imaginable form:

[a] Present address: Department of Physics, Temple University, Philadelphia, Pennsylvania 19122.

 0022-2488/90/061445-11$03.00

$$\mu(\omega) = \begin{cases} \mu_s, & \omega \leqslant \omega_0, \\ 1, & \omega > \omega_0, \end{cases} \tag{1.2}$$

where μ_s and ω_0 are constants. The corresponding permittivity follows from (1.1). Our calculational procedure is mainly the same as the one worked out in earlier papers[9,10] for the case of spherical shells. Another feature of our numerical calculations is that we consider in the main text only the special case of geometrically very *thin* shells. This feature brings an element of singularity into our calculations that in turn implies that terms occur in the force that diverge when summed up to $\ell = \infty$. On physical grounds we have to truncate the sum at a finite upper limit ℓ_0. The sum–integral interchange problem mentioned above is accordingly removed as the series contains only a finite number of terms.

In the following section we work out the general formalism for the Casimir force. Because of the complexity, numerical work is generally required for a complete evaluation. The important special case in which $\mu_s \to \infty$ or 0 considered in Sec. III is to a large extent amenable to an analytic treatment. We find in this case that for a very thin shell the force is the sum of two terms: First, there is a term $\mathscr{F}^{(0)}$ that is precisely of the form that we would expect for an ordinary perfectly conducting shell in electrodynamics, consisting of an attractive dispersion-induced part and a repulsive nondispersive part. Secondly there is a term $\mathscr{F}^{(1)}$ that diverges when summed over ℓ. This term, absent in the case of a compact sphere has to be a consequence of the geometrical singularity of the shell. The extraction of physically meaningful results is discussed. Section IV is considered with the case of dilute media. Also in this case, divergent terms are found in the force expression.

In order to elucidate the role played by the geometrical singularity, we consider in Sec. V the Casimir force on a shell whose width is not necessarily small. If the outer radius is infinite (at fixed inner radius), the force becomes easily calculable. Taking into account the Debye expansion, we can write the force expression with excellent accuracy in a very simple way. When the outer radius decreases, numerical calculations indicate that the force stays finite down to a surprisingly low value of the outer radius/inner radius ratio, viz. to about 1.1. Only when a shell is thinner than this, do we become confronted with the peculiarities of thin-shell theory.

In Appendix A we compile some information about the generation of Riccati–Bessel functions on a pc computer. Appendix B contains an analysis of how the essentials of the Casimir shell theory can be generalized to the case of finite temperatures.

In this paper, $\hbar$ and c are put equal to unity.

II. GENERAL FORMALISM FOR THE SPHERICAL SHELL

A. Basics

The geometry of the shell is sketched in Fig. 1. The inner radius is a, the outer is b, and the medium in between satisfies (1.1). In the two regions $r < a$, and $r > b$, we assume there to be a vacuum. When dealing with the general theory below, we assume the thickness $(b - a)$ of the shell to be arbitrary. To begin with, no explicit choice is made for $\mu(\omega)$.

If one makes use of the conventional point-splitting formalism in field theory, one can write the surface force density F (F_1 on the inner surface, F_2 on the outer surface) as a Fourier integral

$$F = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega\tau} F(\omega), \tag{2.1}$$

with $\tau = t - t'$ denoting the time splitting between the two space-time points x and x'. This is because the stationarity of the problem means that the force depends only on the time difference between the two points. In nondispersive theory, τ plays the role of a cutoff parameter, just as it does in conventional field theory. In the present dispersive theory, there is no need to keep this parameter in the formalism. We accordingly put $\tau = 0$ in the following.

Making use of Maxwell's stress tensor, we obtain for the Fourier component $F_1(\omega)$ of the surface force density F_1:

$$F_1(\omega) = \frac{\mu(\omega) - 1}{2} \Big[\langle E_r^2(a-) \rangle_\omega + \frac{1}{\mu(\omega)} \times \langle E_\perp^2(a-) \rangle_\omega - \frac{1}{\mu(\omega)} \langle H_r^2(a-) \rangle_\omega - \langle H_\perp^2(a-) \rangle_\omega \Big]. \tag{2.2}$$

Here $r = a-$ is the position just inside of the inner shell, and the subscripts r and $\perp$ refer to the radial and the orthogonal direction.

The expectation values of the products of field components are evaluated by means of Scwinger's source theory.[11] The electric field $\mathbf{E}(x)$ is related to the polarization $\mathbf{P}(x)$ through a dyad $\mathbf{\Gamma}(x,x')$:

$$\mathbf{E}(x) = \int dx' \; \mathbf{\Gamma}(x,x')\mathbf{P}(x'), \tag{2.3}$$

where the Fourier component of $\mathbf{\Gamma}$ satisfies the governing equation

$$-\text{curl curl } \mathbf{\Gamma}(\mathbf{r},\mathbf{r}',\omega) + \omega^2 \mathbf{\Gamma}(\mathbf{r},\mathbf{r}',\omega) = -\mu(\omega)\omega^2 \mathbf{1}\delta(\mathbf{r} - \mathbf{r}'). \tag{2.4}$$

The effective product of two electric field components is

$$i\langle E_i(\mathbf{r}) E_k(\mathbf{r}') \rangle_\omega = \Gamma_{ik}(\mathbf{r},\mathbf{r}',\omega). \tag{2.5}$$

The solution of (2.4) contains two scalar Green's functions F_ℓ and G_ℓ, which can be expressed in terms of spherical Bessel and Hankel functions. We write down the expressions for the effective products of the electric field components at $r = r' = a-$:

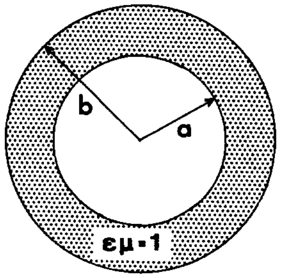


FIG. 1. Geometry of the spherical shell.

$$i\langle E_r^2(a-)\rangle_\omega = \frac{1}{a^2}\sum_{\ell=1}^{\infty}\frac{2\ell+1}{4\pi}\ell(\ell+1)G_\ell(a-,a-),$$

$$i\langle E_\perp^2(a-)\rangle_\omega = \sum_{\ell=1}^{\infty}\frac{2\ell+1}{4\pi}\left[\omega^2F_\ell+\frac{1}{a^2}\frac{\partial}{\partial r}r\frac{\partial}{\partial r'}r'G_\ell\right]_{r=r'=a-}. \quad (2.6)$$

Note that since the fields are evaluated in the vacuum region on the inside of the shell, the material permittivity or permeability do not occur in these expressions. The analogous magnetic field products $i\langle H_r^2(a-)\rangle_\omega$ and $i\langle H_\perp^2(a-)\rangle_\omega$ are obtained from (2.6) upon the substitutions $G_\ell \leftrightarrow F_\ell$. Inserting the four effective field products in (2.2), we obtain the force density on the inner surface expressed in terms of F_ℓ and G_ℓ.

Analogous considerations apply to the calculation of the force density $F_2(\omega)$ on the outer surface. It is now convenient to exploit the electromagnetic boundary conditions across the surface $r=b$ so as to permit the force to be expressed in terms of the vacuum fields on the *outside*:

$$F_2(\omega) = \frac{\mu(\omega)-1}{2}\left[-\langle E_r^2(b+)\rangle_\omega - \frac{1}{\mu(\omega)}\times\langle E_\perp^2(b+)\rangle_\omega + \frac{1}{\mu(\omega)}\langle H_r^2(b+)\rangle_\omega + \langle H_\perp^2(b+)\rangle_\omega\right]. \quad (2.7)$$

We write down the effective products for the electric field components on the outside:

$$i\langle E_r^2(b+)\rangle_\omega = \frac{1}{b^2}\sum_{\ell=1}^{\infty}\frac{2\ell+1}{4\pi}\ell(\ell+1)G_\ell(b+,b+),$$

$$i\langle E_\perp^2(b+)\rangle_\omega = \sum_{\ell=1}^{\infty}\frac{2\ell+1}{4\pi}\left[\omega^2F_\ell+\frac{1}{b^2}\frac{\partial}{\partial r}r\frac{\partial}{\partial r'}r'G_\ell\right]_{r=r'=b+}. \quad (2.8)$$

The magnetic field products are analogous. These expressions are to be inserted in Eq. (2.7).

We shall be interested in the total surface force $\mathscr{F}$ on the shell. It is defined as

$$\mathscr{F} = 4\pi(a^2F_1+b^2F_2). \quad (2.9)$$

Its Fourier component $\mathscr{F}(\omega)$ can in view of the above expressions be written

$$\mathscr{F}(\omega) = \frac{\mu(\omega)-1}{2i}\sum_{\ell=1}^{\infty}(2\ell+1)\Bigg\{\left([\ell(\ell+1)-\omega^2a^2]\left[G_\ell-\frac{F_\ell}{\mu(\omega)}\right]+\frac{\partial}{\partial r}r\frac{\partial}{\partial r'}r'\left[\frac{G_\ell}{\mu(\omega)}-F_\ell\right]\right)_{r=r'=a-}$$
$$-\left([\ell(\ell+1)-\omega^2b^2]\left[G_\ell-\frac{F_\ell}{\mu(\omega)}\right]+\frac{\partial}{\partial r}r\frac{\partial}{\partial r'}r'\left[\frac{G_\ell}{\mu(\omega)}-F_\ell\right]\right)_{r=r'=b+}\Bigg\}. \quad (2.10)$$

From this equation the advantage of expressing $F_1(\omega)$ and $F_2(\omega)$ in terms of the effective products on the inside, respectively on the outside of the shell is apparent: We need only to use explicitly the scalar Green's functions in the two *vacuum* regions,

$r,r'<a$,

$$F_\ell,G_\ell = ikj_\ell(kr_<)[h_\ell^{(1)}(kr_>)-A_{F,G}j_\ell(kr_>)],$$

$r,r'>b$, (2.11)

$$F_\ell,G_\ell = ik[j_\ell(kr_<)-B_{F,G}h_\ell^{(1)}(kr_<)]h_\ell^{(1)}(kr_>),$$

with $k=|\omega|$. The constants $A_{F,G}$ and $B_{F,G}$ have to be determined by the electromagnetic boundary conditions across the surfaces $r=a$ and $r=b$. This procedure makes it necessary to invoke the Green's function in the intermediate region $a<r<b$ also. We abstain from going into detail here—the reader is referred to an earlier paper[9]—and we confine ourselves to writing down the expressions for those terms that are needed in (2.10):

$$[\ell(\ell+1)-\omega^2a^2]\left[G_\ell-\frac{F_\ell}{\mu(\omega)}\right](a-,a-) = ks_\ell''(1)\left\{\left[\frac{1}{N}-\frac{\mu(\omega)}{\tilde N}\right]e_\ell(2)Q_\ell'(2) + \left[\frac{1}{\tilde N}-\frac{\mu(\omega)}{N}\right]e_\ell'(2)Q_\ell\right\}, \quad (2.12)$$

$$\frac{\partial}{\partial r}r\frac{\partial}{\partial r'}r'\left[\frac{G_\ell}{\mu(\omega)}-F_\ell\right]_{r=r'=a-} = ks_\ell'(1)\left\{\left[\frac{1}{N}-\frac{\mu(\omega)}{\tilde N}\right]e_\ell(2)Q_\ell''(12) + \left[\frac{1}{\tilde N}-\frac{\mu(\omega)}{N}\right]e_\ell'(2)Q_\ell'(1)\right\}, \quad (2.13)$$

$$[\ell(\ell+1)-\omega^2b^2]\left[G_\ell-\frac{F_\ell}{\mu(\omega)}\right](b+,b+) = -ke_\ell''(2)\left\{\left[\frac{1}{N}-\frac{\mu(\omega)}{\tilde N}\right]s_\ell(1)Q_\ell'(1) + \left[\frac{1}{\tilde N}-\frac{\mu(\omega)}{N}\right]s_\ell'(1)Q_\ell\right\}, \quad (2.14)$$

$$\frac{\partial}{\partial r}r\frac{\partial}{\partial r'}r'\left[\frac{G_\ell}{\mu(\omega)}-F_\ell\right]_{r=r'=b+} = -ke_\ell'(2)\left\{\left[\frac{1}{N}-\frac{\mu(\omega)}{\tilde N}\right]s_\ell(1)Q_\ell''(12) + \left[\frac{1}{\tilde N}-\frac{\mu(\omega)}{N}\right]s_\ell'(1)Q_\ell'(2)\right\}. \quad (2.15)$$

In these expressions $s_\ell(z)=zj_\ell(z)$ and $e_\ell(z)=zh_\ell^{(1)}(z)$ are the Riccati–Bessel functions; for notational convenience $s_\ell(1)\equiv s_\ell(ka)$, $s_\ell(2)\equiv s_\ell(kb)$, etc. Prime means differentiation with respect to the whole argument. The symbol Q and its various derivatives are defined as

$$Q_\ell = s_\ell(1)e_\ell(2)-e_\ell(1)s_\ell(2),$$
$$Q_\ell'(1) = s_\ell'(1)e_\ell(2)-e_\ell'(1)s_\ell(2), \quad (2.16)$$
$$Q_\ell'(2) = s_\ell(1)e_\ell'(2)-e_\ell(1)s_\ell'(2),$$

$$Q''_\ell(12) = s'_\ell(1)e'_\ell(2) - e'_\ell(1)s'_\ell(2).$$

Finally, the symbols N and $\tilde{N}$ are defined as

$$N = e_\ell(2)[s_\ell(1)Q''_\ell(12) - \mu(\omega)s'_\ell(1)Q'_\ell(2)] - \mu(\omega)e'_\ell(2)[s_\ell(1)Q'_\ell(1) - \mu(\omega)s'_\ell(1)Q_\ell],$$
$$\tilde{N} = e'_\ell(2)[s'_\ell(1)Q_\ell - \mu(\omega)s_\ell(1)Q'_\ell(1)] - \mu(\omega)e_\ell(2)[s'_\ell(1)Q'_\ell(2) - \mu(\omega)s_\ell(1)Q''_\ell(12)]. \quad (2.17)$$

When deriving these expressions we made use of the basic differential equation satisfied by the Riccati–Bessel functions.[12]

B. Frequency rotation. General expression for the force

Inserting the expressions (2.12)–(2.15) in (2.10) we can calculate the Fourier component of the force. The physical force on the shell is in accordance with (2.1) equal to

$$\mathscr{F} = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \mathscr{F}(\omega), \quad (2.18)$$

when the cutoff parameter is equal to zero. As in the nondispersive case, for a medium satisfying the condition $\epsilon\mu = 1$, there is no need of taking into account contact terms.[13]

The frequency integral in (2.18), which implies the Feynman path of integration, can in view of symmetry about the origin be replaced by twice the integral from zero to infinity. We perform a complex frequency rotation, $\omega \to i\hat{\omega}$, and integrate along the imaginary frequency axis. Since only positive frequencies are now involved, the frequency rotation implies

$$k \to i\hat{k} = i\hat{\omega}, \quad (2.19)$$

where the last equality holds because the refractive index of the medium is equal to unity; cf. (1.1). It is convenient to define nondimensional frequencies x and y:

$$x = \hat{k}a, \quad y = \hat{k}b \quad (2.20)$$

(they were called x_1 and x_2 in Refs. 9 and 10). The quantities defined in (2.16) and (2.17) now become functions of ix and iy. As in Ref. 3 we retain the symbols unchanged, implying that $s_\ell(ix) \to s_\ell(x)$, etc. The Riccati–Bessel functions are, with $\nu = \ell + 1/2$,

$$s_\ell(x) = (\pi x/2)^{1/2} I_\nu(x), \qquad e_\ell(x) = (2x/\pi)^{1/2} K_\nu(x), \quad (2.21)$$

corresponding to the Wronskian $W\{s_\ell, e_\ell\} = -1$. The quantities Q_ℓ and N are transformed similarly; for definiteness we write down here the new versions of the two first of Eqs. (2.16) and the first of Eqs. (2.17):

$$Q_\ell = s_\ell(x)e_\ell(y) - e_\ell(x)s_\ell(y),$$
$$Q'_\ell(x) = s'_\ell(x)e_\ell(y) - e'_\ell(x)s_\ell(y), \quad (2.22)$$
$$N = e_\ell(y)[s_\ell(x)Q''_\ell(x,y) - \mu(x)s'_\ell(x)Q'_\ell(y)] - \mu(x)e'_\ell(y)[s_\ell(x)Q'_\ell(x) - \mu(x)s'_\ell(x)Q_\ell].$$

Introducing the magnetic susceptibility

$$\chi(x) = \mu(x) - 1, \quad (2.23)$$

we arrive after some calculation at the following general expression for the force:

$$\mathscr{F} = \frac{1}{2\pi a^2} \int_0^\infty dx\, x\chi(x) \sum_{\ell=1}^{\infty} (2\ell+1) \left\{ \left[\frac{1}{N} - \frac{\mu(x)}{\tilde{N}} \right] [(s_\ell(x)e'_\ell(y) + s'_\ell(x)e_\ell(y))Q''_\ell(x,y) + s_\ell(x)e''_\ell(y)Q'_\ell(x) + s''_\ell(x)e_\ell(y)Q'_\ell(y)] + \left[\frac{1}{\tilde{N}} - \frac{\mu(x)}{N} \right] [(s'_\ell(x)e''_\ell(y) + s''_\ell(x)e'_\ell(y))Q_\ell + s'_\ell(x)e'_\ell(y)(Q'_\ell(x) + Q'_\ell(y))] \right\}. \quad (2.24)$$

The force expression written in this way is convenient for further numerical processing. Assume, for instance, that both functions s_ℓ and e_ℓ and their first derivatives are accessible from a computer library. Then (2.24) is in principle directly computable, for a given dispersion relation $\chi(x)$, when in addition one takes into account that the second derivatives are calculable from the basic differential equation for the Riccati–Bessel functions in the form (2.21):

$$\begin{Bmatrix} s''_\ell(x) \\ e''_\ell(x) \end{Bmatrix} = \left[1 + \frac{\ell(\ell+1)}{x^2} \right] \begin{Bmatrix} s_\ell(x) \\ e_\ell(x) \end{Bmatrix}. \quad (2.25)$$

For analytic purposes it is usually more convenient to rewrite $\mathscr{F}$ in a more compact form by introducing the operator L:

$$L = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}, \quad (2.26)$$

x and y being considered as independent variables in a function on which L acts. We can then write

$$\mathscr{F} = \frac{1}{2\pi a^2} \int_0^\infty dx\, x\chi(x) \sum_{\ell=1}^{\infty} (2\ell+1) \times \left\{ \left[\frac{1}{N} - \frac{\mu(x)}{\tilde{N}} \right] L\, [s_\ell(x)e_\ell(y)Q''_\ell(x,y)] + \left[\frac{1}{\tilde{N}} - \frac{\mu(x)}{N} \right] L\, [s'_\ell(x)e'_\ell(y)Q_\ell] \right\}. \quad (2.27)$$

This formula gives the total surface force as defined in (2.9) on the shell. Let us summarize here the basic assumptions on which (2.27) rests: The formula holds at zero temperature for a medium satisfying condition (1.1). No explicit choice of the dispersion relation $\chi(x)$ is made so far. Nor is there at this stage any restriction on the thickness $(b-a)$ of the shell.

For the remainder of this paper we shall chiefly be concerned with numerical calculations under certain restrictive conditions implying, in the first place, adoption of the dispersion relation (1.2). It corresponds to

$$\chi(x) = \begin{cases} \chi_s, & x \leqslant x_0, \\ 0, & x > x_0, \end{cases} \quad (2.28)$$

with $\chi_s = \mu_s - 1$. Secondly, we specialize to the case of a geometrically *thin* shell. Thus

$$b = a(1+\xi), \quad \text{with } \xi \ll 1. \tag{2.29}$$

In the following section we consider the case where μ_s is either a very large or a very small quantity.

III. EXTREME PERMEABILITIES

We examine the case where μ_s satisfies one of the following two conditions:

$$\mu_s \to \begin{cases} \infty, \\ 0. \end{cases} \tag{3.1}$$

These two possibilities can be treated analytically on the same footing; they will lead to the same expression for the force. Note that in view of the condition (1.1) a very large value of μ_s corresponds to a very small value of ϵ_s ($= 1/\mu_s$).

For convenience we may start from the general force expression in the compact form (2.27). It is apparent that the two terms having the form $L[\cdots]$ are independent of $\mu(x)$. The permeability turns up only in these terms' prefactors:

$$\chi(x)\left[\frac{1}{N} - \frac{\mu(x)}{\tilde{N}}\right] \to \frac{-1}{s_\ell(x)e_\ell(y)Q''_\ell(x,y)},$$
$$\chi(x)\left[\frac{1}{\tilde{N}} - \frac{\mu(x)}{N}\right] \to \frac{-1}{s'_\ell(x)e'_\ell(y)Q_\ell}, \tag{3.2}$$

valid whenever one of the conditions (3.1) is satisfied. Taking into account the dispersion relation we write the total force $\mathscr{F}$ as a sum of two terms:

$$\mathscr{F} = \mathscr{F}^{(0)} + \mathscr{F}^{(1)}, \tag{3.3}$$

where the first term can be expressed in the following two alternative ways:

$$\mathscr{F}^{(0)} = \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty} (2\ell+1) \times \left[\frac{s'_\ell(x)}{s_\ell(x)} + \frac{s''_\ell(x)}{s'_\ell(x)} + \frac{e'_\ell(y)}{e_\ell(y)} + \frac{e''_\ell(y)}{e'_\ell(y)}\right]$$
$$= \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty} (2\ell+1) L \times \ln[-s_\ell(x)s'_\ell(x)e_\ell(y)e'_\ell(y)]. \tag{3.4}$$

The second term can analogously be expressed as

$$\mathscr{F}^{(1)} = \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty} (2\ell+1)\left[\frac{Q'_\ell(x)}{Q_\ell} + \frac{Q'_\ell(y)}{Q_\ell} + \frac{s''_\ell(x)}{s_\ell(x)}\frac{Q'_\ell(y)}{Q''_\ell(x,y)} + \frac{e''_\ell(y)}{e_\ell(y)}\frac{Q'_\ell(x)}{Q''_\ell(x,y)}\right]$$
$$= \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty} (2\ell+1) L \ln[-Q_\ell Q''_\ell(x,y)]. \tag{3.5}$$

It is seen that the permeability μ_s has dropped out explicitly from these expressions. Note that the logarithmic arguments in (3.4) and (3.5) are both positive. The latter expressions are in agreement with the nondispersive expressions given in Eq. (10) of Ref. 10 if the cutoff parameter in that paper is put equal to zero and the frequency integration terminated at x_0.

We now consider the two terms in the force separately.

A. The term $\mathscr{F}^{(0)}$

This term is, as we will see, the "normal" term since it gives results that are in accordance with what we would expect in the electrodynamic theory of an ordinary dielectric medium. Also, no divergences are encountered. The term is at once seen to possess particularly simple properties in the limiting case when the thickness of the shell shrinks to zero: there is no presence of Q_ℓ or any of its derivatives in (3.4). As is seen from the definition Eqs. (2.22), the Q_ℓ are thickness quantities which go to zero when $y \to x$. This means that $\mathscr{F}^{(0)}$ does not require any expansion in the thickness parameter, and we can simply replace y by x in (3.4). This property makes it relatively easy to calculate $\mathscr{F}^{(0)}$ analytically, making use of the Debye expansion of the Riccati–Bessel functions.[12] For our purpose it is convenient to quote from Ref. 14 the following expansion for the product that is needed in (3.4):

$$-s_\ell(x)s'_\ell(x)e_\ell(x)e'_\ell(x) = \tfrac{1}{4}[1 - t^6/4\nu^2 + O(\nu^{-4})], \tag{3.6}$$

where

$$t(z) = (1+z^2)^{-1/2}, \quad z = x/\nu. \tag{3.7}$$

Expanding the logarithm of expression (3.6) to the lowest order, and using that $L(t) = -zt^3/\nu$ we obtain, writing the force as $\mathscr{F}^{(0)}(a,a)$ for clarity,

$$\mathscr{F}^{(0)}(a,a) = \frac{-3}{2\pi a^2}\int_0^{x_0} dx\, x^2 \sum_{\ell=1}^{\infty} \frac{\nu^5}{(\nu^2+x^2)^4}. \tag{3.8}$$

The sum over ℓ can be calculated using the Euler–Maclaurin formula[12] in the form

$$\sum_{\ell=1}^{\infty} \frac{\nu^5}{(\nu^2+x^2)^4} = \frac{\frac{1}{6}x^4}{(w^2+x^2)^3} + \frac{\frac{1}{2}w^2}{(w^2+x^2)^2} + \sum_{\ell=1}^{4} \frac{\nu^5}{(\nu^2+x^2)^4} + \frac{\frac{7}{3}w^4}{(w^2+x^2)^4} + \frac{\frac{2}{3}w^6}{(w^2+x^2)^5}, \quad w = 11/2. \tag{3.9}$$

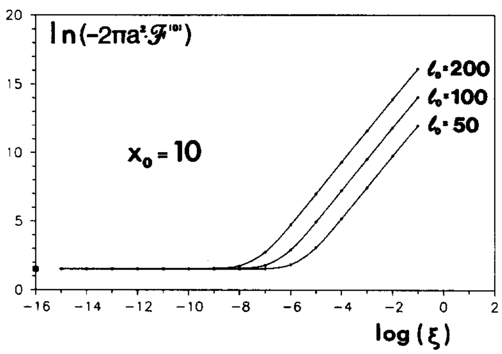


FIG. 2. Force term $\mathscr{F}^{(0)}$ for an extreme-permeability shell, as calculated from the basic Eq. (3.4). Here, ℓ_0 denotes the upper limit in the ℓ sum and ξ is the thickness parameter. The square point on the left ordinate axis is calculated from (3.12). Here, log denotes the logarithm with base 10 whereas ln is the natural logarithm.

As discussed in connection with Eq. (3.3b) of Ref. 8, the form (3.9) is quite accurate even for low values of x. Inserting (3.9) in (3.8) we obtain

$$\mathscr{F}^{(0)}(a,a) = \frac{-3}{2\pi a^2}\left\{\int_0^{x_0/w} dz\left[\frac{w}{6}\frac{z^6}{(1+z^2)^3} + \frac{w}{2}\frac{z^2}{(1+z^2)^2} + \frac{7}{3w}\frac{z^2}{(1+z^2)^4} + \frac{2}{3w}\frac{z^2}{(1+z^2)^5}\right] + \sum_{\ell=1}^{4}\int_0^{x_0/\nu}\frac{dz\, z^2}{(1+z^2)^4}\right\}. \tag{3.10}$$

It is here of interest to examine the limiting case $x_0 \to \infty$. The first term in (3.10) then diverges; the other terms remain finite. To show the structure of the divergent term explicit, we first rewrite it as

$$\frac{w}{6}\int_0^{x_0/w}\frac{dz\, z^6}{(1+z^2)^3} = \frac{x_0}{6} + \frac{x_0}{48}\frac{7+9x_0^2/w^2}{(1+x_0^2/w^2)^2} - \frac{55}{32}\arctan\left(\frac{x_0}{w}\right), \tag{3.11}$$

valid for arbitrary x_0 (cf. for instance, formulas 2.213 in Ref. 15). The divergence occurring when $x_0 \to \infty$ is thus seen to be *linear*. The remaining terms in (3.10) simplify to beta functions, and we obtain altogether:

$$\mathscr{F}^{(0)}(a,a) = -\frac{x_0}{4\pi a^2} + \frac{3}{64a^2}, \quad \text{for large } x_0. \tag{3.12}$$

This simple expression makes the main structure of the particular force term $\mathscr{F}^{(0)}(a,a)$ explicit: There is an attractive part in the force, being due to dispersion. Although the mathematical condition for the simple proportionality spelled out in the first term in (3.12) is, strictly speaking, that x_0 is a large quantity, it turns out numerically that the formula is surprisingly accurate even when x_0 is not very much larger than unity. In fact we have made a direct numerical calculation of the basic expression (3.4) (first version) for $\mathscr{F}^{(0)}$, with $y = x$ inserted, without invoking the Debye expansion at all. (How the Riccati–Bessel functions can be generated on a pc computer is discussed in Appendix A.) It turned out that even for a value of x_0 as low as 1.4, the error in the formula (3.12) amounted to less than 1%. For $x_0 = 1$, the error was 10%, and for $x_0 = 2$ it was about 0.1%. This is a striking demonstration of the usefulness of the analytic formula (3.12). (The physical importance of the term $\mathscr{F}^{(0)}(a,a)$ is larger than one might expect at the present stage; we will return to this point in connection with the discussion on thick shells in Sec. V.) The attractiveness of the dispersion—induced part of the force—and also the proportionality with x_0 inferred at large or moderate x_0—are properties that are in accordance with those found in Ref. 3 in the case of a *compact* spherical ball.

The second finite term in (3.12) is a repulsive term. It is interesting to note that this term, which is not related to dispersion, is just the term that was found in the earlier calculations based upon a *nondispersive* material model from the outset. Adopting such a model, the dispersive term was simply missed. We ought to mention that the term $3/64a^2 = 0.09375/2a^2$ is the result of an approximate calculation; we have made use of the lowest-order expansions of the Riccati–Bessel functions only. The most accurate calculation of the dispersion nonrelated term for a singular shell was given by the calculation of Milton, DeRaad, and Schwinger[7] (MDS):

$$\mathscr{F}_{\rm MDS} = 0.09235/2a^2. \tag{3.13}$$

The agreement with the second term in (3.12) is thus quite good.

How "thin" must a shell be before the basic expression (3.4) for $\mathscr{F}^{(0)}$ reduces the expression $\mathscr{F}^{(0)}(a,a)$ characterizing a singular shell? To investigate this point, we have calculated (3.4) numerically for various input values ξ (not using the Debye expansion). Figure 2 shows the calculated results for the case $x_0 = 10$, which is a typical value for the frequency cutoff. It is seen that it is in fact necessary to go to quite low values of ξ before one with sufficient accuracy is within the "singular shell" region, corresponding to $y = x$. It is necessary that the calculated results are independent with respect to variations in ξ. The figure shows that under the conditions chosen we are on safe ground when

$$\xi \lesssim 10^{-10}.$$

This estimate depends slightly on x_0. It is only when ξ reaches these extremely small values that we can replace $\mathscr{F}^{(0)}$ with $\mathscr{F}^{(0)}(a,a)$. Without computer assistance, this required smallness of ξ would not have been so easy to recognize.

There are two other useful observations to be made from this figure. First, the result corresponding to $x_0 = 10$ is in excellent agreement with the analytic high-x_0 approximation given in (3.12). [The point, marked with a square, on the left coordinate axis is calculated from (3.12).] Secondly, the figure indicates how far it is necessary to extend the ℓ summation in (3.4) in order to represent "infinity" with reasonable accuracy. It appears that an upper limit of

$$\ell_0 \sim 5x_0,$$

is sufficient for this purpose. The same conclusion is obtained more clearly from an inspection of Fig. 3, which shows a direct calculation of the force term $\mathscr{F}^{(0)}(a,a)$ for a singular shell, for various values of ℓ_0. [The two curves are terminated at the limits of the computer capacity, whereas

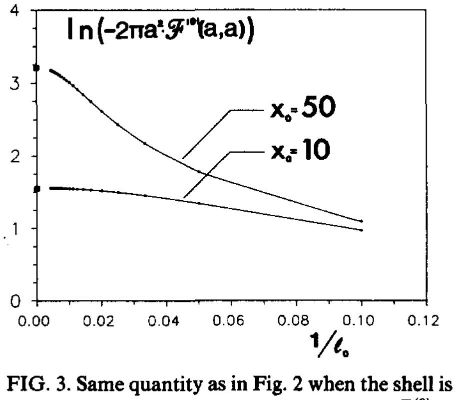


FIG. 3. Same quantity as in Fig. 2 when the shell is geometrically singular, i.e., $y = x$ or $\xi = 0$. This case is referred to as $\mathscr{F}^{(0)}(a,a)$. Square points to the left are calculated from (3.12).

the square points to the left are calculated from (3.12).] Again, $\ell_0 \sim 5x_0$, appears to be adequate as an upper limit.

B. The term $\mathscr{F}^{(1)}$

The structure of this term is different from that of $\mathscr{F}^{(0)}$ since Q_ℓ as well as its derivatives approach zero when y approaches x. We therefore have to perform a limiting procedure in the expression (3.5) for $\mathscr{F}^{(1)}$, using ξ as a smallness parameter. To this end we may again use the Debye expansion. From Eq. (27) in Ref. 10 we quote the expansion

$$\ln[-Q_\ell Q''_\ell(x,y)] = 2\ln(\nu\xi/t) - (t^2/4\nu^2) + O(\xi) + O(\nu^{-3}), \quad (3.14)$$

whereby after application of the operator L

$$L\ln[-Q_\ell Q''_\ell(x,y)] = -\frac{2t^2}{\nu z} + \frac{zt^4}{2\nu^3} + O\left(\frac{\xi}{\nu}\right) + O(\nu^{-4}). \quad (3.15)$$

Insertion in (3.5) yields, upon neglect of the higher-order terms,

$$\mathscr{F}^{(1)} = \frac{1}{\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty}\left[\frac{2\nu^3}{x(\nu^2+x^2)} - \frac{\nu x}{2(\nu^2+x^2)^2}\right]. \quad (3.16)$$

This expression, in which ξ is nowhere present, possesses the remarkable property that the sum over ℓ is diverging. This kind of behavior contrasts that found earlier for a compact spherical ball made of the same kind of material;[3] in that case all sums over ℓ were found to be finite. The divergence of (3.16) therefore has to be related to the *thin shell* geometry. However, in spite of this behavior it is still possible to use (3.16) to make an estimate of the force term $\mathscr{F}^{(1)}$. The important point here is that on physical grounds the large contribution from the higher values of ℓ has to be a spurious effect. Because of the dispersion relation, the presence of the medium cannot be felt for photons having frequencies above ω_0. If a photon of limiting frequency ω_0 just touches the surface of the sphere, its angular momentum is equal to $\omega_0 a$, i.e., to x_0. If the photon impact parameter is much larger than a, we do not expect it to be physically significant. (Arguments of the same kind were given also by Candelas.[2]) Thus, we may estimate the magnitude of the physical force component $\mathscr{F}^{(1)}$ by truncating the sum in (3.16) at $\ell_0 = fx_0$, where f is a factor. In analogy with what we found in the previous subsection, we expect that $f \sim 5$ is a sufficient value for most practical purposes.

Integrating over x in (3.16) we obtain finally

$$\mathscr{F}^{(1)} = \frac{2}{\pi a^2}\sum_{\ell=1}^{\ell_0}\left[\left(\nu^2 - \frac{1}{8}\right)\arctan\left(\frac{x_0}{\nu}\right) + \frac{x_0\nu}{8(\nu^2+x_0^2)}\right]. \quad (3.17)$$

When x_0 and ℓ_0 are given, this sum is easily evaluated numerically. The curve (b) in Fig. 4 shows how $\mathscr{F}^{(1)}$, calculated from (3.17), varies with ℓ_0 if $x_0 = 10$. The curve (a) shows for comparison the same result as calculated numerically from the first line in Eq. (3.5). It is seen that if we choose $f = 5$, which in the present case corresponds to $1/\ell_0 = 0.02$, the agreement between the two curves (a) and

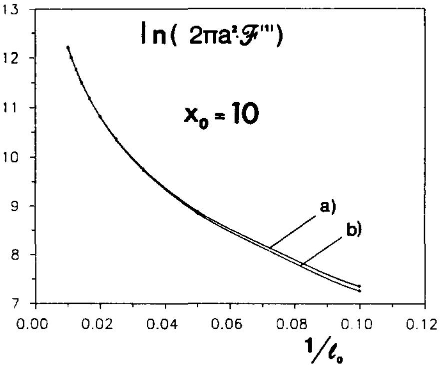


FIG. 4. Repulsive term $\mathscr{F}^{(1)}$ versus upper limit ℓ_0. Curve (a) is calculated numerically from Eq. (3.5); curve (b) follows from the semianalytic approximation (3.17).

(b) is very good, so that the formula (3.17) is adequate. The force term $\mathscr{F}^{(1)}$ is strong, and *repulsive*. For $x_0 = 10$ and $f = 5$, it is seen from Figs. 4 and 3 that $\mathscr{F}^{(1)}$ dominates completely over $\mathscr{F}^{(0)}(a,a)$.

As noted above, ξ has dropped out from the analytic approximation (3.16). This is a consequence of the fact that ξ has been assumed small in the derivation of this expression. One may wonder how small it is necessary to make ξ in the basic expression (3.5) for $\mathscr{F}^{(1)}$ before the singular shell regime is attained, i.e., before the expression becomes insensitive with respect to variations in ξ. Numerical trials indicate that $\xi \sim 10^{-3}$ is sufficient. In this sense the term $\mathscr{F}^{(1)}$ is seen to be not so delicate at small values of ξ as is the term $\mathscr{F}^{(0)}$.

Finally, it is of interest to compare the above result for $\mathscr{F}^{(1)}$ with that obtained on the basis of *nondispersive* theory. In the latter case we quote from Eq. (30) of Ref. 10:

$$\mathscr{F}^{(1)} = \frac{1}{a^2}\left(-\frac{1}{8} + \frac{2}{\delta^3}\right) \quad \text{(nondispersive theory)}, \quad (3.18)$$

where $\delta \to 0^+$ is the time-splitting cutoff parameter. We can recover the finite part of this expression from our dispersive result (3.17) above, if we first simply put $x_0 = \infty$ (for a finite value of ℓ_0) and thereafter put $\ell_0 = \infty$. This procedure leads to the sum

$$\mathscr{F}^{(1)} = \frac{1}{a^2}\sum_{\ell=1}^{\infty}\left(\nu^2 - \frac{1}{8}\right), \quad (3.19)$$

which can be further processed in a very simple way by making use of the standard analytic continuation of Riemann's zeta function:

$$\sum_{\ell=0}^{\infty}\nu^s = (2^{-s} - 1)\zeta(-s). \quad (3.20)$$

This expression implies the following substitutions:

$$\sum_{\ell=1}^{\infty}\nu^0 = -1, \quad \sum_{\ell=1}^{\infty}\nu^2 = -\frac{1}{4}, \quad (3.21)$$

which, upon insertion in (3.19), yield the answer

$$\mathscr{F}^{(1)} = -1/8a^2. \quad (3.22)$$

The comparison between the dispersive and the nondispersive theories is thus seen to give a satisfactory result: If we eliminate the influence from x_0 in the dispersive theory by putting $x_0=\infty$, we recover the expression (3.22) that is identical with the finite part of the nondispersive expression (3.18). Moreover, the presence of the cutoff parameter in the second term in (3.18) reflects in an indirect way the cutoff dependence of the result that is shown in an explicit way in our basic expression (3.17) for $\mathscr{F}^{(1)}$.

IV. DILUTE MEDIA

This case is characterized by a small susceptibility

$$\chi_s \ll 1. \tag{4.1}$$

We go back to the general force expression (2.27). It is useful to note the general relationship

$$N-\tilde{N}=[\mu^2(x)-1][s_\ell'(x)e_\ell(x)-s_\ell'(y)e_\ell(y)], \tag{4.2}$$

cf. (2.22), and also that

$$N=\tilde{N}=1, \quad \text{when } \mu(x)=1. \tag{4.3}$$

Expanding in $\chi(x)$ as a smallness parameter, we obtain using Eqs. (4.2) and (4.3)

$$\begin{aligned}1/N-\mu(x)/\tilde{N} &= -\chi(x)[1+2s_\ell'(x)e_\ell(x)-2s_\ell'(y)e_\ell(y)+O(\chi)],\\ 1/\tilde{N}-\mu(x)/N &= -\chi(x)[1-2s_\ell'(x)e_\ell(x)+2s_\ell'(y)e_\ell(y)+O(\chi)].\end{aligned} \tag{4.4}$$

Comparing with (2.27), it is already at this stage apparent that the force varies quadratically with the susceptibility χ. This feature is characteristic for the case of dilute media.

The smallness of the thickness parameter ξ has so far not been invoked. Turn now to an expansion in ξ. To this end it is convenient to use again the Debye expansion. Since[10]

$$s_\ell' e_\ell = \tfrac{1}{2}[1+t^3/2\nu+O(\nu^{-3})], \tag{4.5}$$

we obtain

$$\begin{aligned}\frac{1}{N}-\frac{\mu(x)}{\tilde{N}} &= -\chi(x)\left\{1+\frac{3\xi z^2t^5}{2\nu}[1+O(\xi)+O(\nu^{-2})]+O(\chi)\right\},\\ \frac{1}{\tilde{N}}-\frac{\mu(x)}{N} &= -\chi(x)\left\{1-\frac{3\xi z^2t^5}{2\nu}[1+O(\xi)+O(\nu^{-2})]+O(\chi)\right\}.\end{aligned} \tag{4.6}$$

There are thus three expansion parameters in this problem: shell thickness ξ, susceptibility χ, and Debye parameter $1/\nu$. To the lowest order, both expressions in (4.6) reduce to $-\chi(x)$. Then (2.27) yields

$$\mathscr{F}=\frac{-\chi_s^2}{2\pi a^2}\int_0^{x_0}dx\,x\sum_{\ell=1}^{\infty}(2\ell+1)L \times[s_\ell(x)e_\ell(y)Q_\ell''(x,y)+s_\ell'(x)e_\ell'(y)Q_\ell]. \tag{4.7}$$

The operator L when applied to the square parenthesis in (4.7) may be rewritten as

$$\begin{aligned}L[\cdots] &= \xi x L(s_\ell e_\ell''-s_\ell' e_\ell')\\ &= \xi x L[(s_\ell' e_\ell)'-2s_\ell' e_\ell'],\end{aligned} \tag{4.8}$$

to the lowest order in ξ. Making use of (4.5), together with

$$\begin{aligned}s_\ell' e_\ell' &= (-1/2tz)[1+O(\nu^{-2})],\\ L(t) &= -zt^3/\nu, \quad L(tz)=t^3/\nu,\end{aligned} \tag{4.9}$$

we obtain

$$L[\cdots]=(-\xi t/z)[1+O(\nu^{-2})]. \tag{4.10}$$

Then the force becomes

$$\mathscr{F}=\frac{\chi_s^2\xi}{\pi a^2}\int_0^{x_0}dx\sum_{\ell=1}^{\infty}\frac{\nu^3}{\sqrt{\nu^2+x^2}}[1+O(\nu^{-2})]. \tag{4.11}$$

The same kind of divergence is seen to occur here as in the previous case when we were dealing with $\mathscr{F}^{(1)}$. We terminate the sum at $\ell_0=fx_0$ as before, omit the $O(\nu^{-2})$ term, and integrate over x to obtain

$$\mathscr{F}=\frac{\chi_s^2\xi}{\pi a^2}\sum_{\ell=1}^{\ell_0}\nu^3\ln\left\{\frac{x_0}{\nu}+\sqrt{1+\frac{x_0^2}{\nu^2}}\right\}, \tag{4.12}$$

to the lowest order in ξ.

A noteworthy feature of this expression is that ξ appears as a factor, so that the force becomes significantly suppressed for very thin shells.

Figure 5 shows results for $\mathscr{F}$, calculated from (4.12), for some different value of the factor f. The force is seen to be repulsive; this behavior is analogous to that of $\mathscr{F}^{(1)}$ above.

V. CONCLUSION AND FINAL REMARKS

We have assumed throughout that the medium satisfies the condition (1.1). The theory in the main text applies to the case $T=0$ (some remarks on the case of finite temperatures are made in Apendix B).

We may conclude this work as follows.

(1) Using Schwinger's source theory, together with Maxwell's stress tensor, we have calculated the total surface force $\mathscr{F}$, as defined in (2.9), on the shell. The general expression is given in (2.24). For a further analytic processing the compact form (2.26) may be more convenient. At this stage no restriction is imposed on the thickness $(b-a)$ of the shell; nor is the dispersion relation specified. No approximation is so far made on the Riccati–Bessel functions.

When proceeding to numerical calculations we have made two further assumptions; first that the susceptibility χ

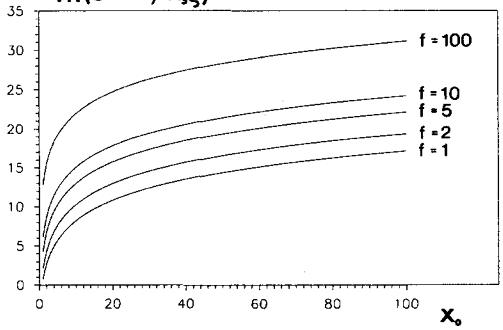


FIG. 5. Force $\mathscr{F}$ on a shell when the medium is dilute. The parameter f is the factor occurring in the angular momentum effective cutoff $\ell_0=fx_0$. The curves are calculated from Eq. (4.12).

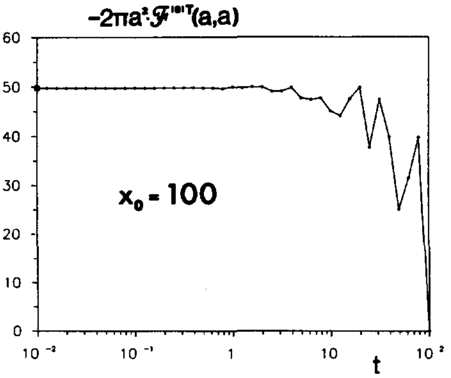


FIG. 6. Finite temperature force $\mathscr{F}^{(0)T}(a,a)$ on a geometrically singular shell, as calculated from Eq. (B4). The flat plateau shows the extent of the low-temperature region. The square point to the left is calculated from (3.12).

satisfies the simple equation (2.28); secondly that the thickness parameter ξ is small.

(2) The limiting case of $\mu_s \to \infty$ or 0 is of main interest, both because of the formalism's analytic manageability, and because we are able to draw lines to results that are expected for ordinary perfectly conducting media. The surface force $\mathscr{F}$ is in this case naturally written as the sum of two terms, $\mathscr{F} = \mathscr{F}^{(0)} + \mathscr{F}^{(1)}$; cf. (3.4) and (3.5). The physical meaning of this decomposition is most clear cut when ξ is very small, $\xi \lesssim 10^{-10}$. Then $\mathscr{F}^{(0)}$ reduces to the singular-shell quantity $\mathscr{F}^{(0)}(a,a)$, as defined by (3.4) with $y = x$, and represented with surprisingly high accuracy by the analytic approximation (3.12), derived by means of the Debye expansion. Here, $\mathscr{F}^{(0)}(a,a)$ shows just the dispersion-induced attractiveness and the nondispersive repulsiveness that is expected for an ordinary perfectly conducting medium in electrodynamics. The second term $\mathscr{F}^{(1)}$ is however strange, in that it diverges when summed over ℓ. On physical grounds we expect that it is appropriate to truncate the sum at an upper limit ℓ_0, being equal to a moderate factor f times x_0 (a photon at frequency ω_0 and impact parameter a has angular momentum equal to x_0). Numerical trials indicate that $f \sim 5$ is suitable in order to extract the order of magnitude of the physical part of $\mathscr{F}^{(1)}$. Figures 2–4 show numerical and analytical results for this case of extreme permeabilities. The agreement between numerics and analysis is generally good. The cutoff *independent* part of $\mathscr{F}^{(1)}$, which may be derived by first letting $x_0 \to \infty$ and thereafter $\ell_0 \to \infty$, is given in (3.22) and is in agreement with the cutoff independent part of our earlier result given in Ref. 10, based upon nondispersive theory from the outset.

(3) One may wonder if the strange behavior of $\mathscr{F}^{(1)}$ is related to the fact that we have assumed an extreme permeability of the medium. (We recall from Ref. 10 that in such a case the total electromagnetic energy within the shell is finite, i.e., the energy *density* is infinite in the interior.) However, this does not appear to be so: The calculation of Sec. IV shows that the repulsive angular momentum divergence persists even if the medium is assumed to be *dilute*. We thus arrive at the conclusion that it is the geometrical singularity of the shell as such which is responsible for the strange behavior of the Casimir force.

(4) To elucidate the role played by the geometrical singularity more explicitly, we shall finally make some comments on the theory of *thick* shells. We assume again that the permeability is extreme, and apply formulas (3.4) and (3.5) to the case of an infinitely thick shell, i.e., a finite and $b \to \infty$. Since we can substitute

$$s_\ell(y) = \tfrac{1}{2}e^y, \quad e_\ell(y) = e^{-y}, \tag{5.1}$$

when $y \to \infty$, we obtain in this way

$$\mathscr{F}^{(0)}(a,b\to\infty) = \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty}(2\ell+1)\left[\frac{s_\ell'(x)}{s_\ell(x)} + \frac{s_\ell''(x)}{s_\ell'(x)} - 2\right], \tag{5.2}$$

$$\mathscr{F}^{(1)}(a,b\to\infty) = \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty}(2\ell+1)\left[\frac{e_\ell'(x)}{e_\ell(x)} + \frac{e_\ell''(x)}{e_\ell'(x)} + 2\right]. \tag{5.3}$$

In each of these expressions, the last term is seen to be divergent when summed over ℓ. However, the important point here is that these infinities cancel when the expression for the total force $\mathscr{F} = \mathscr{F}^{(0)} + \mathscr{F}^{(1)}$ is formed:

$$\mathscr{F}(a,b\to\infty) = \frac{-1}{2\pi a^2}\int_0^{x_0} dx\, x \sum_{\ell=1}^{\infty}(2\ell+1) \times\left[\frac{s_\ell'(x)}{s_\ell(x)} + \frac{s_\ell''(x)}{s_\ell'(x)} + \frac{e_\ell'(x)}{e_\ell(x)} + \frac{e_\ell''(x)}{e_\ell'(x)}\right]. \tag{5.4}$$

This expression is exactly the same as the expression for the term $\mathscr{F}^{(0)}(a,a)$ occurring in the theory for extremely thin shells, $\xi \lesssim 10^{-10}$; cf. Eq. (3.4). Also, the analytic formula (3.12) is immediately applicable for very thick shells. There is thus no need of imposing a finite upper limit in the ℓ summation in this case.

It ought to be mentioned that the force (5.4) is equal to the force $\mathscr{F}_1 = 4\pi a^2 F_1$ that acts on the *inner* surface, $r = a$. This can be verified explicitly by starting from the expression (2.2) for $F_1(\omega)$ and calculating herefrom the expression for $\mathscr{F}_1$. Actually this fact can also be seen directly, without calculation, since the force on a surface is inversely proportional to the square of the radius, so that the force $\mathscr{F}_2$ on the outer surface tends to zero when the radius $b \to \infty$.

For decreasing outer radii the force $\mathscr{F}$ gradually changes. This paper does not discuss the evaluation of $\mathscr{F}$ for a moderately thick shell (this will be done elsewhere). We have merely made some numerical trials in order to test for how small values of ξ the above-mentioned cancellation between divergent terms ceases to exist. It is indicated that $\xi \sim 0.1$ marks the transitional region in this sense. As regards the total Casimir force on the shell, we thus expect to encounter the thin shell peculiarities when $\xi \lesssim 0.1$.

ACKNOWLEDGMENT

One of us (RS) would like to thank the Royal Norwegian Council for Scientific and Industrial Research (NTNF) for financial support during part of this work.

APPENDIX A: NUMERICAL METHOD AND RESULTS

We are faced with the task of numerically evaluating the following expressions for the two parts (3.4) and (3.5) of the Casimir force

$$-2\pi a^2 \mathscr{F}^{(0)} = \int_0^{x_0} dx\, x \sum_{\ell=1}^{\ell_0} (2\ell+1)\left[\frac{s'_\ell(x)}{s_\ell(x)} + \frac{s''_\ell(x)}{s'_\ell(x)} + \frac{e'_\ell(y)}{e_\ell(y)} + \frac{e''_\ell(y)}{e'_\ell(y)}\right], \tag{A1}$$

$$-2\pi a^2 \mathscr{F}^{(1)} = \int_0^{x_0} dx\, x \sum_{\ell=1}^{\ell_0} (2\ell+1)\left[\frac{Q'_\ell(x)}{Q_\ell} + \frac{Q'_\ell(y)}{Q_\ell} + \frac{s''_\ell(x)}{s_\ell(x)}\frac{Q'_\ell(y)}{Q''_\ell(x,y)} + \frac{e''_\ell(y)}{e_\ell(y)}\frac{Q'_\ell(x)}{Q''_\ell(x,y)}\right].$$

Here, Q_ℓ and its derivatives are defined in Eq. (2.16).

The main problem is to generate the Riccati–Bessel functions $s_\ell(x)$ and $e_\ell(x)$, and their first derivatives, for $x \in [0, x_0]$ and $\ell = 1, ..., \ell_0$. These are in principle available from libraries of numerical algorithms via the standard *Modified Spherical Bessel functions* of first and third kind. However, the vast range of the parameters x_0 and ℓ_0 required in the present problem makes these routines inaccessible, mainly due to the use of single instead of double precision in these old routines.

To generate the functions, two simple recursion relations are applicable

$$\begin{aligned} s_{\ell-1}(x) &= s_{\ell+1}(x) + [(2\ell+1)/x]s_\ell(x), \\ e_{\ell+1}(x) &= e_{\ell-1}(x) + [(2\ell+1)/x]e_\ell(x). \end{aligned} \tag{A2}$$

These two recursion relations are of slightly different nature. For $e_\ell(x)$ we start from low values $\ell = 0,1$ and iterate upwards to higher values. For $s_\ell(x)$ the recursion goes downwards, starting from $\ell = \ell_0$, $\ell_0 - 1$. Trying to start an upward recursion for $s_\ell(x)$ by using $s_{\ell+1}(x) = s_{\ell-1}(x) - (2\ell+1)s_\ell(x)/x$ will in general lead to an unstable recursion and wrong results.

The initial values at $\ell = \ell_0, \ell_0 - 1$ may, for low values of x, be generated by the expansion

$$s_\ell(x) = \frac{x^{\ell+1}}{(2\ell+1)!!} \sum_{n=0}^{N} \frac{(2\ell+1)!!}{(2\ell+2n+1)!!} \frac{(\frac{1}{2}x^2)^n}{n!}, \tag{A3}$$

where N is a parameter to be chosen. We use $N = 20$ and can thereby utilize this expansion for values of x up to $x = 10$. This may seem as an overshoot, but we avoid the problems of Ref. 3, where the x axis had to be divided into three intervals. The calculations were performed on an IBM PS2, model 80, which was so fast that summing up to lower values than $N = 20$ would not make any big difference in CPU time. For higher values of x, i.e., $x > 10$, we used a Debye expansion to the fourth order. This Debye expansion is really not necessary. One could start from arbitrary values, do the recursion and normalize afterwards (p. 385 in Ref. 12).

The functions have been generated and checked with the tables in Ref. 12. We reproduce essentially all digits given at pp. 469–475 (at most we see a discrepancy in the last digit for a few cases). Some unimportant misprints are also detected.

The first derivatives of the Riccati–Bessel functions are calculated by the recursion relations

$$\begin{aligned} s'_\ell(x) &= s_{\ell-1}(x) - (\ell/x)s_\ell(x), \\ e'_\ell(x) &= -e_{\ell-1}(x) - (\ell/x)e_\ell(x), \end{aligned} \tag{A4}$$

which involves no numerical difficulties. A check performed is that the Wronskian is -1:

$$W\{s_\ell, e_\ell\} \equiv s_\ell(x)e'_\ell(x) - s'_\ell(x)e_\ell(x) = -1. \tag{A5}$$

The second derivatives are naturally given by the governing differential Eq. (2.25).

One numerical problem that occurs is the over- and underflows of the Riccati–Bessel functions for extreme values of x and ℓ. This sets the limit of ℓ and x. Let us first consider $\mathscr{F}^{(0)}$. Using Eq. (2.25) we can rewrite the first of Eq. (A1) on a form that only involves the ratios s_ℓ/s'_ℓ and e_ℓ/e'_ℓ. These ratios stay finite for a large range of the parameters x and ℓ. It is convenient therefore to generate the logarithm of the functions, thereby avoiding the over- and underflows of the Riccati–Bessel functions that occur for extreme values of x and ℓ. This is not so easily done for $\mathscr{F}^{(1)}$, and therefore the values of x_0 and ℓ_0 are more restricted in this situation. We plan to extend these numerical computations elsewhere, in connection with calculations on a thick shell.

APPENDIX B: THE CASE OF FINITE TEMPERATURES

The results derived in this paper hold at zero temperature. It is worthwhile to comment upon the finite temperature generalization of the theory, not least so because the range of validity of the $T = 0$ formulation will be of experimental interest. We shall however not enter into a complete study of all the finite temperature force terms here. To exhibit the main behavior of the theory, we restrict ourselves to the specific term $\mathscr{F}^{(0)}(a,a)$ in the case of extreme permeability, $\mu_s \to \infty$ or 0. The reason why this term is chosen is of course that it is the "normal" term: it is finite, dispersionally attractive, and is moreover of physical interest for thick shells.

We thus start from the expression (3.8) for $\mathscr{F}^{(0)}(a,a)$, which is a $T = 0$ result obtained with recourse to the Debye expansion. The transition to finite temperatures is accomplished by means of a discretization of the imaginary frequency

$$\hat{\omega} \to \hat{\omega}_n = 2\pi n/\beta, \quad x \to x_n = \hat{\omega}_n a, \tag{B1}$$

where n is an integer in the range $\langle -\infty, \infty \rangle$ and $\beta = 1/k_B T$ (cf. for instance, Ref. 8). The general rule for going from frequency integral to sum is

$$\int_0^\infty dx \to t \sum_{n=0}^{\infty}{}', \tag{B2}$$

with

$$t = 2\pi a/\beta \tag{B3}$$

denoting a nondimensional temperature and the prime in (B2) meaning that the $n = 0$ term is counted with half-weight.

Thus the finite temperature generalization $\mathscr{F}^{(0)T}(a,a)$ of $\mathscr{F}^{(0)}(a,a)$ is

$$\mathscr{F}^{(0)T}(a,a) = \frac{-3t^3}{2\pi a^2} \sum_{n=0}^{n_0}{}' n^2 \sum_{\ell=1}^{\infty} \frac{\nu^5}{(\nu^2 + t^2 n^2)^4}, \quad \text{(B4)}$$

where $n_0 = [x_0/t]$, i.e., the largest integer smaller than or equal to x_0/t. The sum over ℓ in (B4) has been given earlier, in (3.9). The remaining sum over n in (B4) is easily calculated numerically, and so we arrive at an expression for the temperature dependent force with t and x_0 as input parameters.

In the special case then $t \to 0$, and when x_0 is large, the result (B4) must necessarily be in agreement with our previous expression (3.12). It is of interest to investigate how far the low temperature region extends. In this context the following simple argument may be given. We expect to stay within the low temperature region as long as the most significant frequencies in the thermal radiation field are much smaller than the cutoff frequency ω_0 introduced by the dispersion relation. It is natural to identify the most significant frequency with the value ω_m corresponding to the maximum of the blackbody energy distribution. From Wien's displacement law we have $\omega_m = 2.8/\beta$. Our low temperature condition $\omega_m \ll \omega_0$ thus implies $\beta \gg 2.8/\omega_0$, which is equivalent to $t \ll 2x_0$. This simple argument is supported by a direct numerical calculation of the expression (B4). Figure 6 shows how $\mathscr{F}^{(0)T}(a,a)$ varies with t in the case of $x_0 = 100$. The square dot on the left ordinate axis is calculated from (3.12). The $T = 0$ approximation is seen to be adequate until $t \sim 2$, which is much less than $2x_0 = 200$. [The irregular variation to the right in the figure is due to the discontinuous variation of n_0 in (B4).]

Note that, in dimensional terms, $t = (2\pi a/\hbar c) k_B T$. From Ref. 8 we recall, as a useful rule of thumb, that in order to employ the low temperature approximation for the nondimensional temperature t, the radius a must at room temperature be smaller than about 1 μm.

Consider finally the special case $t \gg 1$. Since[8]

$$\sum_{\ell=1}^{\infty} \frac{\nu^5}{(\nu^2 + t^2 n^2)^4} = \frac{1}{6t^2 n^2} [1 + O(t^{-4})], \quad \text{(B5)}$$

for $n \gg 1$, we obtain

$$\mathscr{F}^{(0)T}(a,a) = \frac{-t}{4\pi a^2} \sum_{n=1}^{n_0} [1 + O(t^{-4})] = \frac{-t}{4\pi a^2} \left[\frac{x_0}{t}\right], \quad t \gg 1. \quad \text{(B6)}$$

If, as an additional assumption, x_0 is much larger than t, we can simply approximate $[x_0/t]$ by x_0/t and so obtain

$$\mathscr{F}^{(0)T}(a,a) = -x_0/4\pi a^2, \quad t \gg 1, \quad x_0/t \gg 1. \quad \text{(B7)}$$

This force term is attractive and temperature independent, and is actually seen to be identical with the first term in the zero temperature result (3.12). This means, physically, that the cutoff frequency x_0 is assumed so high that it washes out the influence from the thermal field.

If, on the contrary, x_0 is so small that it lies below t, then $[x_0/t] = 0$ and so the force (B6) vanishes. [Note that in the general expression (B4) there is no contribution from $n = 0$.] The Casimir force does not survive in this approximation. Physically, this means that the dominant frequencies in the thermal radiation field are higher than the cutoff frequency ω_0, and under these conditions the medium behaves essentially like a vacuum.

[1] H. B. G. Casimir, Proc. K. Ned. Akad. Wet. **51**, 793 (1948). For a recent review of the Casimir effect, see G. Plunien, B. Müller, and W. Greiner, Phys. Rep. **134**, 87 (1986).

[2] P. Candelas, Ann. Phys. (N. Y.) **143**, 241 (1982).

[3] I. Brevik and G. Einevoll, Phys. Rev. D **37**, 2977 (1988).

[4] T. H. Boyer, Phys. Rev. **174**, 1764 (1968).

[5] B. Davies, J. Math. Phys. **13**, 1324 (1972).

[6] R. Balian and B. Duplantier, Ann. Phys. (N.Y.) **112**, 165 (1978).

[7] K. A. Milton, L. L. DeRaad, Jr., and J. Schwinger, Ann. Phys. (N.Y.) **115**, 388 (1978).

[8] I. Brevik and I. Clausen, Phys. Rev. D **39**, 603 (1989).

[9] I. Brevik, Can. J. Phys. **61**, 493 (1983).

[10] I. Brevik and H. Kolbenstvedt, Can. J. Phys. **63**, 1409 (1985).

[11] J. Schwinger, *Particles, Sources, and Fields* (Addison-Wesley, Reading, MA, 1970), Vol. I.

[12] *Handbook of Mathematical Functions*, Natl. Bur. Stand. Appl. Math. Ser. No. 55, edited by M. Abramowitz and I. A. Stegun (U. S. GPO, Washington, DC, 1964). Reprinted by Dover, New York, 1972.

[13] I. Brevik and H. Kolbenstvedt, Ann. Phys. (N.Y.) **143**, 179 (1982).

[14] I. Brevik, J. Phys. A **20**, 5189 (1987).

[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, London, 1980).

Anisotropic cosmological models with energy density dependent bulk viscosity

Wung-Hong Huang
Department of Physics, National Cheng Kung University, Tainan, Taiwan 70101, Republic of China

(Received 28 December 1988; accepted for publication 24 January 1990)

An analysis is presented of the Bianchi type I cosmological models with a bulk viscosity when the universe is filled with the stiff fluid $p = \epsilon$ while the viscosity is a power function of the energy density, such as $\eta = \alpha|\epsilon|^n$. Although the exact solutions are obtainable only when the $2n$ is an integer, the characteristics of evolution can be clarified for the models with arbitrary value of n. It is shown that, except for the $n = 0$ model that has solutions with infinite energy density at initial state, the anisotropic solutions that evolve to positive Hubble functions in the later stage will begin with Kasner-type curvature singularity and zero energy density at finite past for the $n \geqslant 1$ models, and with finite Hubble functions and finite negative energy density at infinite past for the $n < 1$ models. In the course of evolution, matters are created and the anisotropies of the universe are smoothed out. At the final stage, cosmologies are driven to infinite expansion state, de Sitter space-time, or Friedman universe asymptotically. However, the de Sitter space-time is the only attractor state for the $n < \frac{1}{2}$ models. The solutions that are free of cosmological singularity for any finite proper time are singled out. The extension to the higher-dimensional models is also discussed.

I. INTRODUCTION

The investigation of relativistic cosmological models usually has the energy momentum tensor of matter as that due to a perfect fluid. To consider more realistic models one must take into account the viscosity mechanisms, which have already attracted the attention of many investigators. Misner[1] suggested that strong dissipative due to the neutrino viscosity may considerably reduce the anisotropy of the blackbody radiation. Viscosity mechanism in the cosmology can explain the anomalously high entropy per baryon in the present universe.[2,3] Bulk viscosity associated with the grand-unified-theory phase transition[4] may lead to an inflationary scenario.[5–7]

An exactly soluble isotropic cosmological model of the zero curvature Friedman model in the presence of bulk viscosity has been examined by Murphy.[8] The solutions that he found exhibit an interesting feature that the big bang type singularity appears in infinite past. Exact solutions of the isotropic homogeneous cosmology for the open, closed and flat universe have been found by Santos *et al.*,[9] when the bulk viscosity is the power function of energy density. However, in some cases, the big bang singularity occurs at finite past. It is thus shown that Murphy's conclusion that the introduction of bulk viscosity can avoid the initial singularity at finite past is not, in general, valid. (The extensive collections of exact isotropic solutions are those found in Ref. 10.)

Belinskii and Khalatnikov[11] analyzed the Bianchi type I cosmological models under the influence of viscosity. They then found the remarkable property that near the initial singularity the gravitional field creates matters. Using a certain simplifying assumption, Banerjee and Santos[12,13] obtained some exact solutions for the homogeneous anisotropic model. Recently Banerjee *et al.*[14] obtained some Bianchi type I solutions for the case of stiff matter by using the assumption that shear viscosity is the power function of the energy density. However, the bulk viscosity coefficients adopted in their model are zero or constant.

In this paper, without introducing the shear viscosity, we shall examine the Bianchi type I cosmological models with bulk viscosity (η) that is a power function of energy density (ϵ), i.e., $\eta = \alpha|\epsilon|^n$, when the universe is filled with the stiff matter $p = \epsilon$. We are interested in the cosmological solutions that will eventually go to the states of positive Hubble functions. The exact solutions are obtained when $2n$ is an integer. Furthermore, through some analyses, we can know how evolutions of the models with arbitrary value of n will be. We prove that the isotropic de Sitter space-time is a stable attractor state as $t \to \infty$ if $n < \frac{1}{2}$. It is thus in accord with the "cosmic no hair" theorem[15–17] even though the strong energy condition[18] is violated.[19] (The weak energy condition question in the anisotropic viscous models has been discussed by Barrow.[20]

We also find that, for the models of $n < 1$ there are solutions (in fact, all solutions in $0 < n < \frac{1}{2}$ models) that can avoid the cosmological singularity at any finite proper time. Our show models that the anisotropies of the universe are smoothed out and matters are created by the gravitational field in the course of the evolution, in agreement with the results found by others.[11,14]

The models of $n = 0$ and 1 have been discussed in our previous paper.[21] However, the method invented there cannot be used to solve the models with other n, and the analyses of the $n = 0$ model were incomplete. Barrow[22] had also given further discussion of bulk viscous models in theories possessing quadratic curvature.

It is worth mentioning that the $n = 3/2$ model can be used to describe the quantum production of infinitely thin Witten strings[23] on super-horizon scales in the very early universe (see the arguments in the paper of Turok[24]), which has been analyzed by Barrow[10] recently.

The outline of this paper is as follows. We first discuss

 0022-2488/90/061456-07$03.00

the axially symmetric Bianchi type I model in which there are ony two cosmic scale functions. In Sec. II we reduce the Einstein's field equation to a pair of coupled differential equations that become an integratable equation as we define two suitable variables in Sec. III. The exact solutions for the models with integer $2n$ are then found. The analyses about the initial and final states of the models with any n are given in Sec. IV. To discuss the generic Bianchi type I model with multiple cosmic scale functions we then in Sec. V present a simple method that can also be used to analyze the higher-dimensional models. Section VI is devoted to conclusions.

II. EINSTEIN'S FIELD EQUATIONS AND SOME ANALYSES

We first consider the axially symmetric Bianchi type I model with a metric in the form

$$ds^2 = g_{\mu\nu}\, ds^\mu\, dx^\nu$$
$$= -dt^2 + X(t)^2\, dx^2 + Y(t)^2(dy^2 + dz^2), \quad (2.1)$$

where X and Y are functions of cosmic time t alone. The field equations to be solved are

$$R_{\mu\nu} = ((\epsilon - \bar{p})/2)g_{\mu\nu} + (\epsilon + \bar{p})u_\mu u_\nu, \quad (2.2)$$

where ϵ is the energy density, and u_μ is the four-velocity that satisfies

$$u_\mu u^\mu = -1. \quad (2.3)$$

The total pressure $\bar{p}$ is defined by

$$\bar{p} = p - \eta u^\mu{}_{;\mu}, \quad (2.4)$$

where p is the pressure coming from the perfect fluid and η is the bulk viscosity. Choosing a comoving frame where $u_\mu = \delta^0{}_\mu$, the Einstein's field equations (2.2) lead to

$$\frac{dH}{dt} + HW = (\epsilon - \bar{p})/2, \quad (2.5)$$

$$\frac{dh}{dt} + hW = (\epsilon - \bar{p})/2, \quad (2.6)$$

$$W^2 - (2H^2 + h^2) = 2\epsilon, \quad (2.7)$$

where H and h are the Hubble functions defined by

$$H \equiv \frac{dY}{dt} Y^{-1}, \quad (2.8a)$$

$$h \equiv \frac{dX}{dt} X^{-1}, \quad (2.8b)$$

and W is the total expansion function

$$W \equiv 2H + h. \quad (2.8c)$$

We also have a relation

$$\epsilon - \bar{p} = (2 - \gamma)\epsilon + \eta W, \quad (2.9)$$

where γ is defined by the equation of state

$$p = (\gamma - 1)\epsilon, \quad 1 \leqslant \gamma \leqslant 2. \quad (2.10)$$

In this paper we only consider the stiff matter, i.e., $\gamma = 2$, which is the possible relevance of the equation of state $p = \epsilon$ regarding the matter content of the early universe.[25,26] Using the assumption that the bulk viscosity is a power function of energy density,

$$\eta = \alpha \epsilon^n, \quad n \geqslant 0. \quad (2.11)$$

Equations (2.5) and (2.6) then become

$$\frac{dH}{dt} + HW = \frac{\alpha}{2}[H(H+2h)]^n W, \quad (2.12)$$

$$\frac{dh}{dt} + hW = \frac{\alpha}{2}[H(H+2h)]^n W. \quad (2.13)$$

The work that remains is to analyze the above coupled differential equations and find their exact solutions. We will express these solutions as the flows in the phase space $H \times h$ and thus find the dynamical evolutions of cosmology. We only consider the physical plane where the trajectories shall evolve to the positive Hubble functions in the latter stage. The regions where $\epsilon < 0$ that violate the weak energy condition[18] are needed for the solutions of $n < 1$ models as discussed in next section. We then find there that the evolutions of the cosmology are confined in the regions of $H \geqslant 0$ and $H + 2h \geqslant 0$ for the $n \geqslant 1$ models, while these for the $n < 1$ models are $W \geqslant 0$.

Let us first determine the singular points, fixed points, and cosmic time in the phase space.

A. Singular points

Equation (2.7) tells us that the energy density becomes infinite only if H and/or h are infinite. From Eqs. (2.12) and (2.13) we also know that dH/dt and dh/dt can be infinite only if H and/or h are infinite. As the Riemann scalar curvature can be written as

$$R = 2\frac{dW}{dt} + W^2 + 2H + h^2,$$

we see that R can be infinitely large only if H and/or h are infinite. Therefore, *the singularity of diverge R and ϵ could occur only if H and/or h is infinite.*

B. Fixed points

The fixed points are the solutions of Eqs. (2.12) and (2.13) once we let $dH/dt = dh/dt = 0$. They are

$$(1)\, W = 0 \Rightarrow H = h = 0 \text{ or } 2H = -h \neq 0, \quad (2.14a)$$

$$(2)\, H = h = H_D \equiv [(3^n/2)\alpha]^{1/(1-2n)}, \quad n \neq \tfrac{1}{2}, \quad (2.14b)$$

$$(3)\, H = h = \forall \text{ value if } \alpha = \alpha_c \equiv 2/\sqrt{3}, \quad n = \tfrac{1}{2}. \quad (2.14c)$$

Using Eqs. (2.12) and (2.13) we can determinate the signs of dH/dt and dh/dt in the neighborhoods of the de Sitter space-time of $H = h = H_D$ and thus determine the stability of the de Sitter state. It is then found that the *isotropic cosmologies with $n < \frac{1}{2}$ display inflationary behavior* but those with *$n > \frac{1}{2}$ shall exhibit the deflationary behavior* as found in Ref. 9. *When $n = \frac{1}{2}$, then $h = h = 0$ and $H = h = \infty$ is the attractor state for $\alpha < \alpha_c$ and $\alpha > \alpha_c$, respectively.* The investigation appears to display the division into $n > \frac{1}{2}$, $n = \frac{1}{2}$, and $n < \frac{1}{2}$, also discussed by Barrow.[19,10]

C. Cosmic time

The solutions expressed as the trajectories on the phase plane do not explicitly depend on the cosmic time. However, through a simple analysis we can determine whether the proper time in a solution is finite or infinite.

Equation (2.12) can lead to

$$\int^{H} \frac{d\tilde{H}}{[(\alpha/2)[\tilde{H}(\tilde{H}+2\tilde{h})]^n - \tilde{H}]\tilde{W}} = \int^{t} d\tilde{t}. \quad (2.15)$$

We then see that the left-hand side in the above equation will become infinite only if H is on a fixed point, and one can show that the *cosmic time of a point on a trajectory corresponding to a solution will be infinite if the trajectory has already met a fixed point.* On the contrary, *the trajectory starting with diverge H will have finite proper time.*

III. EXACT SOLUTIONS

We now begin to solve Eqs. (2.12) and (2.13). After dividing the former one by the latter one, we obtain

$$\frac{dH}{dh} = \frac{\alpha[H(H+2h)]^n - 2H}{\alpha[H(H+2h)]^n - 2h}. \quad (3.1)$$

The above equation can lead to

$$\frac{dL}{dh} = \frac{3\alpha[HL]^n - 2L}{\alpha[HL]^n - 2h}, \quad (3.2a)$$

$$\frac{dB}{dh} = \frac{-2B}{\alpha[HL]^n - 2h}, \quad (3.2b)$$

where

$$L \equiv H + 2h, \quad B \equiv 2(H-h). \quad (3.3)$$

Dividing Eq. (3.2a) by Eq. (3.2b) one gets

$$\frac{dL}{dB} = \frac{L}{B}\left[1 - \frac{\alpha}{2} 3^{1-n} L^{2n-1}\left(1 + \frac{B}{L}\right)^n\right]. \quad (3.4)$$

Using a new variable

$$A = L/B, \quad (3.5)$$

then Eq. (3.4) gives a simple form as

$$dA/(A^2+A)^n = -(\alpha/2)3^{1-n}B^{2n-2}\,dB \quad (\text{for } \epsilon \geqslant 0). \quad (3.6)$$

The variables are thus separated and the equation becomes integratable. Through the integration by part, the exact solutions can be found when $2n$ is an integer.

Since we will describe our solutions in the phase plane $H \times h$, we express H and h in the variables r and θ:

$$H \equiv r\cos\theta, \quad (3.7a)$$

$$h \equiv r\sin\theta, \quad (3.7b)$$

in terms of which A and B become

$$A = \frac{1 + 2\tan\theta}{2(1-\tan\theta)}, \quad (3.8)$$

$$B = 2r(\cos\theta - \sin\theta). \quad (3.9)$$

Therefore, if Eq. (3.6) can be integrated exactly, the solutions that relate function r to the variable θ will be found, and trajectories in phase plane can be plotted exactly, which in turn determine the evolutions of the cosmology. Various integration constants chosen in integrating Eq. (3.6) will produce the various trajectories which correspond to the solutions of the same model but with different initial states. The arrows in the trajectories, which tell us the directions of the evolution of the cosmology, are easily determined from Eqs. (2.12) and (2.13).

As argued below, we also need to consider the solutions of negative energy density. A natural extension is the models with the viscosity function $\eta = \alpha(-\epsilon)^n$. Through the same procedure we can find the following relations:

$$dA/(-A^2-A)^n = -(\alpha/2)3^{1-n}B^{2n-2}\,dB \quad (\text{for } \epsilon<0;\ H,W>0), \quad (3.10a)$$

$$dA/(-A^2+A)^n = (\alpha/2)3^{1-n}B^{2n-2}\,dB \quad (\text{for } \epsilon, H \leqslant 0;\ W \geqslant 0). \quad (3.10b)$$

The variables are separated and the exact solutions can be found when $2n$ is an integer.

As examples, we will give some explicit solutions:

(1) $n = 0$

$$h - H_D = C(H - H_D), \quad (3.11)$$

where C is an integration constant, and H_D is defined in Eq. (2.14b). The solutions are thus all the straight lines that pass through the point of $H = h = H_D$. This means that the initial state of cosmology shall begin with H and/or $h \to \pm\infty$. Therefore, initial singularity will arise. However, we must be careful now. The analyses in the above section show that the points on the line of $W = 0$ are the fixed points. Therefore some anisotropic solutions will begin on these points (in which ϵ is negative) at *infinite past*; during the evolution the cosmologies are isotropized and driven to the de Sitter state asymptotically. Using the relation of Eq. (3.11) we can from Eq. (2.12) find the function $H(t)$ which then explicitly shows this fact.

Note that the energy density of these solutions is negative in the early stage; this will violate both the weak energy condition and the strong energy condition, and there is no singularity at any finite proper time.

Although it is difficult for the negative energy density to appear classically, it could be found in the quantized matter field. Also, as Hu[27] has discussed, the quantum dissipative process of the particle production could be formulated in terms of relativistic imperfect fluid. Accordingly, it seem that there will be, more or less, some quantum senses in these solutions. It is interesting to mention that the introducing of quantized matter field into the energy momentum tensor can sometimes lead to avoidance of the cosmological singularity, as found by Parker Fulling.[28]

(2) $n = 1/2$

$$r = C[\pm(\cos\theta - \sin\theta)]^{(\alpha_c/\alpha)-1} \times [\sqrt{\cos\theta + 2\sin\theta} + 3\sqrt{\cos\theta}]^{-2\alpha_c/\alpha} \quad (\text{for } \epsilon \geqslant 0), \quad (3.12a)$$

where C is an integration constant and α_c is the constant defined in Eq. (2.14c). The solution shows that H and/or h can go to infinity only if $H = h$, i.e., $\cos\theta = \sin\theta$. Therefore, this model is with finite Hubble functions at the states of zero energy density ($H = 0$ or $H + 2h = 0$).

The experiences from the analyses of the $n = 0$ model tell us that, as the points on the lines of zero energy density are not the fixed point, it is now also needed to discuss the regions of $\epsilon < 0$. (Note that we only discuss the expanding solutions, i.e., $W \geqslant 0$.) We then must analyze the extended model in which the bulk viscosity is $\eta = \alpha(-\epsilon)^{1/2}$. Equation (3.10) can give the explicit solutions:

$$r = \frac{C}{(\sin\theta - \cos\theta)}\exp\left[\pm\frac{\alpha_c}{\alpha}\sin^{-1}\left(\frac{2+\tan\theta}{1-\tan\theta}\right)\right]$$
$$(\text{for } \epsilon \leqslant 0) \qquad (3.12\text{b})$$

where C is an integration constant.

After plotting the flows in the phase plane (see Fig. 1, in which, for clarity we adopt the nonuniform scale), we then find that anisotropic cosmologies shall always begin with finite negative energy density at the states with $2H + h = 0$ in infinite past; the Riemann scalar curvature and Hubble functions are finite in the initial phase. During the evolution, the energy density is increasing subsequently and the anisotropies of the universe are smoothed out. At the final stage as $t \to \infty$, depending on the value of α (and not on the value of C), there are three classes of states that may be approached asymptotically:

(1) Both the energy density and Hubble functions go to infinity if $\alpha > \alpha_c$.

(2) Energy density is finite and space-time is attracted to a de Sitter universe (which is a function of the value of C) if $\alpha = \alpha_c$.

(3) Both the energy density and Hubble functions decrease to approach zero and the model is driven to the isotropic Friedman universe if $\alpha < \alpha_c$.

The last two cases provide us with the solutions that are free cosmological singularity for all finite proper time.

(3) $n = 1$

$$r = [\alpha(\cos\theta - \sin\theta)]^{-1}[C - \ln(1 + 2\tan\theta)], \qquad (3.13)$$

where C is an integration constant. After plotting the flows in the phase plane (see Fig. 2, where, for clarity, we adopt the nonuniform scale), we then find that the anisotropic cosmologies shall start from the vacuum states and end in another fixed point or infinite expansion state. It is then found that, except in the isotropic model ($h = H$) that has been investigated by Murphy,[1] the cosmologies shall always begin with zero energy density at the initial phase of singularity. During the evolution, the energy density is increasing subsequently and the anisotropies of the universe are smoothed out. At the final stage as $t \to \infty$, depending on the integration constant C

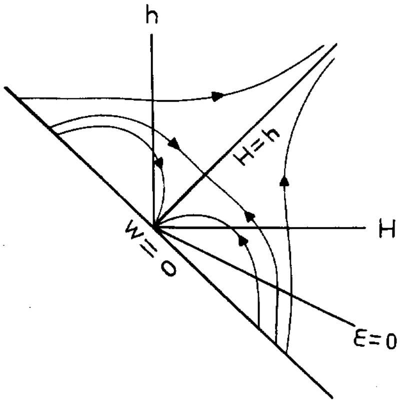


FIG. 1. The phase plane trajectories of the $n = \frac{1}{2}$ model. All solutions will begin with $W = 0$, $\epsilon < 0$ at $t \to -\infty$.

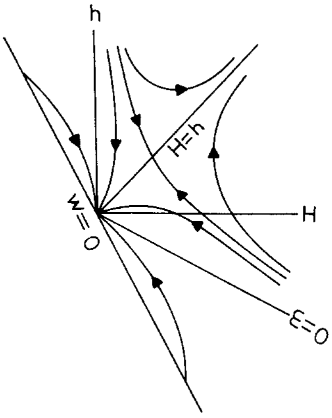


FIG. 2. The phase plane trajectories of the $n = 1$ model. The solutions that evolve into positive Hubble functions at later stage will begin with $\epsilon = 0$ at finite past.

(and not on the value of α), there are three classes of state that may be approached asymptoptically:

(1) Both the energy density and Hubble functions go to infinity if $H > h$, $C > \ln 3$ or $h > H$, $\ln 3 > C > 0$.

(2) Energy density is finite and space-time is attracted to a de Sitter universe [which is determined by Eq. (2.14b)] if $C = \ln 3$.

(3) Both the energy density and Hubble functions decrease to approach zero and model is driven to the isotropic Friedmann universe if $H > h$, $\ln 3 > C > 0$ or $h > H$, $C > \ln 3$.

The solutions that start at fixed point on the line of $W = 0$ are also described by Eq. (13) if one lets $\alpha \to -\alpha$. However, the trajectories that begin with $W = 0$ do not go into positive Hubble functions states at the final stage.

(4) $n = 3/2$

$$r^2 = \frac{1}{\alpha(\cos\theta - \sin\theta)^2}\left[\frac{4(2+\tan\theta)}{\sqrt{1+2\tan\theta}} - c\right]. \qquad (3.14)$$

This model can be used to describe the quantum production of infinitely thin Witten strings[23] on super-horizon scales in the very early universe.[24,10]

(5) $n = 2$

$$r^3 = \frac{1}{\alpha(\cos\theta - \sin\theta)^3}\left[\frac{3(2+\tan\theta)(1-\tan\theta)}{1+2\tan\theta} + \frac{9}{2}(1 + 2\tan\theta) - c\right] \qquad (3.15)$$

and so on.

All the models of $n \geqslant 1$, as will be proved in the next section (for any n), possess the same characteristics such as the isotropization of the cosmology, beginning with zero energy density and with infinite Riemann scalar curvature, creating the matters in the course of evolution, and having three classes of final state.

IV. ANALYSES OF INITIAL AND FINAL STATES

We give in this section the analyses of the initial and final states. The results can clarify the characteristics of the cosmological evolutions for the models of any values of n.

A. Initial states

The key equation to analyze the initial and final states of our model is Eq. (3.6). It can lead to

$$C_1 \int^A (\tilde{A}^2+\tilde{A})^\sigma d\tilde{A} + C_2(2A+1)(A^2+A)^\sigma + C_3\frac{(2A+1)}{(A^2+A)^{1-\sigma}} + C_4\frac{(2A+1)}{(A^2+A)^{n-1}} + C_5\frac{(2A+1)}{(A^2+A)^{n-2}} + \cdots$$
$$= C_0 B^{2n-1} = C_0[2r(\cos\theta - \sin\theta)]^{2n-1}, \tag{4.1}$$

where C_i are constant numbers which only depend on n, and σ is chosen to satisfy $0<\sigma<1$. It is important to notice that C_3 is nonzero only if $n\geqslant 1$, C_4 is nonzero only if $n\geqslant 2,\ldots,$ and so on. If the universe is in the initial state $H+2h\to 0$, then $A\to 0$ and $A^2+A\to 0$, and we can easily prove that the first term (neglects the integration constant) and second term on the left-hand side of Eq. (4.1) shall approach zeros while the other terms become infinite. Therefore the value of r in Eq. (4.1) is finite for the states that have vanish energy density, if and only if $0<n<1$.

However, it is now necessary to discuss the solutions in the regions of $\epsilon\leqslant 0$, as the $\epsilon=0$ state is not the fixed point on the phase plane. From Eq. (3.10) one can show that r is finite as $W\to 0$. Hence we have proved that the *cosmologies of* $0<n<1$ *shall begin with finite negative energy density and zero total expand function at infinite past.* They will then go into the positive energy density state. Although the *models of* $n\geqslant 1$ *shall begin with zero energy density and diverge values of* H *and* h, one can from Eq. (2.15) conclude that *the models with* $n\geqslant 1$ *shall begin at finite past.*

Furthermore, for the $n\geqslant 1$ models we can from Eqs. (2.5) and (2.6) prove that near the initial phase the curvature singularity is the Kasner type, although the energy density is zero. This fact was first found in the letter of Belinskii and Khalatnikov,[29] in which only the anisotropic $n=1$ was analyzed.

B. Final states

To analyze the final states we consider three cases separately:

Case 1: $n>1$

The key equation can lead to

$$C_1\int^A \frac{d\tilde{A}}{(\tilde{A}^2+\tilde{A})^\sigma} + C_2\frac{(2A+1)}{(A^2+A)^{n-1}} + C_3\frac{(2A+1)}{(A^2+A)^{n-2}}$$
$$+\cdots = C_0[2r(\cos\theta-\sin\theta)]^{2n-1}, \tag{4.2}$$

where C_i are constant numbers that depend only on n, and σ is chosen to satisfy $1<\sigma<2$. The value of C_2 is nonzero only if $n\geqslant 2$, C_3 is nonzero only if $n\geqslant 3,\ldots,$ and so on. As universe approaches isotropic state, i.e., $H\to h$, then $A\to\infty$ and $B\to 0$, which in turn implies that the first term (neglects the integration constant) on the left-hand side of Eq. (4.2) becomes zero. Therefore, depending on the chosen integration constant, r may be $\pm\infty$. This situation is like that in the model of $n=1$. Therefore, depending on the initial state, the cosmology may be driven to infinite expansion state, de Sitter space-time, or isotropic Friedmann universe at the final stage.

Case 2: $\frac{1}{2}<n<1$

The key equation can lead to

$$C_1\int^A \frac{d\tilde{A}}{(\tilde{A}^2+\tilde{A})^\sigma} + C_2\frac{(2A+1)}{(A^2+A)^n}$$
$$= C_0[2r(\cos\theta-\sin\theta)]^{2n-1}, \tag{4.3}$$

where C_i are constant numbers that depend only on n, and $\sigma\equiv n+1$, thus $1<\sigma<2$. Using the arguments like that in *Case 1,* we can also show that the models with $1>n>\frac{1}{2}$ have three classes of final states. Hence we have proved that, *depending on the initial state, the cosmologies for the models of* $n>\frac{1}{2}$ *may be driven to infinite expansion state, de Sitter space-time, or isotropic Friedmann universe at the final stage.* The model of $n=\frac{1}{2}$ was discussed in Sec. III.

Case 3: $n<\frac{1}{2}$

The key equation can lead to

$$-\frac{n}{2(n-1)}\left[\int_A \frac{d\tilde{A}}{(\tilde{A}^2+\tilde{A})^{n+1}} + \frac{2A+1}{n(A^2+A)^n}\right]$$
$$= \frac{-3^{1-n}\alpha}{2(2n-1)}B^{2n-1}. \tag{4.4}$$

As universe approaches isotropic state, i.e., $H=h$ (it implies $A\to\infty$ and $B\to 0$), it is easy to prove that the first term in the bracket becomes zero (neglects the integration constant) and

$$r\to[3^n 2^{-(n+\frac{1}{2})}\alpha]^{1/(1-2n)}, \tag{4.5}$$

no matter what the value of integration constant that will be chosen. The state corresponding to Eq. (4.5) can be easily checked to be just the fixed point defined in Eq. (2.14b). Hence, we have proved that *solutions of the models of* $n<\frac{1}{2}$ *shall always be attracted to an isotropic de Sitter state at final stage.*

For the $n\geqslant 1$ models there are the solutions that start with $W=0$ at infinite past and then are attracted to the original point on the phase plane. However, they never go into the state of positive Hubble function.

V. MODELS WITH MULTIPLE HUBBLE FUNCTIONS

The methods described in the above sections can only be used to study the models with two Hubble functions. We will now give a simple algorithm that can be used to analyze the models with multiple Hubble functions.

Let us consider the $D+1$-dimensional Binachi type I models. The Einstein's field equation Eq. (2.2) can lead to

$$\frac{dH_i}{dt} + H_i W = (\epsilon - \bar{p})/2, \quad i=1,2,\ldots,D, \tag{5.1}$$

$$W^2 - \sum_i H_i^2 = \epsilon, \tag{5.2}$$

where H_i are the Hubble functions and W is the total expansion function. With the relation Eq. (2.9), and letting $\gamma = 2$ in there, we can from Eq. (5.1) find

$$\frac{d\ln(H_i - H_j)}{dt} = \frac{d\ln(H_i - H_k)}{dt}. \quad (5.3)$$

This equation tells us that one can express all other D–2 Hubble functions in terms of only two Hubble functions. One can see that this is a very general property and may be used to analyze many other anisotropical cosmological models. As examples, we will now describe the procedure of how to determine the three Hubble functions in the four-dimensional model.

Equation (5.3) can yield a relation

$$H_3 = (1 - C)H + Ch, \quad (5.4)$$

in which C is an integration constant, and for simplification, H_1 and H_2 are denoted as H and h. Using Eq. (5.4) we can from Eq. (5.2) obtain

$$\epsilon = (1 - C)H^2 + 2Hh + Ch^2 = KL, \quad (5.5)$$

$$K \equiv [(1 - C)H + ah], \quad (5.6)$$

$$L \equiv [H + bh], \quad (5.7)$$

where the constants a and b are the functions of C. Substituting the relation (5.5) into the viscosity function Eq. (2.11), then the Einstein's field equations (5.1) lead to

$$\frac{dK}{dt} = \frac{\alpha}{2}(1 - C + a)(KL)^n W - KW, \quad (5.8)$$

$$\frac{dL}{dt} = \frac{\alpha}{2}(1 + b)(KL)^n W - LW. \quad (5.9)$$

Dividing Eq. (5.8) by Eq. (5.9), one gets

$$\frac{dK}{dL} = \frac{\alpha(1 - C + a)(KL)^n - 2K}{\alpha(1 + b)(KL)^n - 2L}. \quad (5.10)$$

After defining the variables

$$A \equiv L/B, \quad (5.11a)$$

$$B \equiv (1 + b)K - (1 - C + a)L, \quad (5.11b)$$

Eq. (5.10) gives a simple form

$$\frac{dA}{[A^2 + (A/(1 - C + a))]^n} = -\frac{\alpha(1 - C + a)}{2} B^{2n-2}\, dB. \quad (5.12)$$

The variables are now separated and the equation is integratable. (One can prove that $1 - C + a$ is nonzero.) Using the methods described in the above sections we can therefore analyze any dimensional Bianchi type I cosmological models with energy density dependent bulk viscosity. The results show the same characteristics as those in the models with two Hubble functions.

VI. CONCLUSIONS

We have analyzed in detail the anisotropic cosmological models with bulk viscosity (η) which is a power-law dependence upon energy density (ϵ), i.e., $\eta = \alpha|\epsilon|^n$, when the universe is filled with stiff matter $p = \epsilon$. We are interested in the cosmological solutions that will eventually go to the states of positive Hubble functions in the latter stage. Although the exact solutions could be obtained only when the $2n$ is an integer, we are able to clarify the characteristics of evolution for the models of any n.

Let us give a summary. (1) There have been two kinds of solutions in the $n = 0$ model, which start either with diverge Hubble functions and infinite energy density at finite past or with finite Hubble functions and negative energy density at infinite past. However, both solutions are driven to a de Sitter space-time asymptotically. (2) All the solutions in the $0 < n < \frac{1}{2}$ models will start with finite Hubble functions and negative energy density at infinite past and then are driven to a de Sitter space-time asymptotically. (3) For the $\frac{1}{2} \leqslant n < 1$ models, the initial state is with finite Hubble functions and negative energy density at infinite past; however, they can go to the infinite expansion state, de Sitter space-time, or Friedmann universe at final stage. (4) The $n \geqslant 1$ models will always begin with Kasner-type curvature singularity at final past in which the energy density is zero, however; and then they are driven to the above-mentioned three kinds of states asymptotically. All the solutions that begin with $\epsilon < 0$ and then are attracted to a de Sitter space-time or Friedmann universe are free of cosmological singularity for any finite proper time.

Historically, Murphy[8] presented the exact solution of the $n = 1$ model, and showed that the bulk viscosity can eliminate the big band singularity at any finite proper time. Belinskii and Khalatnikov[29] then analyzed the $n = 1$ Bianchi I model; they found that the cosmology is with the vanish energy density in the initial phase in which the Kasner singularity will arise. Now, the investigation in this paper shows that all the $n \geqslant 1$ Bianchi I models will share the same characteristics of initial singularity, and that the $n < 1$ Bianchi models could give us some cosmological solutions that are free of singularity for all finite proper time. However, these singular-free solutions have negative energy density in the early epoch.

The models discussed in this paper are only for the stiff matter. As the case of stiff matter is special because the shear and matter density behave in the same way in the absence of viscosity, and vacuum and nonvacuum perfect fluid solutions are formally similar, the same models, while with other matter fields are certainly interesting, remain to be studied.

Finally, we want to mention that the prescription adopted in this paper can also be used to analyze the Bianchi type I cosmological models with energy density dependent shear viscosity.[30] The details will be presented elsewhere.

[1] C. W. Misner, Nature **214**, 40 (1967); Astrophys. J. **151**, 431 (1968).
[2] S. Weinberg, Astrophys. J. **168**, 175 (1972).
[3] S. Weinberg, *Gravitation and Cosmology* (New York, Wiley, 1972).
[4] P. Langacker, Phys. Rep. **72**, 185 (1981).
[5] I. Waga, R. C. Falcao, and R. Chanda, Phys. Rev. D **33**, 1839 (1986).
[6] T. Pacher, J. A. Stein-Schabes, and M. S. Turner, Phys. Rev. D **36**, 1603 (1987).
[7] A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[8] G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).
[9] N. O. Santos, R. S. Dias, and A. Banerjee, J. Math. Phys. **26**, 878 (1985).
[10] J. D. Barrow, Nucl. Phys. B **310**, 743 (1988).
[11] V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1976); Sov. Phys. JETP **45**, 1 (1977).
[12] A. Banerjee and N. O. Santos, J. Math. Phys. **24**, 2689 (1983).

[13]A. Banerjee and N. O. Santos, Gen. Relativ. Gravit. **16**, 217 (1984); **18**, 461 (1986).
[14]A. Banerjee, S. B. Duttachoudhury, and A. K. Sanyal, J. Math. Phys. **26**, 3010 (1985).
[15]G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[16]S. W. Hawking and I. G. Moss, Phys. Lett. B. **110**, 35 (1982).
[17]R. M. Wald, Phys. Rev. D **28**, 2118 (1983); L. G. Jensen and A. Stein-Schabes, Phys. Rev. D **35**, 1146 (1987).
[18]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure Of Spacetime* (Cambridge U.P., Cambridge, 1973).
[19]J. D. Barrow, Phys. Lett. B. **180**, 335 (1986).
[20]J. D. Barrow, in *The Early Universe*, edited by W. Unruh and G. W. Semenoff (Reidel, Dordrecht, 1988), pp. 191–194.
[21]W. -H. Huang, Phys. Lett. A. **129**, 429 (1988).
[22]J. D. Barrow, Phys. Lett. B. **183**, 285 (1987).
[23]E. Witten, Phys. Lett. B. **153**, 243 (1985).
[24]N. Turok, Phys. Rev. Lett. **60**, 548 (1988).
[25]Y. B. Zeldorich, Sov. Phys. JETP **14**, 1143 (1962).
[26]J. D. Barrow, Nature **272**, 211 (1978).
[27]B. L. Hu, Phys. Lett. A **90**, 375 (1982).
[28]L. Parker and S. A. Fulling, Phys. Rev. D **7**, 2357 (1973).
[29]V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP Lett. **21**, 99 (1975).
[30]W. -H. Huang, "Effects of the shear viscosity on the character of cosmological evolution," J. Math Phys. **31**, 659 (1990).

Killing vectors of static spherically symmetric metrics

Ashfaque H. Bokhari and Asghar Qadir
Mathematics Department, Quaid-i-Azam University, Islamabad, Pakistan

(Received 17 February 1989; accepted for publication 3 January 1990)

An error in a previous theorem [J. Math. Phys. **28**, 1019 (1987); **29**, 525 (1988)] is corrected and the theorem is extended.

It was pointed out by Professor Stephani[1] that there was an error in our earlier claim[2] that the only spherically symmetric static space-times with more than four KV's were the de Sitter, Minkowski, and anti-de Sitter spaces. In fact, the well-known Einstein universe, possessing seven KV's, is a clear counterexample.[3] Here we correct our previous theorem and include a spherically symmetric case of an unusual type in that the area subtended by a solid angle is independent of the radial coordinate. The most general spherically symmetric static metric is[4]

$$ds^2 = e^{\nu(r)}\,dt^2 - e^{\lambda(r)}\,dr^2 - R(r)^2(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{1}$$

There are two possibilities here; either R^2 is a constant function or it is not. In the latter case, a redefinition of variables enables us to replace it by r^2. The former case will be discussed later.

The procedure adopted was to eliminate all possibilities in solving the Killing equations. However, this elimination was not made explicit. The procedure requires that when a separation, or integration, constant occurs all possibilities be enumerated and tabulated (e.g., it being zero, positive, or negative, or possibly possessing some specific numeric value like unity). In doing so, the case of seven KV's was omitted. This includes the Einstein universe and its counterpart with a negative energy density, the anti-Einstein universe. The elimination procedure yields 25 separate cases of which 20 are uninteresting as they only possess the minimal symmetry required, while five have additional symmetries. The metrics and Killing vectors of these five are explicitly known.

In the case that R^2 is replaced by a positive constant again there are generically four KV's. However, there are three cases that have six KV's. They include the Bertotti–Robinson metric.[3,5] One of these cases was not included in Petrov's classification, though a special subcase is.[6] Details of these metrics are available elsewhere.[7]

We have, therefore, the following corrected version of the theorem: Spherically symmetric, static space-times have either 10 KV's (corresponding to de Sitter, Minkowski, and anti-de Sitter metrics), or seven KV's (corresponding to the Einstein and anti-Einstein metrics), or six KV's (incorporating the Bertotti–Robinson and two other metrics), or four KV's (the minimal symmetry).

ACKNOWLEDGMENTS

We are grateful to Professor Stephani for pointing out our earlier omission and to the editors for passing it on to us. We are also indebted to H. Azad and M. Ziad for several useful discussions.

[1] H. Stephani (private communication).
[2] A. H. Bokhari and A. Qadir, J. Math. Phys. **28**, 1019 (1987); **29**, 525 (1988).
[3] D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herit, *Exact Solutions to Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[4] N. Straumann, *General Relativity and Relativistic Astrophysics* (Springer, Berlin, 1984). This depends on "spherical symmetry" being taken to imply that a two-sphere is left invariant under the action of the group SO(3,R). There can be cases in which the killing vectors satisfy an SO(3,R) but where there is no two-sphere left invariant.
[5] We are grateful to an unknown referee for pointing out that these include the Bertotti–Robinson metric.
[6] A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).
[7] A. Qadir and M. Ziad, J. Math. Phys. **29**, 2473 (1988).

 0022-2488/90/061463-01$03.00

Conformal Birkhoff and Taub theorems in higher dimensions

Can-bin Liang
China Center of Advanced Science and Technology (World Laboratory), P.O. Box 8730, Beijing, People's Republic of China and Department of Physics, Beijing Normal University, Beijing, People's Republic of China

(Received 4 August 1989; accepted for publication 17 January 1990)

The result of a recent paper [Kuang and Liang, J. Math. Phys. **29**, 2475 (1988)] is generalized to higher dimensions, i.e., it is proved that any conformally spherically (resp. plane-) symmetric solution to vacuum Einstein equations in higher dimensions must be the generalized Schwarzschild (resp. generalized Taub–Kasner or flat) metric.

I. INTRODUCTION

A recent paper by Kuang and Liang[1] has generalized Birkhoff and Taub theorems to Ricci-flat metrics with conformal symmetry. It has also developed a new method (based on conformal transformation) for proving these two old theorems. It is shown in this paper that this method is also powerful in proving the validity of these two theorems and their conformal generalization in higher dimensions.

An $(N+2)$-dimensional space-time is said to be spherically (resp. plane-) symmetric if its isometry group contains an $N(N+1)/2$-dimensional subgroup isomorphic to $SO(N)$ (resp. the Euclidean group) with spacelike N-spheres (resp. N-planes) as its orbits. The orbits are then constant curvature spaces with Gaussian curvature 1 (resp. 0), and hence admit orthogonal two-surfaces. Therefore, the general form of an $(N+2)$-dimensional spherically (plane-) symmetric metric can be written as

$$dS^2 = E(t,r)\,(-dt^2+dr^2) + G^2(t,r)d\sigma^2, \tag{1}$$

where $E(t,r)$ and $G(t,r)$ are arbitrary functions and

$$d\sigma^2 = d\vartheta_1^2 + (1-\zeta\cos^2\vartheta_1)\,d\vartheta_2^2 + \cdots + \prod_{m=1}^{N-1}(1-\zeta\cos^2\vartheta_m)d\vartheta_N^2, \tag{2}$$

with $\zeta = 1$ for the spherically symmetric case and $\zeta = 0$ for the plane-symmetric case. A conformally spherically (plane-) symmetric metric can then be expressed as

$$d\hat{S}^2 = \Omega_1^2\, dS^2 = \Omega^2[H(t,r)\,(-dt^2+dr^2)+d\sigma^2] \equiv \Omega^2\, dS_H^2,$$

where Ω_1 is an arbitrary function of $t,r,\vartheta_1,...,\vartheta_N$, $\Omega\equiv\Omega_1 G$, and $H(t,r)\equiv E(t,r)/G^2(t,r)$. Our essential task is to find out all such $d\hat{S}^2$ with vanishing Ricci tensor.

II. THE PROOF OF THE GENERALIZED THEOREM

The nonvanishing Christoffel symbols and Ricci tensor components of dS_H^2 are

$$\Gamma^r_{rr} = \Gamma^r_{tt} = \Gamma^t_{rt} = \frac{1}{2}\left(\frac{\partial \ln|H|}{\partial r}\right),$$

$$\Gamma^t_{rr} = \Gamma^t_{tt} = \Gamma^r_{rt} = \frac{1}{2}\left(\frac{\partial \ln|H|}{\partial t}\right),$$

$$\Gamma^i_{ji} \text{ (short for } \Gamma^{\vartheta_i}_{\vartheta_j\vartheta_i}) = \zeta\cot\vartheta_j \quad (j<i),$$

$$\Gamma^j_{ii} = -\zeta\sin\vartheta_j\cos\vartheta_j \prod_{m=j+1}^{i-1}\sin^2\vartheta_m \quad (j<i),$$

$$R_{rr} = -R_{tt} = \frac{1}{2}\left(\frac{\partial^2 \ln|H|}{\partial t^2} - \frac{\partial^2 \ln|H|}{\partial r^2}\right),$$

$$R_{ii} = \zeta(N-1)\prod_{m=1}^{i-1}\sin^2\vartheta_m$$

$$\left(\prod_{m=1}^{0}\sin^2\vartheta_m \text{ is understood to be } 1\right).$$

The conformality between $d\hat{S}^2$ and dS_H^2 and the Ricci flatness of $d\hat{S}^2$ require

$$\Omega^{-1}R_{\mu\nu} + N(\Omega^{-1})_{;\mu\nu} - \Psi g_{\mu\nu} = 0, \tag{3}$$

with

$$\Psi = \Omega^{-3}(\Omega^2)_{;\rho\sigma}g^{\rho\sigma}/2, \tag{4}$$

where $g_{\mu\nu}$ are metric components of dS_H^2. Expression (3) represents $(n+2)(N+3)/2$ equations restricting the unknown function Ω:

$$\frac{\partial^2\Omega^{-1}}{\partial t\,\partial\vartheta_i} = \frac{\partial^2\Omega^{-1}}{\partial r\,\partial\vartheta_i} = 0 \quad (i=1,...,N), \tag{5}$$

$$\frac{\partial^2\Omega^{-1}}{\partial\vartheta_j\,\partial\vartheta_i} - \zeta(\cot\vartheta_j)\frac{\partial\Omega^{-1}}{\partial\vartheta_i} = 0 \quad (j=1,...,N-1, i>j), \tag{6}$$

$$\frac{\partial^2\Omega^{-1}}{\partial t\,\partial r} - \Gamma^r_{rt}\,\Omega^{-1}_{,r} - \Gamma^t_{rt}\,\Omega^{-1}_{,t} = 0, \tag{7}$$

$$\Omega^{-1}R_{rr} + N(\Omega^{-1})_{;rr} - \Psi H = 0, \tag{8}$$

$$\Omega^{-1}R_{tt} + N(\Omega^{-1})_{;tt} + \Psi H = 0, \tag{9}$$

$$\Omega^{-1}R_{ii} + N(\Omega^{-1})_{;ii} - \Psi\prod_{m-1}^{i-1}(1-\zeta\cos^2\vartheta_m) = 0 \quad (i=1,...,N). \tag{10}$$

The same argument of Ref. 1 shows that

$$\Omega^{-1} = L(t,r) + S(\vartheta_1,...,\vartheta_N),$$

and the coordinates $t,r,\vartheta_1,...,\vartheta_N$ can be so chosen that

$$L = L(t), \quad H = \frac{dL}{dt},$$

and hence $d\hat{S}^2 = \Omega^2[(dL/dt)(-dt^2+dr^2)+d\sigma^2]$.

Applications of Eq. (6), respectively, to $j=1$ and $j=2$ result in

$$\frac{\partial S}{\partial\vartheta_i} = e_{12}(\vartheta_3,...,\vartheta_N)(\sin\vartheta_2\sin\vartheta_1)^\zeta \quad (i=3,4,...,N)$$

(with e_{12} an arbitrary function of ϑ_j's except for $j=1,2$),

 0022-2488/90/061464-03$03.00

hence

$$S = f_{12}(\vartheta_3,\dots,\vartheta_N)\,(\sin\vartheta_2 \sin\vartheta_1)^{\zeta} + h_i\,(\vartheta_1,\dots,\vartheta_{i-1},\vartheta_{i+1},\dots,\vartheta_N), \tag{11}$$

where $f_{12} \equiv \int e_{12}\, d\vartheta_i$, and h_i is an arbitrary function of ϑ_j's except for $j = 1$. Application of (6) to $j = 1$, $i = 2$ yields

$$\frac{\partial S}{\partial \vartheta_2} = \gamma_1(\vartheta_2,\dots,\vartheta_N)\,(\sin\vartheta_1)^{\zeta}, \tag{12}$$

with γ_1 an arbitrary function of ϑ_j's except for $j = 1$. It follows from (11) and (12) that

$$S = f_{12}(\sin\vartheta_2 \sin\vartheta_1)^{\zeta} + \rho_{1i}\,(\sin\vartheta_1)^{\zeta} + \sigma_2, \tag{13}$$

with ρ_{1i} an arbitrary function of ϑ_j's except for $j = 1,i$, and σ_2 an arbitrary function of ϑ_j's except for $j = 2$. Applications of Eq. (10), respectively, to $i = 1$ and $i = 2$, taking account of (13), yield

$$\zeta\rho_{1i} + \frac{\partial^2 \rho_{1i}}{\partial \vartheta_2^2} = \frac{\partial^2 \sigma_2}{\partial \vartheta_1^2}(\sin\vartheta_1)^{\zeta} - \zeta\frac{\partial \sigma_2}{\partial \vartheta_1}\cos\vartheta_1.$$

Since this holds for $i = 3,4,\dots,N$, the two sides must be a constant:

$$\frac{\partial^2 \sigma_2}{\partial \vartheta_1^2}\,(\sin\vartheta_1)^{\zeta} - \zeta\frac{\partial \sigma_2}{\partial \vartheta_1}\cos\vartheta_1 = A, \quad A \text{ const.} \tag{14}$$

For $\zeta = 0$, (13) and (14) lead to

$$\frac{\partial^2 S}{\partial \vartheta_1^2} = A. \tag{15a}$$

For $\zeta = 1$, the general solution to (14) is $\sigma_2 = -B_{12}\cos\vartheta_1 - A\sin\vartheta_1 + C_{12}$, where B_{12} and C_{12} are arbitrary functions of ϑ_j's except for $j = 1,2$. Substitution into (13) then yields

$$\frac{\partial^2 S}{\partial \vartheta_1^2} = C_{12} - S. \tag{15b}$$

Combining (15a) and (15b) one gets

$$\frac{\partial^2 S}{\partial \vartheta_1^2} = -\zeta S + a_{12}, \quad a_{12} \equiv \zeta(C_{12} - A) + A. \tag{16}$$

It follows from (10) (with $i = 1$) that

$$\Omega^{-1}\zeta(N-1) + N\frac{\partial^2 S}{\partial \vartheta_1^2} - \Psi = 0. \tag{17}$$

Substitution of (16) into (17) and added by (9) then gives

$$S(\zeta H - R_{tt}) = L\,[\zeta(N-1)H + R_{tt}] + N(L_{;tt} + a_{12}H). \tag{18}$$

Except for the trivial case $\zeta H - R_{tt} = 0$, where the Weyl tensor of dS_H^2 vanishes and hence $d\hat{S}^2$ is flat, (18) can be rewritten as

$$S = \{L\,[\zeta(N-1)H + R_{tt}] + N(L_{;tt} + a_{12}H)\}/(\zeta H - R_{tt}),$$

showing $\partial S/\partial\vartheta_1 = 0$ which, on acount of Eq. (6) with $j = 1$, implies that S is a constant and hence can be absorbed into $L(t)$ to give $\Omega^{-1} = L(t)$. The unknown function $\Omega(t)$ is now restricted only by Eqs. (8), (9), and (17) containing Ψ.

Taking account of (4), Eq. (17) can be rewritten as

$$L\frac{d^2 L}{dt^2} - (N+1)\left(\frac{dL}{dt}\right)^2 - \zeta(N-1)L^2\frac{dL}{dt} = 0. \tag{19}$$

or

$$\frac{dF}{d\Omega} = -\frac{(N-1)(F+\zeta)}{\Omega}, \tag{20}$$

where $F \equiv L^{-2}\, dL/dt = -d\Omega/dt$. Equation (20) can be integrated to yield

$$F = -\zeta + \alpha\Omega^{1-N}, \quad \alpha \text{ const.} \tag{21}$$

It is straightforward to check that (21) also satisfies Eqs. (8) and (9), and consequently is the general solution to the original equations (5)–(10). Therefore, the resultant metric $d\hat{S}^2$ can finally be expressed as follows.

(A) $\zeta = 1$.

$$d\hat{S}^2 = (1 - \alpha\Omega^{1-N})^{-1}\, d\Omega^2 - (1 - \alpha\Omega^{1-N})dr^2 + \Omega^2\, d\sigma^2,$$

which can be cast into the standard form of the generalized Schwarzschild metric[2,3] by setting $T = r$ and $R = \Omega$:

$$d\hat{S}^2 = -(1 - \alpha R^{1-N})dT^2 + (1 - \alpha R^{1-N})^{-1}\, dR^2 + R^2\, d\sigma^2.$$

(B) $\zeta = 0$.

$$d\hat{S}^2 = -\alpha^{-1}\Omega^{N-1}\, d\Omega^2 + \alpha\Omega^{1-N}\, dr^2 + \Omega^2\, d\sigma^2.$$

There are two subcases to be distinguished as follows.

(a) $\alpha < 0$. Setting

$$Z = (-N^2\alpha)^{-N/(1+N)}\Omega^N, T = (-N^{1-N}\alpha)^{1/(1+N)}r,$$
$$X_i = (-N^2\alpha)^{1/(1+N)}\vartheta_i \quad (i = 1,2,\dots,N),$$

one obtains

$$d\hat{S}^2 = Z^{(1-N)/N}(-dT^2 + dZ^2) + Z^{2/N}(dX_1^2 + \cdots + dX_N^2).$$

This will be referred to as the $(N+2)$-dimensional Taub metric.

(b) $\alpha > 0$. Setting

$$T = (N^2\alpha)^{-N/(1+N)}\Omega^N, \quad Z = (N^{1-N}\alpha)^{1/(1+N)}r,$$
$$X_i = (N^2\alpha)^{1/(1+N)}\vartheta_i \quad (i = 1,2,\cdots,N),$$

one obtains

$$d\hat{S}^2 = T^{(1-N)/N}(-dT^2 + dZ^2) + T^{2/N}(dX_1^2 + \cdots + dX_N^2).$$

This will be referred to as the $(N+1)$-dimensional Kasner metric. Therefore, we have the following theorem.

Theorem: Any conformally spherically (plane-) symmetric solution to the vacuum Einstein equations in $(N+2)$ $(N \geqslant 2)$ dimensions must be the generalized Schwarzschild (Taub–Kasner or flat) metric.

This theorem can be viewed as the generalization of the Birkhoff and Taub theorems in two different senses: (1) it is a generalization from four dimensions to $(N+2)$ dimensions; (2) it is a generalization from spherical (plane) symmetry to conformally spherical (plane) symmetry. Note that the conformal method developed in Ref. 1 has many advantages even for the first sense of generalization.

ACKNOWLEDGMENTS

I would like to thank Professor Kuang zhi-quan for his careful reading of the manuscript and his comments on it.

This work has been supported by the Natural Science Foundation of China.

[1]Z. Kuang and C. Liang, J. Math. Phys. **29**, 2475 (1988).
[2]F. R. Tangherlini, Nuovo Cimento **77**, 636 (1963).
[3]R. C. Myers and M. J. Perry, Ann. Phys. **172**, 304 (1986).

Infrared and vacuum structure in two-dimensional local quantum field theory models. The massless scalar field

G. Morchio
Dipartimento di Fisica dell'Università, Pisa, Italy
D. Pierotti and F. Strocchi
International School for Advanced Studies, Trieste, Italy

(Received 25 September 1988; accepted for publication 31 January 1990)

A systematic and rigorous treatment of the massless scalar field in two dimensions is presented by carefully taking into account the maximal (Krein) state space associated to the Wightman functions. This allows a simple and rigorous answer to controversial statements appearing in the literature about the uniqueness of the translation-invariant state, the construction of translation-invariant operators (infrared operators), the scale and special conformal transformations of the fields, the construction of the dual field and the breaking of the Lorentz transformations, and, more generally, the status of symmetry breaking in the model.

I. INTRODUCTION

The massless scalar field in two dimensions is the simplest example of a nonpositive Wightman theory (positivity is violated by the two-point function) and also of a theory invariant under gauge transformations. Interest in the model would be rather modest if it did not exhibit some of the typical structural problems of indefinite metric quantum field theories. The lack of sufficient appreciation of such delicate points is at the origin of the contradictory and often incorrect conclusions drawn about this (relatively simple!) model. We list some of the controversial points.

(1) In most of the treatments (see, however, Ref. 1, which follows Ref. 2, it is asserted that the theory has only one translation-invariant state (namely, the Wightman vacuum); this, however, is incompatible with the existence of translation-invariant "field" operators like $\int dx\, \partial_0\varphi(x)$ and $\int dx\, \partial_0\widetilde{\varphi}(x)$ (which do not leave the vacuum invariant) widely used in the literature.[3]

(2) It has been stated that the scale (and special conformal) transformations of the fields require the *ad hoc* introduction of translation-invariant "auxiliary annihilation and creation operators" a, a^* (see Ref. 4); however, the status of these operators is rather unclear and, in fact, incompatible with the uniqueness of the translation-invariant state (this problem has become particularly interesting in relation to two-dimensional conformal models).

(3) The construction of the dual field $\widetilde{\varphi}$ in terms of the field φ involves a very delicate procedure since $\widetilde{\varphi}$ is nonlocal with respect to φ; the lack of attention to this problem is at the origin of the (incorrect) statement that the Lorentz transformations are broken by the introduction of $\widetilde{\varphi}$ (see, e.g., Ref. 5).

(4) More generally, the discussion of the spontaneous breaking of symmetries in this model is far from settled in the existing literature.

The aim of the present paper is to present a systematic and rigorous treatment of the model, which will provide a simple answer to the above controversial points.

As we will see, the origin of the controversial statements is that implicitly the various treatments make reference to different spaces of states associated to the Wightman functions. In Wightman theories satisfying positivity the space of states is obtained as the closure of the local states with respect to the Hilbert topology defined by the Wightman functions. In the case of indefinite metric theories, the construction of the space of states associated to the Wightman functions is not automatic and, in general, not unique, and different representations of the field algebra may arise corresponding to different Hilbert closures of the local states, namely, to different infrared properties of the states (see Refs. 2 and 6). This problem is not overcome (nor solved) by the algebraic approach to indefinite metric theories (see Ref. 7, and references therein) since the GNS construction only provides a vector space without a Hilbert structure. [In general, the strategy of identifying a positive space of states by applying nonlocal gauge-invariant operators to the vacuum does not uniquely identify a space of physical states, as is particularly clear in the quantum electrodynamics (QED) case.[6,8]]

A further motivation for the present paper is that a rigorous treatment of such a model of indefinite metric quantum field theory may shed light on general mathematical structures arising in realistic theories in local gauges. In particular, we believe that general features such as the essential uniqueness of the vacuum, the existence of infinitely delocalized (translational-invariant) operators and their role in symmetry breaking, the mechanism of symmetry realization in the physical space, and the identification of nonlocal physical states through a subsidiary condition,[9] which can be simply seen and controlled in this model, should also arise in realistic theories when formulated in local and covariant gauges. For the general structure of the Wightman theory as well as for some standard notation, we refer to Ref. 10.

At first, one might think to obtain a state space associated to the Wightman functions as a "closure" of the local states with respect to the *Wightman topology* $\tau_{\mathscr{W}}$ uniquely defined by the following family of seminorms:

$$p_g(\Psi_{\underline{f}}) = |\langle \underline{g}\underline{f}\rangle| ,$$

where $\underline{f}$ and $\underline{g}$ belong to the Borchers algebra of the test functions and $\langle\ ,\ \rangle$ is the inner product defined by the Wightman functions. In this topology, however, the inner product $\langle\ ,\ \rangle$ is not jointly continuous and therefore it cannot be extended by continuity to all the states obtained through $\tau_{\mathscr{W}}$ limits. One needs, therefore, a stronger topology than the Wightman topology. The most natural requirement is to look for a topology induced by a Hilbert scalar product,[2,11] and to consider the Hilbert closure $\overline{\mathscr{D}}_0$ (with $\mathscr{D}_0$ the vector space of the local states), where the indefinite inner product can be extended by joint continuity.

The most interesting among the possible Hilbert structures are those that associate to the Wightman functions a maximal space K (see Refs. 2 and 11), namely, such that one cannot find a Hilbert structure (majorizing the same Wightman functions) whose space of states properly contains K. It can be shown[2,11] that in a maximal[12] Hilbert space one can always define the positive scalar product $(\ ,\ )$ in such a way that the *metric operator* η defined by $\langle\ ,\ \rangle = (\ ,\eta)$, satisfies $\eta^2 = 1$ (Krein space[13]). For a general discussion, see Refs. 2 and 11.

II. THE KREIN REALIZATION

A. The Krein topology. The one-particle space

To get a (maximal) Hilbert space realization of the massless scalar field, it is enough to define a seminorm p_1 on $\mathscr{S}(\mathbb{R}^2)$, majorizing the inner product defined by the two-point function,[14]

$$\langle f,g\rangle \equiv \int d^2x\, d^2y\, \bar f(x) g(y) \mathscr{W}(x-y)\,,$$

$$\mathscr{W}(x) = -(1/4\pi)\log(-x^2 + i\epsilon x_0)\,.$$

We want to choose p_1 in such a way that the corresponding closure $\overline{\mathscr{S}}$ is maximal.

The solution of this problem is essentially unique;[2,11] we decompose $\mathscr{S}(\mathbb{R}^2)$ as

$$\mathscr{S} = \mathscr{S}_0 + V, \quad \mathscr{S}_0 \equiv \{f\in\mathscr{S},\ \hat f(0) = 0\}\,,$$

where $\hat{}$ denotes the Fourier transform and V is a one-dimensional space generated by a real symmetric function $\chi\in\mathscr{S}$ with $\hat\chi(0) = 1$ and $\langle\chi,\chi\rangle = 0$.

Thus we can write an arbitrary $f\in\mathscr{S}(\mathbb{R}^2)$ as $\hat f(p) = \hat f(0)\hat\chi(p) + \hat f_0(p)$, with $f_0\in\mathscr{S}_0$. On $\mathscr{S}_0$, the indefinite inner product $\langle\ ,\ \rangle$ is positive. The Krein seminorm is then given on $\mathscr{S}$ by[2,11]

$$p_1(f)^2 \equiv p_K(f)^2 = \langle f_0, f_0\rangle + |\langle f,\chi\rangle|^2 + |\hat f(0)|^2\,. \tag{2.1}$$

By writing the indefinite inner product in the form

$$\langle f,g\rangle = \langle f_0, g_0\rangle + \bar{\hat f}(0)\langle\chi,g\rangle + \hat g(0)\langle f,\chi\rangle\,,$$

one easily verifies that

$$|\langle f,g\rangle| \leqslant p_K(f) p_K(g)\,.$$

We denote by $K^{(1)}$ the completion $\overline{\mathscr{S}}^K$ of $\mathscr{S}$, in the Hilbert topology defined by (2.1). The indefinite product is extended by continuity and, since $f\to\hat f(0)$ is continuous in the topology (2.1), the Hilbert scalar product is given on $K^{(1)}$ by

$$(f,g)_K = \langle f_0, g_0\rangle + \langle f,\chi\rangle\langle\chi,g\rangle + \bar{\hat f}(0)\hat g(0)\,. \tag{2.2}$$

In the following we will often omit the index K.

Lemma 2.1: The linear functional on $K^{(1)}$ defined by

$$F_\chi(f) \equiv \langle\chi,f\rangle\,,$$

has norm equal to 1, and therefore it defines a normalized element v_0 of $K^{(1)}$, such that $(v_0,f)_K = \langle\chi,f\rangle$. Furthermore, $v_0\in\overline{\mathscr{S}}_0{}^K$ and, $\forall f\in\mathscr{S}$,

$$\langle v_0, f\rangle = \hat f(0)\,. \tag{2.3}$$

Proof: Obviously, we have

$$|\langle\chi,f\rangle|/\|f\|_K \leqslant 1, \quad \forall f\in\mathscr{S}\,,$$

and therefore $\|v_0\|\leqslant 1$. We shall prove that the equality actually holds. Let us define the sequence

$$\hat f_n^\chi(p) = \theta_n(p_0)\hat\chi(p)\,, \tag{2.4}$$

where $\theta_n(p_0) = \vartheta(np_0)$, $\vartheta\in\mathscr{C}^\infty$, $\vartheta(p) = 1$ for $p\geqslant 1$, $\vartheta(p) = 0$ for $p\leqslant 0$, $0\leqslant\vartheta\leqslant 1$. From (2.2), we have

$$\frac{|\langle\chi, f_n^\chi\rangle|^2}{\|f_n^\chi\|_K^2} = \left(1 + \frac{\langle f_n^\chi, f_n^\chi\rangle}{|\langle\chi, f_n^\chi\rangle|^2}\right)^{-1}\,.$$

On the other hand, we obtain

$$\pi\frac{\langle f_n^\chi, f_n^\chi\rangle}{|\langle\chi, f_n^\chi\rangle|^2} = \frac{\int (dp/|p|)|\vartheta(n|p|)\hat\chi(p,|p|)|^2}{|\int (dp/|p|)\vartheta(n|p|)\hat\chi(p,|p|)^2|^2} \leqslant \left|\int\frac{dp}{|p|}\,\vartheta(n|p|)\hat\chi(p,|p|)^2\right|^{-1}\,.$$

Since this last term decreases as $(\log n)^{-1}$ for $n\to\infty$, we get

$$\sup_n \frac{|\langle\chi, f_n^\chi\rangle|^2}{\|f_n^\chi\|_K^2} = 1\,,$$

which in turn implies

$$\|v_0\|_K = 1\,.$$

Moreover, setting

$$v_n \equiv (\langle\chi, f_n^\chi\rangle)^{-1} f_n^\chi\,,$$

we have

$$(v_n, v_0)_K = \langle v_n, \chi\rangle = 1\,,$$

$$(v_n, v_n)_K \underset{n\to\infty}{\to} 1, \quad \langle v_n, v_n\rangle \underset{n\to\infty}{\to} 0\,.$$

Hence the v_n converge strongly to v_0 in $K^{(1)}$ and converge to zero in $L^{(2)}(dp_1/|p_1|, C_+)$, $C_+ \equiv \{p\in\mathbb{R}^2,\ |p_1| = p_0\}$. In particular, $\langle v_n, \mathscr{S}_0\rangle \to 0$ and therefore $\forall g\in\mathscr{S}$, $\langle v_n, g\rangle \to \hat g(0)$.

Proposition 2.2: $K^{(1)}$ is a Krein space, in fact, a Pontryagin space with one negative dimension:

$$K^{(1)} = L^2\left(\frac{dp_1}{|p_1|}, C_+\right)\oplus V_0\oplus V, \quad V_0 \equiv \{\lambda v_0,\ \lambda\in\mathbb{C}\}\,.$$

The metric operator $\eta^{(1)}$ defined by $\langle\cdot,\cdot\rangle = (\cdot,\eta^{(1)}\cdot)$ is given by $\eta^{(1)} = 1$ on $L^{(2)}$, $\eta^{(1)} v_0 = \chi$, $\eta^{(1)}\chi = v_0$.

Proof: Clearly, from (2.2) and $\langle\chi,\chi\rangle = 0$, one has $(\chi, \mathscr{S}_0) = 0$ and therefore $K^{(1)} = \overline{\mathscr{S}}_0^K \oplus V$. Moreover, $\forall f,g\in\mathscr{S}_0$,

$$(f,g)_K = \langle f,g\rangle + (f, v_0)(v_0, g) \tag{2.5}$$

and

$$\langle f,g\rangle = \pi\int_{C_+}\frac{dp_1}{|p_1|}\bar f g\,,$$

so that, if $f_n\in\mathscr{S}_0$ and f_n converge in the $\|\ \|_K$ norm, also

(f_n, v_0) converge and the f_n converge strongly in $L^2(dp_1/|p_1|, C_+)$. Since the v_n converge to zero in L^2, $\forall f \in \mathscr{S}_0$, $f_n = f - (v_0 f)(v_n, v_0)^{-1} v_n$ converge to f in L^2, i.e., the subspace of $\mathscr{S}_0$ orthogonal to v_0 is dense in L^2. Cauchy sequences f_n of elements of $\mathscr{S}_0$ in the $\| \, \|_K$ norm can therefore be identified with pairs (f, λ) with $f \in L^2(dp_1/|p_1|, C_+)$, $\lambda = \lim_n (f_n, v_0) \in \mathbb{C}$. Hence $\bar{\mathscr{S}}_0^K = L^2(dp_1/|p_1|, C_+) \oplus V_0$.

From Eq. (2.5) it follows that $\eta^{(1)} = 1$ on L^2; furthermore, $(v_0 f)_K = \langle \chi f \rangle$ and $\langle v_0 f \rangle = \tilde{f}(0) = (\chi f)$ [by Lemma 2.1 and Eq. (2.2)] imply $\eta^{(1)} \chi = v_0$ and $\eta^{(1)} v_0 = \chi$, respectively.

Proposition 2.3: The representation $U(a, \Lambda)$ of the Poincaré group has a unique continuous extension from $\mathscr{S}$ to $K^{(1)}$ and

$$U(a, \Lambda) v_0 = v_0 . \tag{2.6}$$

Proof: By using the above results we can write the Krein scalar product in the form, $\forall f, g \in \mathscr{S}$,

$$(f, g) = \langle f, g \rangle + (\overline{\hat{f}(0) - \langle f, \chi \rangle})(\hat{g}(0) - \langle \chi, g \rangle) \tag{2.7}$$

and obtain

$$\|Uf\|^2 = \|f\|^2 + |\langle Uf, \chi \rangle|^2 - |\langle f, \chi \rangle|^2$$
$$\leqslant \|f\|^2 + \|f\| \, \|U^{-1}\chi - \chi\| (\|Uf\| + \|f\|) ,$$

which implies

$$\|Uf\| \leqslant \|f\| (1 + \|U^{-1}\chi - \chi\|) .$$

Thus the unique extension of $U(a, \Lambda)$ by continuity yields a representation of the Poincaré group that preserves the inner product $\langle \cdot, \cdot \rangle$. Equation (2.6) then follows from Eq. (2.3).

From the explicit representation of $K^{(1)}$ given in Proposition 2.2 it follows that there are no other Poincaré-invariant vectors in $K^{(1)}$.

B. Fock–Krein space

Given a set of Wightman functions $\mathscr{W}_n$ whose truncated parts vanish for $n > 2$ ($\mathscr{W}_1 = 0$), one can define the positive and negative energy parts $\varphi_\pm(f)$:

$$\varphi_+(f)\varphi(f_1)\cdots\varphi(f_n)\Psi_0$$
$$= \varphi(f_1)\cdots\varphi(f_n)\varphi(f)\Psi_0$$
$$- \sum_j \langle \varphi(f_j)\varphi(f) \rangle_0 \varphi(f_1)\cdots\widehat{\varphi(f_j)}\cdots\varphi(f_n)\Psi_0$$

(where the symbol $\frown$ means the depletion of the variable) and $\varphi_-(f) = \varphi(f) - \varphi_+(f)$. Then, the set of vectors

$$\Psi^{(n)}_{f_1, \ldots, f_n} = (n!)^{-1/2} \varphi_+(f_1)\cdots\varphi_+(f_n)\Psi_0$$

generate $\mathscr{D}_0$, and the indefinite inner product

$$\langle \Psi^{(n)}_{f_1, \ldots, f_n}, \Psi^{(m)}_{g_1, \ldots, g_m} \rangle = \delta_{n,m} (n!)^{-1} \sum_P \langle f_1, g_{i_1} \rangle \cdots \langle f_n, g_{i_n} \rangle \tag{2.8}$$

vanishes if $n \neq m$ [the sum in the above expression is over all permutations $(i_1 \cdots i_n)$ of the n indices]. Hence, given a Hilbert seminorm p_1 on $\mathscr{S}$ majorizing the two-point function, the set of seminorms on $\mathscr{D}_0$,

$$p_n(\Psi^{(n)}_{f_1, \ldots, f_n}) = \left((n!)^{-1} \sum_P (f_1, f_{i_1}) \cdots (f_n, f_{i_n}) \right)^{1/2} ,$$

where (,) denotes the Hilbert product that defines p_1, defines a majorant Hilbert topology with a Hilbert scalar product

$$(\Psi^{(n)}_{f_1, \ldots, f_n}, \Psi^{(m)}_{g_1, \ldots, g_m}) = \delta_{n,m} (n!)^{-1} \sum_P (f_1, g_{i_1}) \cdots (f_n, g_{i_n}) , \tag{2.9}$$

preserving the factorization property (2.8).

We denote by K the completion of $\mathscr{D}_0$ with respect to the topology introduced above. Clearly, K has the form of a direct sum of symmetric tensor products:

$$K = \sum_n (\otimes_s^n K^{(1)}) \equiv \sum_n K^{(n)} .$$

Then, from Eqs. (2.8) and (2.9) we have the following proposition.

Proposition 2.4: The space K is a Krein space with the metric operator η given by

$$\eta K^{(n)} = \otimes_s^n \eta^{(1)} K^{(n)} ,$$

satisfying

$$\eta^2 = 1 .$$

The representation $U(a, \Lambda)$ of the Poincaré group extends from $\mathscr{D}_0$ to the dense domain of finite particle states of K; on each subspace $K^{(n)}$, the operators $U(a, \Lambda)$ are bounded operators. It should be stressed, however, that the norm of $U(a, \Lambda)$, restricted to $K^{(1)}$, is larger than 1 and therefore $U(a, \Lambda)$ are unbounded operators in K.

An important feature of the Krein space is given by the following proposition.

Proposition 2.5: The space K contains an infinite-dimensional subspace V_0 of vectors invariant under the Poincaré transformations. However, the vacuum is still essentially unique (i.e., any strictly positive subspace of invariant vectors is one dimensional[2]).

Clearly, all the vectors

$$\Phi_n^{(o)} = \otimes_s^n v_0$$

are Poincaré invariant and generate a Poincaré-invariant space V_0. Since $U(a, \Lambda) K^{(n)} \subset K^{(n)}$, any invariant vector Ψ must have an invariant component Ψ_n in each $K^{(n)}$; since v_0 is the unique invariant vector in $K^{(1)}$, $\Psi_n = \lambda \Psi_n^{(0)}$ and $\Psi \in V_0$.

The essential uniqueness of the vacuum follows from the fact that $\langle \Phi_n^{(0)}, \Phi_n^{(0)} \rangle = 0$, $\forall n$, as a consequence of $\langle v_0, v_0 \rangle = \langle \chi, \chi \rangle = 0$ (vectors of zero η norm). □

It is important to stress that in order to majorize the Wightman functions and therefore *dominate* their infrared singularities, one needs to introduce *infinitely delocalized* and therefore *infinitely extended states*. In particular, $K^{(1)}$ and therefore K are not function spaces. It should be clear from the above discussion that this phenomenon is directly related to the infrared singularities of the Wightman functions (which are not measures in momentum space), and therfore it is expected to occur also in realistic four-dimensional models exhibiting infrared singularities of the confining type.[2] One of the points of the above discussion is, in fact, that of isolating general structure properties independently of the exact solubility of the model.

Finally, the representation of the fields on $\mathscr{D}_0$ is given as in the standard case[15] and the field $\varphi(f)$ transforms covariantly under $U(a,\Lambda)$. The above representation of the field operators gives the following estimate, $\forall\Psi\in K^{(m)}$:

$$\|\varphi(f)\Psi\|_K \leqslant (m+1)^{1/2}|||f|||\,\|\Psi\|_K,$$

where

$$|||f||| = \|f^+\|_K + \|f^-\|_K,$$
$$\hat{f}^+(p) = \hat{f}(p)|_{C_+}, \quad \hat{f}^-(p) = \hat{f}(-p)|C_+. \qquad (2.10)$$

This allows one to construct a class of analytic functions of the field, in particular, the exponential function $\exp[i\lambda\varphi(f)]$, for any complex λ and $f\in\mathscr{S}^{|||\,|||}$.

C. Extension of the field algebra. Infrared operators

As we have seen in the previous section, the (minimal) Hilbert topology majorizing the infrared singularities of the Wightman functions leads to the existence of infinitely delocalized states and it is reasonable to ask what is the counterpart at the level of field algebras. As we will see, in contrast to the standard (positive metric) case, the strong closure of the field algebra contains *infinitely delocalized field operators* that are translational invariant. This feature is related to the minimality of the Hilbert topology (Krein topology) and is not shared by other realizations (with η^{-1} unbounded).

These operators appeared in the literature with different motivation and in different contexts.[3,4,16–18] They appear as *ad hoc* ingredients not related to the infrared structures of the quantum field theory. Their mathematical status is unclear since they map $\mathscr{D}_0$ into vectors that exist only in the maximal, i.e., Krein, completion of $\mathscr{D}_0$.[2] The aim of the present treatment is to clarify their origin and to show that they are intrinsic features of the space of states associated to quantum field theories with nonpositive infrared singularities.

By the estimate before Eq. (2.10), the field operator $\varphi(f)$, $f\in\mathscr{S}$, has an extension, by strong continuity on the dense subspace $\cup_n K^{(n)}$, to $f\in\mathscr{S}^{|||\,|||}$, where $|||\;|||$ is defined in Eq. (2.10). Since the operator product is continuous with respect to the strong topology (recall that the field operators are bounded operators from $K^{(n)}$ to $K^{(n\pm 1)}$, and therefore we actually deal with the norm operator topology), we obtain a well defined extension $\mathscr{A}_{\rm ext}$ of the field algebra $\mathscr{A}$.

Clearly, for $\hat{f}$ vanishing in the neighborhood of the origin the splitting $f=f^+ + f^-$ is well defined and in analogy with Proposition 2.2,

$$\mathscr{S}_0^{|||\,|||} = L^2\left(\frac{dp}{|p|}, C_+\cup C_-\right)\oplus\{V_0^+\}\oplus\{V_0^-\},$$
$$\{V_0^\pm\} = \{\lambda v_0^\pm,\ \lambda\in\mathbb{C}\},$$

and $v_0^\pm$ are limits of the sequences $v_n^\pm(p) = v_n(\pm p)$, with v_n defined as in Lemma 2.1.

Proposition 2.6: The algebra $\mathscr{A}_{\rm ext}$ contains the following infinitely delocalized operators, briefly *infrared operators*:

$$\varphi(v_0) = \varphi_+(v_0) + \varphi_-(v_0),$$

$$Q = i\pi[\varphi_+(v_0) - \varphi_-(v_0)],$$

with $\varphi_\pm(v_0)\equiv\varphi(v_0^\pm)$. They are η symmetric (actually essentially η adjoint), invariant under Poincaré transformations, and satisfy the following commutation relations, $\forall f\in\mathscr{S}$:

$$[Q,\varphi(f)] = -2\pi i\hat{f}(0), \qquad (2.11)$$

$$[\varphi(v_0),\varphi(f)] = 0, \quad [Q,\varphi(v_0)] = 0. \qquad (2.12)$$

Proof: The representation of the Poincaré group on the test functions of $\mathscr{S}$ is strongly continuous in the topology generated by $|||\;|||$, and $v_0^\pm$ are Poincaré invariant as in Proposition 2.3. The second of Eqs. (2.12) follows from the first by strong continuity.

It is worthwhile to remark that the operators $\varphi_\pm(\chi)$ defined in Sec. II B do not belong to the strong closure $\mathscr{A}_{\rm ext}$ of the field algebra, since, by Eqs. (2.12) and the strong continuity of the operator product, $\mathscr{A}_{\rm ext}$ commutes with $\varphi(v_0)$, whereas

$$[\varphi(v_0),\varphi_\pm(\chi)] = \text{s-lim}\langle[\varphi(v_n),\varphi_\pm(\chi)]\rangle_0 = \pm 1.$$

(Note that $\chi_\pm$ do not belong to $\mathscr{S}^{|||\,|||}$; only their sum does.)

D. Physical states. Subsidiary conditions

To identify the physical states, we will characterize the one-particle physical space $K'^{(1)}$.

Proposition 2.7: The only nontrivial maximal subspace $K'^{(1)}$ of $K^{(1)}$, which is invariant under Poincaré transformations and non-negative, is

$$K'^{(1)} = \mathscr{S}_0^K = L^2\left(\frac{dp}{|p|}, C_+\right)\oplus\{V_0\}.$$

Furthermore, the subspace $K''^{(1)}\subset K'^{(1)}$ of "null" vectors coincides with the subspace generated by v_0, and therefore the one-particle physical Hilbert space $\mathscr{H}^{(1)}_{\rm phys}$ is $L^2(dp/|p|, C_+)$.

As usual (see, e.g., the free QED case), the subspace $K'\subset K$ corresponding to states with physical interpretation is defined by taking symmetric tensor products of the corresponding one-particle subspaces. The physical Hilbert space $\mathscr{H}_{\rm phys}$ is then obtained by closing K'/K'' with respect to the topology induced by the semidefinite inner product $\langle\ ,\ \rangle$.

Actually, in this case, K'/K'' is already complete and therefore

$$\mathscr{H}_{\rm phys} = \frac{K'}{K''} = \sum_n L_s^2\left(\frac{dp_1}{|p_1|}\cdots\frac{dp_n}{|p_n|}, C_+\times\cdots\times C_+\right).$$

As usual, it is convenient to characterize the physical vectors by an operator equation, or subsidiary condition, like the Gupta–Bleuler condition in QED.

Theorem 2.8: A dense set of vectors of K' can be characterized as the solution of the following subsidiary condition:

$$\varphi_-(v_0)\Psi = 0. \qquad (2.13)$$

Proof: Clearly, for any Ψ of the form $\varphi(f_1)\cdots\varphi(f_n)\Psi_0$, with $f_j\in\mathscr{S}_0^K$, we have $\varphi_-(v_0)\Psi=0$, as a consequence of $[\varphi_-(v_0),\varphi(f_j)]=0$. Conversely, if $\Psi\in\mathscr{D}_0$ and $\varphi_-(v_0)\Psi=0$, then $(\varphi_-(v_0)\Psi)^{(n)}=0$, i.e., $\langle\Psi^{n+1},\varphi_+(v_0)\Phi^n\rangle = 0$, $\forall\Phi^n\in\mathscr{S}(\mathbb{R}^{2n})$, so that by using Proposition 2.2 one has that Ψ belongs to the closure of $\otimes_s^n(\mathscr{S}_0^K)$.

Remark: For any $f_0\in\mathscr{S}_0$, $\varphi(f_0)$ is a quotientable field operator,[27] but $\varphi(\chi)$ is not; the infrared operators Q and

$\varphi(v_0)$ are quotientable operators, and their images on $\mathscr{H}_{\rm phys}$ vanish.

The estimate (2.10) easily passes to the strong closures $\mathscr{P}^{\|\|\ \|\|}$, and therefore $\mathscr{D}_0$ is a set of analytic vectors also for the infrared operators. As we will see in Sec. IV the operator Q can be interpreted as the generator of gauge transformations, and therefore it has the meaning of a "charge." The result is that such a charge is *unbroken* in K, since Q and its exponentials are well defined on $\mathscr{D}_0$, but it is *bleached*, namely, all the physical states have zero charge.

III. THE DUAL FIELD. LEFT AND RIGHT MOVERS

A. Gauss' law and charged field. The dual field

Much of the usefulness of the massless scalar scalar field for the discussion of two-dimensional field theory models relies on the introduction of the so-called dual field $\tilde{\varphi}$.[3,15,16] It is not necessary to recall the connection between the dual field and the fermion bosonization, the discussion of the chiral symmetry and its breaking, and the operator solution of the Schwinger model.[5,19–25] However, apart from the common motivations, the introduction of $\tilde{\varphi}$ in the literature appears with very different mathematical justifications and, actually, in our opinion, a careful analysis on how it could be introduced in terms of the massless scalar field φ seems to be lacking. As we will see, the arbitrariness involved in the definition of $\tilde{\varphi}$ is at the root of the peculiar results appearing in the literature.

The aim of this section is to provide a careful mathematical analysis of this problem, and to show that complete control on the corresponding field theory can be achieved only when the fields φ and $\tilde{\varphi}$ are realized as operator-valued distributions in a Hilbert space.

The reason why this problem is more delicate than it appears is that it amounts to the construction of a field charged with respect to a charge that obeys a *local Gauss law*.[6,26–28]

In fact, the differential equation of $\tilde{\varphi}$ in term of φ can be written as

$$\partial_\mu \tilde{\varphi} = \epsilon_{\mu\nu}\, \partial^\nu \varphi \equiv \partial^\nu F_{\mu\nu}\,, \tag{3.1}$$

where $F_{\mu\nu} = \epsilon_{\mu\nu}\varphi = -F_{\nu\mu}$. The conserved current

$$\tilde{j}_\mu = \partial_\mu \tilde{\varphi}$$

then defines a charge

$$\tilde{Q}_R = \int dx_1 f_R \tilde{j}_0,\quad f_R(x) = f\left(\frac{x}{R}\right),\quad f\in\mathscr{D}(\mathbb{R})\,,$$

which obeys a local Gauss law, and $\tilde{\varphi}$ is not chargeless with respect to it.

It is well known[27,28] that a field carrying a charge that obeys a local Gauss law [Eq. (3.1)] cannot be local with respect to the antisymmetric tensor $F_{\mu\nu}$ which enters in the Gauss law. We are then facing one of the fundamental problems of gauge quantum field theory, namely, the conflict between locality and the local Gauss law, and consequently the nontrivial construction of nonlocal charged fields starting from local fields.[6,26] Here, we are given the local field φ and we want to construct the field $\tilde{\varphi}$, which cannot be local with respect to φ. More precisely, we have to find an operator-valued distribution $\tilde{\varphi}$ that solves Eq. (2.1), namely,

$$\partial_0 \tilde{\varphi} = -\partial_1 \varphi,\quad \partial_1 \tilde{\varphi} = -\partial_0 \varphi\,. \tag{3.2}$$

The compatibility of these two equations requires

$$\Box \varphi = \Box \tilde{\varphi} = 0\,. \tag{3.3}$$

Clearly, on the test functions $f\in\mathscr{S}_0$, $\tilde{\varphi}$ is simply recovered from φ, since $\tilde{\varphi}(\partial_\mu f^\mu) = \varphi(\epsilon_{\mu\nu}\,\partial^\nu f^\mu)$ and the functions $\partial^\mu f_\mu$ exhaust $\mathscr{S}_0$.

The nontrivial problem is to extend $\tilde{\varphi}$ from $\mathscr{S}_0$ to $\mathscr{S}$. This will be done by first determining the Wightman functions of $\tilde{\varphi}$ and then by reconstructing the field out of them. Finally, the realization of $\tilde{\varphi}$ as an operator-valued distribution on a Hilbert space will require an enlargement of the Hilbert space structure needed for the realization of φ.

B. The Wightman functions of the dual field. Left and right movers

To determine the Wightman functions of $\tilde{\varphi}(f)$, $f\in\mathscr{S}$, we start by noticing that, $\forall g\in\mathscr{S}_0$, $\tilde{\varphi}(g) = \varphi(\tilde{g})$, with

$$\tilde{g}|_{C_+\cup C_-} = (\epsilon(p_0)\epsilon(p_1)g)|_{C_+\cup C_-}\,. \tag{3.4}$$

Furthermore, the two-point function of φ has support in C_+, and, when restricted to $\mathscr{S}_0$, can be multiplied by $\epsilon(p_0)\epsilon(p_1)$.

Thus, *on* $\mathscr{S}_0$,

$$\begin{aligned}
\langle \varphi(x)\varphi(y)\rangle_0 &= \langle \tilde{\varphi}(x)\tilde{\varphi}(y)\rangle_0 \\
&= \mathscr{W}_R(x-y) + \mathscr{W}_L(x-y) \\
&\equiv \mathscr{W}(x-y)\,, \\
\langle \varphi(x)\tilde{\varphi}(y)\rangle_0 &= \langle \tilde{\varphi}(x)\varphi(y)\rangle_0 \\
&= \mathscr{W}_R(x-y) - \mathscr{W}_L(x-y) \\
&\equiv \tilde{\mathscr{W}}(x-y)\,,
\end{aligned}$$

where $\mathscr{W}_R$ and $\mathscr{W}_L$ are given by the Fourier (anti) transformation of the distributions

$$\hat{\mathscr{W}}_R(p) = (1/p_-)_+\delta(p_+),\quad \hat{\mathscr{W}}_L(p) = (1/p_+)_+\delta(p_-),$$
$$p_\pm \equiv p_0 \pm p_1\,, \tag{3.5}$$

where $(1/p)_+ = (d/dp)(\theta(p)\log|p|)$.

To define the field $\tilde{\varphi}(f)$, $f\in\mathscr{S}$, we have to characterize the extensions of the above distributions from $\mathscr{S}_0(\mathbb{R}^2)$ to $\mathscr{S}(\mathbb{R}^2)$, satisfying $\partial_\mu \tilde{\mathscr{W}} = \epsilon_{\mu\nu}\,\partial^\nu \mathscr{W}$ [Eq. (3.1)].

To keep the symmetry between φ and $\tilde{\varphi}$ as close as possible we will choose $\langle \tilde{\varphi}\tilde{\varphi}\rangle = \langle \varphi\varphi\rangle$ on $\mathscr{S}$ and require $\langle \tilde{\varphi}\rangle = 0$. The only arbitrariness that remains[29] is an additive constant in $\tilde{\mathscr{W}}$; by Proposition 2.6 and Lemma 2.1, this freedom corresponds to the redefinition of $\tilde{\varphi}$ by $\tilde{\varphi}(f) \to \varphi(f) + a\tilde{f}(0)\varphi(v_0)$. Without loss of generality, we may choose the extension of $\langle \varphi\tilde{\varphi}\rangle$ from $\mathscr{S}_0$ to $\mathscr{S}$, given by Eq. (3.5).

The next problem is then how φ and $\tilde{\varphi}$ can be represented as operators on a suitable vector space. To this purpose, it is convenient to introduce an enlarged Borchers algebra $\mathscr{B}_{\rm ext}$, whose elements are the pairs $\underline{F} = (\underline{f}, \underline{f}')$, and equip it with the indefinite inner product

$$\langle \underline{F}, \underline{G}\rangle = \underline{W}(\underline{F}^* \times \underline{G})\,,$$

where W is the Wightman functional defined by the Wightman functions of φ and $\tilde{\varphi}$ [e.g., for $\underline{F}=(f_1,f_1')$, $\underline{G}=(g_1,g_1')$, we have

$$\langle \underline{F},\underline{G}\rangle = \mathscr{W}_{\varphi\varphi}(\bar{f}_1 g_1) + \mathscr{W}_{\varphi\tilde{\varphi}}(\bar{f}_1 g_1' + \bar{f}_1' g_1) + \mathscr{W}_{\tilde{\varphi}\tilde{\varphi}}(\bar{f}_1' g_1').$$

If we denote by $\mathscr{N}_W^{\text{ext}}$ the Wightman ideal, on the vector space $\mathscr{D}_0^{\text{ext}} = \mathscr{B}_{\text{ext}}/\mathscr{N}_W^{\text{ext}}$ we can represent the fields φ and $\tilde{\varphi}$ as well as the Poincaré transformations.

The explicit characterization of $\mathscr{D}_0^{\text{ext}}$ can be made more easily by introducing the linear combinations

$$\varphi_R = \tfrac{1}{2}(\varphi+\tilde{\varphi}), \quad \varphi_L = \tfrac{1}{2}(\varphi-\tilde{\varphi})$$

(*right and left movers*), which are independent fields in the sense that $\langle \varphi_R,\varphi_L\rangle_0 = 0$, and therefore the corresponding Wightman functions factorize as products of those of φ_R and those of φ_L.

The original vector space $\mathscr{D}_0$ of the field φ can then be identified with the subspace of $\mathscr{D}_0^{\text{ext}}$ obtained by applying to the vacuum polynomials the combination $\varphi_R+\varphi_L$. We also have that $\mathscr{D}_0^{\text{ext}}$ is isomorphic to the vector space $\mathscr{B}_{R,L}/\mathscr{N}_{R,L}$, where $\mathscr{B}_{R,L}$ is the Borchers algebra corresponding to the two fields φ_R, φ_L, i.e., the algebra generated by elements of the form $\underline{F}=(\underline{f}_R,\underline{f}_L)$, with inner product induced by the Wightman functions of φ_R and φ_L.

C. Extension of the Hilbert structure to represent left and right movers

We can now come back to the problem raised in Sec. 3.1, namely, the construction of charged fields (here $\tilde{\varphi}$), or of charged states, starting from the Hilbert space realization of a chargeless field algebra $\mathscr{A}$ [here the local algebra generated by the local field $\varphi(f)$].

The first question of principle is whether the field $\tilde{\varphi}$ can be defined as an operator in the Hilbert space in which φ has been realized. [The field $\tilde{\varphi}$ (as a quadratic form on $\mathscr{D}_0$) could be obtained as a $\tau_{\mathscr{V}}$ limit of φ, but as previously remarked (see, also, Ref. 11), this is not enough to define a field operator, not even on $\mathscr{D}_0$.] The settling of this question is crucial for the solutions of two-dimensional models using φ and $\tilde{\varphi}$ as building blocks; in particular, the nonexistence of the "degree of freedom" associated to $\tilde{\varphi}$ in the (Krein) realization of φ is at the basis of a correct treatment of the vacuum degeneracy in the Schwinger model and the related physical features.[25]

Theorem 3.1: Let K be the Krein space realization of the massless scalar field φ (see Sec. II); then the field $\tilde{\varphi}(f)$, $f\in\mathscr{S}$, cannot be defined as an operator-valued distribution on K.

Proof: It is enough to show that the linear functional on K defined by

$$\tilde{F}(f) = \langle \tilde{\varphi}(\chi)\varphi(f)\rangle_0,$$

is not bounded on K. In fact, consider the sequence

$$h_n(p) = npf(np)/\pi,$$

with $f\in\mathscr{S}(\mathbb{R})$, real, symmetric, and such that $\int dp\, f(p)=1$. Now, $\forall g\in\mathscr{S}(\mathbb{R}^2)$,

$$\langle g,h_n\rangle = \int dp_1\,\epsilon(p_1)f(p_1)\bar{g}\left(\frac{p_1}{n},\frac{|p_1|}{n}\right) \xrightarrow{n\to\infty} \bar{g}(0)\int dp_1\,\epsilon(p_1)f(p_1)=0.$$

Moreover, $(h_n,h_n)_K < C$, since, by the above equation,

$$\lim_{n\to\infty}(h_n,h_n)_K = \lim_{n\to\infty}\langle h_n,h_n\rangle = \frac{1}{\pi}\int dp_1 |p_1|\,|f(p_1)|^2 < \infty,$$

and therefore we have

$$w\text{-}\lim_{n\to\infty} h_n = 0.$$

On the other hand,

$$\tilde{F}(h_n) = \int \frac{dp_1}{|p_1|}\epsilon(p_1)np_1 f(np_1)\chi(p_1,|p_1|) = \int dp\, f(p)\chi\left(\frac{p}{n},\frac{|p|}{n}\right) \xrightarrow{n\to\infty} 1. \qquad \square$$

Thus in the Hilbert space realization of φ there is no room for the field $\tilde{\varphi}$; more precisely, the Krein space closure of $\mathscr{D}_0$ does not contain $\mathscr{D}_0^{\text{ext}}$. The same argument applies to the left and right movers φ_R, φ_L, i.e., *the decomposition of φ into left and right movers cannot be done* in the Hilbert space realization of φ. One has to enlarge the space of states (or introduce "other degrees of freedom"), to allow the splitting into φ_R and φ_L.

A possible and natural solution is to introduce the Wightman functions of the enlarged system $(\varphi,\tilde{\varphi})$, i.e., the vacuum expectation values of the algebra $\tilde{\mathscr{A}}$ generated by φ and $\tilde{\varphi}$, and a Hilbert topology that majorizes them. Since this algebra is also generated by φ_R and φ_L and these two fields are independent, it is enough to majorize $\mathscr{W}_R$ and $\mathscr{W}_L$.

For the Krein space realization of the right and left movers, one proceeds as before separately for φ_R and φ_L. For the one-particle space, one gets

$$K_{\text{ext}}^{(1)} = K_R^{(1)}\oplus K_L^{(1)},$$

$$K_R^{(1)} = L^2\left(\frac{dp}{|p|},C_{+R}\right)\oplus\{V_{0R}\}\oplus\{\chi_R\},$$

$$K_L^{(1)} = L^2\left(\frac{dp}{|p|},C_{+L}\right)\oplus\{V_{0L}\}\oplus\{\chi_L\}.$$

(The old Krein space $K^{(1)}$ for φ is a proper subspace of $K_{\text{ext}}^{(1)}$!)

Similarly, one can introduce the charges

$$Q_{R,L} = i\pi[\varphi^+_{R,L}(v_0) - \varphi^-_{R,L}(v_0)],$$

and define the physical space by a subsidiary condition

$$Q_R^-\Psi = Q_L^-\Psi = 0.$$

IV. SYMMETRIES AND THEIR IMPLEMENTATION

The massless scalar field in two dimensions is not only interesting as an example of QFT with infrared singularities of the type of realistic gauge theories, but also because most of what is known about soluble QFT's in two dimensions (Thirring model, Schroer model, Schwinger model, etc.) relies strongly on the massless scalar field in two dimensions, since the latter is used as a building block for the construction of the solution of such models.

The discussion of the symmetries of the massless scalar field is therefore interesting also in view of the implications for more general two-dimensional local QFT models, in particular for the interplay between infrared singularities (i.e., long distance behavior of the fields) and the occurrence of symmetry breaking.

Especially in view of the implications for models like the Schwinger, Schroer, (and Thirring) models, which are regarded as prototypes of gauge QFT models, the status of symmetries like gauge symmetry and chiral symmetry become relevant for understanding nonperturbative mechanisms which likely have a counterpart in the four-dimensional case (see, e.g., Ref. 25).

The treatment adopted here for the massless scalar field emphasizes the local structure versus positivity, and therefore it is believed to shed light on the local (i.e., renormalizable) and covariant formulations of gauge QFT's, and the various structural mechanisms they are expected to exhibit.

In particular, it will be clear from the following sections that many of the structural properties of the symmetries and their local generators can be rigorously answered *only* by making reference to a Hilbert space topology that allows the realization of the fields as operators in a Hilbert space.

Finally, we want to stress that this type of question is less academic than it might appear, as is indirectly shown by the conflicting statements or results on the subject appearing in the literature.[5,19–24] For example, for the massless scalar field and its dual it has been argued that Lorentz invariance is broken (this question will be settled in Sec. IV A, and for the Schwinger model controversial statements have appeared on the vacuum degeneracy, the implementation of the large gauge transformations, the breaking of chiral symmetry, etc. (the rigorous settling of these questions[25] heavily relies on the contents of this section).

Before entering in the discussion, we recall a few basic facts about *continuous symmetries* of a field algebra, their breaking and their implementation. A one-parameter group of automorphisms β^λ of a local field algebra $\mathscr{F}$ is generated by a local charge Q_R if

$$\frac{d}{d\lambda}\beta^{\lambda}(A) = \lim_{R\to\infty} i[Q_R,A], \quad \forall A\in\mathscr{F},$$

$$Q_R = \int dx_1\, dx_0 f_R(x_1)\alpha_d(x_0) j_0(x_1,x_0),$$

with

$$f_R(x) = f(|x|/R), \quad f\in\mathscr{D}, \quad f(x) = \begin{cases}0, & \text{for } |x|>1,\\ 1, & \text{for } |x|\leqslant 1,\end{cases}$$

$$\alpha_d\in\mathscr{D}, \quad \int \alpha_d(x)dx = 1, \quad \lim_{d\to 0}\alpha_d(x) = \delta(x).$$

Here and in most two-dimensional models, time smearing is actually not necessary.

In a given representation of $\mathscr{A}$, characterized by a translationally invariant ground state Ψ_0, β^λ is said to be *unbroken* if there exists a one-parameter group of operators U^λ, such that $\beta^\lambda(A) = U^\lambda A (U^\lambda)^{-1}$ and

$$\langle U^\lambda\Psi, U^\lambda\Phi\rangle = \langle\Psi,\Phi\rangle.$$

Otherwise, the symmetry is said to be broken.

It should be stressed that, in general, even if a symmetry is unbroken, the Wightman functions need not be invariant, i.e.,

$$\langle\beta^\lambda(A)\rangle_0 = \langle A\rangle_0, \quad \forall A\in\mathscr{A}.$$

Actually, for unbroken symmetries, the invariance of the Wightman functions follows, in general, if U^λ leaves the vacuum invariant, and this may not be the case even if β^λ commutes with space-time translations if the vacuum is not the only translationally invariant state.

This case does not occur in standard QFT's satisfying all the Wightman axioms (including positivity), but it is generally realized when the infrared singularities are of the confining type[2] (as is believed to be the case for non-Abelian gauge QFT's), since in this case the vacuum is only essentially unique. In this case, it could very well be that U^λ exists, but it maps the ground state Ψ_0 in another translationally invariant state Ψ_0^λ, differing from Ψ_0 by a translationally invariant zero norm vector.

As a matter of fact, given a one-parameter group of automorphisms β^λ, $\lambda\in\mathbb{R}$, of a local algebra $\mathscr{A}$, generated by a (local) charge Q_R and commuting with the space time translations α_x, a necessary condition for its implementability in a realization of $\mathscr{A}$ in a Hilbert space $\mathscr{K}$ with an essentially unique vacuum is the existence of an operator Q such that

$$\lim_{R\to\infty}\langle[Q_R,A]\rangle_0 = \langle[Q,A]\rangle_0, \quad \forall A\in\mathscr{A}. \tag{4.1}$$

In the standard case, the vacuum is unique so that $Q\Psi_0 = \lambda\Psi_0$ and the right-hand side of Eq. (4.1) must necessarily vanish, yielding the standard condition for implementability in the differential form. When the Wightman functions exhibit infrared singularities that violate positivity the situation becomes richer; the vacuum is essentially unique and therefore there may be nontrivial operators Q mapping Ψ_0 into another translationally invariant vector, so that the right-hand side of Eq. (4.1) does not vanish. Natural candidates of such operators are infrared operators obtained as weak closures of $\mathscr{A}$. This is, in fact, the case of the Lorentz transformations, the gauge and chiral transformations discussed below. If Eq. (4.1) holds in the stronger form

$$\lim_{R\to\infty}[Q_R,A] = [Q,A], \quad \forall A\in\mathscr{A},$$

and $Q\mathscr{H}'\subset\mathscr{H}''$, then the symmetry β^λ is *bleached* in the physical space $\mathscr{H}_{\text{phys}} = \mathscr{H}'/\mathscr{H}''$, i.e., all the physical states are neutral with respect to η-unitary operators U^λ which implement β^λ. This is the case for gauge and chiral transformations (see Secs. IV B and IV C), whose generators Q and $\widetilde{Q}$ vanish on $\mathscr{H}_{\text{phys}}$ (but not on $\mathscr{H}'$!).

A. Lorentz symmetry and its implementation

We consider the field algebra generated by φ and $\widetilde{\varphi}$ (or by φ_R and φ_L). Then, as can be easily seen, the Wightman function $\widetilde{\mathscr{W}} = \langle\varphi\widetilde{\varphi}\rangle_0$ is not a Lorentz-invariant distribution,

$$\widetilde{\mathscr{W}}(\Lambda(x-y)) = \widetilde{\mathscr{W}}(x-y) - (1/2\pi)\log\lambda, \tag{4.2}$$

where $\Lambda x = \Lambda(x_+,x_-) = (\lambda x_+,\lambda^{-1}x_-)$, $x_\pm = x_0 \pm x_1$.

As a consequence, one can easily show that there cannot be η-unitary operators $U(\Lambda)$ such that (i) φ and $\widetilde{\varphi}$ transform as scalar (pseudoscalar) fields,

$$U(\Lambda)\varphi(x)U(\Lambda)^{-1} = \varphi(\Lambda x)\,,$$
$$U(\Lambda)\widetilde{\varphi}(x)U(\Lambda)^{-1} = \widetilde{\varphi}(\Lambda x)\,; \tag{4.3}$$

(ii) the vacuum is invariant under the Lorentz group,

$$U(\Lambda)\Psi_0 = \Psi_0\,.$$

Now, what is given is the original massless scalar field φ, whose transformation property is fixed by the natural action of the Lorentz group on the Borchers algebra,

$$f(x) \to f(\Lambda^{-1}x) \equiv f_\Lambda(x)\,,$$
$$\alpha_\Lambda(\varphi(f)) = \varphi(f_\Lambda)\,,$$

and whose Wightman functions are Lorentz invariant:

$$\langle \alpha_\Lambda(\varphi(f))\alpha_\Lambda(\varphi(g))\rangle_0 = \langle \varphi(f_\Lambda)\varphi(g_\Lambda)\rangle_0 = \langle \varphi(f)\varphi(g)\rangle_0\,.$$

However, the field $\widetilde{\varphi}$ constructed in terms of φ is not a local function of φ, and the identification of its transformation properties under the Lorentz group is much more delicate, since one has to extend the automorphism α_Λ from the local algebra $\mathscr{A}$ to its nonlocal extension $\widetilde{\mathscr{A}}$ (generated by φ and $\widetilde{\varphi}$).

As is typical of extensions of automorphisms from local algebras to their nonlocal extensions,[30] a certain arbitrariness is involved,[31] which can be reduced by requiring some continuity of the extension. We will use this criterion to fix the transformation properties of $\widetilde{\varphi}$, nanely, by extending α_Λ by $\tau_{\mathscr{W}}$ continuity.

More precisely, we will require

$$\alpha_\Lambda(\widetilde{\varphi}(f)) = \alpha_\Lambda(\tau_{\mathscr{W}}\lim_{n\to\infty} \varphi(f_n)) = \tau_{\mathscr{W}}\lim_{n\to\infty} \alpha_\Lambda(\varphi(f_n))\,.$$

In the following, we shall denote by $\widetilde{\mathscr{A}}_{\rm ext}$ the strong closure (analogous to that defined in Sec. 4.1) of the extended algebra $\widetilde{\mathscr{A}}$.

Theorem 4.1: There exists a unique $\tau_{\mathscr{W}}$-continuous extension of the Lorentz automorphism α_Λ from $\mathscr{A}$ to $\widetilde{\mathscr{A}}_{\rm ext}$, given by

$$\alpha_\Lambda(\widetilde{\varphi}(f)) = \widetilde{\varphi}(f_\Lambda) + 2\pi\hat{f}(0)\log\lambda\varphi(v_0)\,. \tag{4.4}$$

Proof: Since the truncated Wightman functions vanish, it is enough to consider the $\tau_{\mathscr{W}}$-continuity condition on the two-point function. To this purpose, we consider the sequence $\hat{\chi}_n$ defined by

$$\hat{\chi}_n(p) = \tfrac{1}{2}[(1-e^{ip_1n}) + (1-e^{-ip_1n})]\hat{\chi}(p)\,,$$

so that

$$\chi_n(x) = \chi(x) - \tfrac{1}{2}\chi(x+n) - \tfrac{1}{2}\chi(x-n)\,,$$

and let $\tilde{\chi}_n \in \mathscr{S}_0$ be defined in terms of χ_n as in Eq. (3.4). Then, for any $g\in\mathscr{S}(\mathbb{R}^2)$, the $\tau_{\mathscr{W}}$-continuous extension of α_Λ from $\mathscr{A}$ to $\widetilde{\mathscr{A}}$ must satisfy

$$\begin{aligned}\langle \alpha_\Lambda(\widetilde{\varphi}(\chi))\varphi(g)\rangle_0 &= \lim_{n\to\infty} \langle \alpha_\Lambda(\varphi(\tilde{\chi}_n))\varphi(g)\rangle_0 \\ &= \lim_{n\to\infty} \langle(\varphi(\tilde{\chi}_{n\Lambda}))\varphi(g)\rangle_0 \\ &= \lim_{n\to\infty} \langle \varphi(\tilde{\chi}_n)\varphi(g_{\Lambda^{-1}})\rangle_0 \\ &= \langle \widetilde{\varphi}(\chi)\varphi(g_{\Lambda^{-1}})\rangle_0\,.\end{aligned}$$

Then, by Eq. (4.2), we have

$$\langle \alpha_\Lambda(\widetilde{\varphi}(\chi))\varphi(g)\rangle_0 = \langle \widetilde{\varphi}(\chi_\Lambda)\varphi(g)\rangle_0 + 2\pi\log\lambda\hat{g}(0).$$

A similar equation holds for $\langle \varphi(g)\alpha_\Lambda(\widetilde{\varphi}(\chi))\rangle_0$, so that

$$[\alpha_\Lambda(\widetilde{\varphi}(\chi)) - \widetilde{\varphi}(\chi_\Lambda),\mathscr{A}] = 0\,,$$

i.e.,

$$\alpha_\Lambda(\widetilde{\varphi}(\chi)) - \widetilde{\varphi}(\chi_\Lambda) = 2\pi\log\lambda\varphi(v_0) + \delta\widetilde{Q} + \gamma\widetilde{\varphi}(v_0).$$

Finally, the Lorentz invariance of the Wightman function $\langle\widetilde{\varphi}(x)\widetilde{\varphi}(y)\rangle_0$ gives $\delta=\gamma=0$. Equation (3.4) extends from $\mathscr{A}$ to $\widetilde{\mathscr{A}}_{\rm ext}$ as a $\tau_{\mathscr{W}}$-continuous automorphism by strong continuity.

It is worthwhile to remark that the sequences χ_n and $\tilde{\chi}_n$ do not converge in the $\mathscr{S}$ topology, and therefore, even if $\widetilde{\varphi}$ is an operator-valued tempered distribution, the sequence $\widetilde{\varphi}(\chi_{n\Lambda}) \equiv \varphi(\tilde{\chi}_{n\Lambda})$ need not converge to $\widetilde{\varphi}(\chi_\Lambda)$; as a matter of fact,

$$\tau_{\mathscr{W}}\lim \widetilde{\varphi}(\chi_{n\Lambda}) = \varphi(\tilde{\chi}_\Lambda)\ \log\lambda\varphi(v_0).$$

The above transformation property of $\widetilde{\varphi}$ seems to have been missed by the treatments given in the literature,[16–18] which do not realize the relevance of the Hilbert space realization of the fields. As a matter of fact, the existence of the infrared operator $\varphi(v_0)$ crucially relies on the Krein realization,[32] and its introduction is mathematically unclear otherwise.[3]

This also shows the importance of making reference to a Hilbert topology to discuss such structural questions as the breaking of the Lorentz symmetry. Finally, it is important to stress that the above transformation property (4.4) defines an automorphism of $\widetilde{\mathscr{A}}_{\rm ext}$, not of $\widetilde{\mathscr{A}}$ (i.e., one has to add the infrared operators).

As a consequence of the above theorem, we have that *the Lorentz transformations are implemented* by η-unitary operators $U(\Lambda)$ in the Krein space $K_{\rm ext}$, and therefore *unbroken*.

Theorem 4.2: The Lorentz automorphism defined on $\widetilde{\mathscr{A}}_{\rm ext}$ by

$$\alpha_\Lambda(\varphi(f)) = \varphi(f_\Lambda)\,,$$
$$\alpha_\Lambda(\widetilde{\varphi}(f)) = \widetilde{\varphi}(f_\Lambda) + 2\pi\hat{f}(0)\log\lambda\varphi(v_0)\,,$$

can be implemented, in the Krein space K_e, by η-unitary operators $U(\Lambda)$ leaving the vacuum invariant. The operators $U(\Lambda)$ are determined, as usual, by their action in the one-particle space, which is

$$U(\Lambda)\Psi_f = \Psi_{f_\Lambda},\quad \forall f\in\mathscr{S}(\mathbb{R}^2)\,,$$
$$U(\Lambda)\Psi_{\tilde{\chi}} = \Psi_{\tilde{\chi}_\Lambda} + 2\pi\log\lambda v_0\,. \tag{4.5}$$

Proof: The invariance of the vacuum follows by a standard argument from the invariance of the Wightman functions,

$$\langle \alpha_\Lambda(\widetilde{\mathscr{A}})\rangle_0 = \langle\widetilde{\mathscr{A}}\rangle_0\,.$$

This latter equation follows from the invariance of $\langle\varphi\varphi\rangle_0$, $\langle\tilde{\varphi}\tilde{\varphi}\rangle_0$, and Eq. (4.4),

$$\langle\alpha_\Lambda(\varphi(f)\tilde{\varphi}(g))\rangle_0 = \langle\varphi(f)\tilde{\varphi}(g)\rangle_0 .$$

The explicit expression of the operators $U(\Lambda)$ that implement α_Λ is completely determined by their action in the one-particle space, which in turn is determined by the invariance of the two-point function (and the nondegeneracy of the inner product in $K^{(1)}$). □

Remark: In principle, one could consider also the situation in which the fields φ and $\tilde{\varphi}$ are treated symmetrically, in contrast with the general philosophy discussed in Sec. II, where φ is the basic given field and $\tilde{\varphi}$ is a "derived charged field." One could then envisage the case in which $\varphi(x)$ and $\tilde{\varphi}(x)$ transform as

$$\beta_\Lambda(\varphi) = \varphi_\Lambda + \delta \log \lambda \tilde{\varphi}(v_0) ,$$

$$\beta_\Lambda \tilde{\varphi} = \tilde{\varphi}_\Lambda + \tilde{\delta} \log \lambda \varphi(v_0) .$$

For $\delta = \tilde{\delta} = \frac{1}{2}$, we have a symmetric transformation property of φ and $\tilde{\varphi}$. this implies a symmetric transformation property of φ_R and φ_L. For $\delta = 0, \tilde{\delta} = 1$, we recover our formula (4.4).

It is worthwhile to note that all the above β_Λ, for any δ, $\tilde{\delta}$, define automorphisms of the algebra $\tilde{\mathscr{A}}_{\rm ext}$ [it is crucial that one has performed the extension from $\tilde{\mathscr{A}}$ to $\tilde{\mathscr{A}}_{\rm ext}$, by the addition of the infinitely delocalized or infrared operators $\varphi(v_0),\tilde{\varphi}(v_0)$, for which the Krein structure is essential]. All of them are implementable by η-unitary operators $U_{\delta\tilde{\delta}}(\Lambda)$.

However, only β_Λ with $\delta = 0$ maps the (original) local field algebra into itself; this automorphism has therefore a distinguished position among the others, since it *preserves the local structure*, namely, the characterizing property of the whole approach discussed so far.

It should be stressed that the field transformation (considered in most of the literature[5,19])

$$T_\Lambda(\varphi(f)) = \varphi(f_\Lambda), \quad T_\Lambda(\tilde{\varphi}(f)) = \tilde{\varphi}(f_\Lambda), \tag{4.6}$$

also defines an automorphism of $\tilde{\mathscr{A}}$, but it *cannot* be implemented by η-unitary operators that leave the vacuum invariant, and this has sometimes led to the conclusion that it not be implementable at all.

Actually, in the Krein realization of the algebra $\tilde{\mathscr{A}}$, where infrared operators are available, T_Λ is implemented by a unitary operator $V(\Lambda)$ which maps the vacuum state into another translationally invariant state,

$$V(\Lambda)\Psi_0 = \exp(-i \log \lambda \tilde{Q}\varphi(v_0))\Psi_0.$$

Clearly, by putting $U^\lambda_\infty \equiv \exp(-i \log \lambda \tilde{Q}\varphi(v_0))$, one easily checks that

$$V(\Lambda) = U^\lambda_\infty U(\Lambda), \tag{4.7}$$

where $U(\Lambda)$ is defined by Eq. (4.5). We again stress that the transformation law (4.6) is an extension of the Lorentz automorphism from $\tilde{\mathscr{A}}$ to $\tilde{\mathscr{A}}_{\rm ext}$ which is not $\tau_{\mathscr{V}}$-continuous.

B. Gauge transformations

As repeatedly mentioned in the previous sections, the massless field in two dimensions mimics several of the structure properties of gauge field theory. Actually, one can define *gauge transformations* as automorphisms of the field algebra $\mathscr{A}$ (or of $\tilde{\mathscr{A}}$),

$$\gamma^\lambda: \varphi(x) \to \varphi(x) + \lambda, \quad \tilde{\varphi}(x) \to \tilde{\varphi}(x), \quad \lambda \in \mathbb{R}.$$

One easily identifies the gauge-invariant subalgebra $\mathscr{A}_{\rm obs} \subset \mathscr{A}$ as that generated by $\varphi(f), f \in \mathscr{S}_0$, and the gauge-invariant states of K (or of $K_{\rm ext}$).

The above one-parameter group of automorphisms γ^λ is generated on $\mathscr{A}$ by the local charge

$$Q_R = \int d^2x\, \partial_0\varphi(x) f_R(x_1)\alpha_d(x_0). \tag{4.8}$$

To discuss the breaking of such symmetry, it is crucial to make reference to a Hilbert space realization of the field algebra. (For simplicity, we will restrict our attention to $\mathscr{A}$ and to its Krein realization in K, see Sec. 2.B.)

Theorem 4.3: The gauge transformation automorphism γ^λ is implementable (i.e., not broken) in the Krein space K, by the η-unitary operators

$$\Gamma^\lambda = \exp i\lambda Q, \tag{4.9}$$

where Q is the infrared operator defined by

$$Q = i\pi(\varphi_+(v_0) - \varphi_-(v_0)).$$

Actually, one also has, on $\mathscr{D}_0$,

$$Q = \text{weak-}\lim_{R\to\infty} Q_R. \tag{4.10}$$

Proof: The implementation by Γ^λ follows trivially from the commutation relations between Q and the field algebra $\mathscr{A}$ [see Eq. (4.10)], and the fact that $\mathscr{D}_0 = \mathscr{A}\Psi_0$ is a set of analytic vectors for Q (so that the exponential is well defined).

The η unitarity of Γ^λ follows from Q being η symmetric.

Furthermore, for any $g \in \mathscr{S}$,

$$\langle\partial_0\varphi(\alpha_d f_R)\varphi(g)\rangle_0 = -i\pi \int dq\, \hat{f}(q)\hat{g}\left(\frac{q}{R}, \frac{|q|}{R}\right)\hat{\alpha}_d\left(\frac{|q|}{R}\right)$$

$$\to -i\pi\hat{g}(0) = \langle Q\varphi(g)\rangle_0.$$

The vanishing of the truncated Wightman functions implies that Q_R is convergent to Q as a bilinear form on $\mathscr{D}_0 \times \mathscr{D}_0$. Moreover, one can easily check that

$$\langle\partial_0\varphi(\alpha_d f_R)\partial_0\varphi(\alpha_d f_R)\rangle_0$$

$$= \pi \int dq|q|\, |\hat{f}(q)|^2 \left|\hat{\alpha}_d\left(\frac{|q|}{R}\right)\right|^2 < \infty,$$

so that, by using the explicit form of the Krein norm, $\|Q_R\Psi_0\| < C$. By the factorization of the Krein product [Eq. (2.9)], this, in turn, implies

$$\|Q_R\Psi\| < C'_\Psi, \quad \forall\Psi \in \mathscr{D}_0,$$

and therefore Q_R is actually weakly convergent on $\mathscr{D}_0$. □

Remark: This result shows that the Krein realization, which is characterized by associating a maximal set of states to the given Wightman functions, also has the property of having a large set of symmetries unbroken. It is worthwhile to remark that the Wightman functions are not invariant under γ^λ; nevertheless, the automorphism is implementable essentially by the mechanism discussed at the end of the introduction to this section, namely, the vacuum is essentially invariant but not strictly invariant under Γ^λ (it is mapped into another translationally invariant vector of K). The subspace V_0 of translationally invariant vectors of K provides a triangular representation of the gauge group. By strong continuity, one can extend the gauge transformation from the local algebra $\mathscr{A}$ to its (strong) closure $\mathscr{A}_{\rm ext}$, in particular, to the infrared operators discussed in Sec. 2.3. Clearly, they are all neutral under Γ^λ, which follows from the fact that the infrared operators actually belong to the strong closure of the gauge-invariant algebra $\mathscr{A}_{\rm obs}$.

C. Chiral transformations

From the experience with the two-dimensional fermion models (like the Schroer, Thirring, and Schwinger models) one learns that an important symmetry is the *chiral transformation*, which corresponds to a shift of the dual Bose field $\tilde{\varphi}$, occurring in the bosonization of the chiral current $j^5_\mu = (1/\sqrt{\pi})\partial_\mu \tilde{\varphi}$.

Thus it is of some interest to investigate in our framework the status of the one-parameter group of the chiral transformations $\tilde{\gamma}^\lambda$, $\lambda \in \mathbb{R}$, defined on $\tilde{\mathscr{A}}$ by

$$\tilde{\gamma}^\lambda:\ \varphi(x) \to \varphi(x), \quad \tilde{\varphi}(x) \to \tilde{\varphi}(x) + \lambda.$$

They clearly define a one-parameter group of automorphisms of $\tilde{\mathscr{A}}$, which is generated by the local charge

$$\tilde{Q}_R = \int d^2x\, \partial_0 \tilde{\varphi}(x) f_R(x_1) \alpha_d(x_0). \tag{4.11}$$

The situation is very similar to the one discussed in the previous section, and one has the following theorem.

Theorem 4.4: The chiral transformation automorphism is implementable in the enlarged Krein space $K_{\rm ext}$ (see Sec. 4.3) by the η-unitary operators

$$\tilde{\Gamma}^\lambda = \exp i\lambda \tilde{Q}, \tag{4.12}$$

where $\tilde{Q}$ is the infrared operator defined by

$$\tilde{Q}(\tilde{\varphi}_+(v_0) - \tilde{\varphi}_-(v_0)).$$

Actually, one also has

$$\tilde{Q} = \underset{R\to\infty}{\text{weak-lim}}\ \tilde{Q}_R. \tag{4.13}$$

The proof follows the same lines as that of the previous theorem.

The remarks of the previous sections also apply to the present case. In particular, the discussion of chiral symmetry breaking in the Schwinger model cannot be done in a convincing way without taking into account[19,24] the Krein–Hilbert space in which the fields are realized as operators and the above results on the implementability of symmetries.[25]

D. Scale and special conformal transformations

The scale transformations, or dilatations, represent another interesting case of a symmetry leading to a group of automorphisms of the local (extended) algebra of the massless scalar field that are implementable in a way compatible with Eq. (4.1).

In two space-time dimensions, scale transformations can be written in terms of the light cone variables $x_\pm = x_0 \pm x_1$ as

$$(x_+, x_-) \to (sx_+, sx_-), \quad s > 0.$$

If we require that the fields φ and $\tilde{\varphi}$ transform as a scalar and pseudoscalar, respectively, i.e., if we define the automorphism of $\tilde{\mathscr{A}}$,

$$\alpha_s(\varphi(f)) = \varphi(f_s), \quad \alpha_s(\tilde{\varphi}(f)) = \tilde{\varphi}(f_s), \tag{4.14}$$

with $f_s(x) = f(s^{-1}x)$, then the corresponding two-point functions transform in the following way:

$$\mathscr{W}(s(x-y)) = \mathscr{W}(x-y) - (1/2\pi)\log s,$$
$$\tilde{\mathscr{W}}(s(x-y)) = \tilde{\mathscr{W}}(x-y). \tag{4.15}$$

Therefore, as for the case of the Lorentz symmetry, one cannot find a η-unitary operator $U(S)$ implementing the automorphism (4.14) *and* leaving the vacuum invariant.

However, according to Eq. (4.1), this does not necessarily imply symmetry breaking, since one can find a η-unitary operator $U(s)$ mapping the vacuum into another translation-invariant state (see Sec. 4.1).

Alternatively, it may be possible to define an automorphism of the extended algebra $\tilde{\mathscr{A}}_{\rm ext}$ [coinciding with (4.14) on the gauge-invariant subalgebra] which is implementable in the Krein space K_e by η-unitary operators leaving the vacuum invariant. Actually, we have the following theorem (the proof is similar to that of Theorem 4.3).

Theorem 4.5: The automorphism defined on $\tilde{\mathscr{A}}_{\rm ext}$ by

$$\alpha_s(\varphi(f)) = \varphi(f_s) + \pi\hat{f}(0)\log s\varphi(v_0),$$
$$\alpha_s(\tilde{\varphi}(f)) = \tilde{\varphi}(f_s) + \pi\hat{f}(0)\log s\tilde{\varphi}(v_0) \tag{4.16}$$

is implementable in the Krein space K_e by η-unitary operators $U(s)$ leaving the vacuum invariant. The operators $U(s)$ are determined by their action in the one-particle space, which is

$$U(s)\Psi_f = \Psi_{f_s} + \pi\hat{f}(0)\log s\, v_0,$$
$$U(s)\Psi_{\tilde{g}} = \Psi_{\tilde{g}_s} + \pi\hat{f}(0)\log s\, \tilde{v}_0, \tag{4.17}$$

where $\Psi_f = \varphi(f)\Psi_0$, $\Psi_{\tilde{g}} = \tilde{\varphi}(g)\Psi_0$, $f,g \in \mathscr{S}$.

Remark: As can be easily checked, the most general automorphism of the extended field algebra satisfying our requirements is given by

$$\alpha_{s,\gamma}(\varphi(f)) = \varphi(f_s) + \pi\hat{f}(0)\log s(\varphi(v_0) + \gamma\tilde{\varphi}(v_0)),$$
$$\alpha_{s,\gamma}(\tilde{\varphi}(f)) = \tilde{\varphi}(f_s) + \pi\hat{f}(0)\log s(\tilde{\varphi}(v_0) - \gamma\varphi(v_0)),$$

where γ is an arbitrary real parameter. Clearly, only for $\gamma = 0$, i.e., definition (4.16), we get the $\tau_{\mathscr{W}}$-continuous extension to $\tilde{\mathscr{A}}_{\rm ext}$ of the automorphism

$$\alpha_s(\varphi(f)) = \varphi(f_s) + \pi\hat{f}(0)\log s\varphi(v_0), \tag{4.18}$$

defined on the original algebra $\mathscr{A}_{\rm ext}$.

Similarly to the case of the Lorentz symmetry (see Sec. 4.1), the field transformation

$$T_s(\varphi(f)) = \varphi(f_s), \quad T_s(\tilde{\varphi}(f)) = \tilde{\varphi}(f_s) \tag{4.19}$$

defines an automorphism of $\tilde{\mathscr{A}}$ that can be implemented by a (η-unitary) operator mapping the vacuum state into another translation-invariant state. In fact, by putting

$$U^s_\infty = \exp(-is(Q\varphi(v_0) + \tilde{Q}\tilde{\varphi}(v_0))),$$

one can easily check that the transformation (4.19) is implemented by

$$V(s) = U^s_\infty U(s),$$

where $U(s)$ is defined by Eq. (4.17).

We now consider the case of the special conformal transformations,[4,33] defined by

$$x_\mu \to x'_\mu = (x_\mu - b_\mu x^2)/(1 - 2bx + b^2x^2), \qquad (4.20)$$

where b_μ is a real vector.

We recall that in two-dimensional space-time the conformal group is decomposed into a direct product of two groups acting on the light cone variables $x_\pm$ as

$$x_+ \to x'_+ = x_+/(1 + b_- x_+);$$
$$x_- \to x'_- = x_-/(1 + b_+ x_-), \qquad (4.21)$$

where $b_\pm = b_0 \pm b_1$ are the light cone components of the vector defined by (4.20).

Therefore, it is convenient to consider the transformation properties of the two-point functions $\mathscr{W}_R$ and $\mathscr{W}_L$ separately.

For example, by choosing the function

$$\mathscr{W}_R(x_+ - y_+) = -(1/4\pi)\log(x_+ - y_+ + i\epsilon),$$

we find

$$\mathscr{W}_R(x'_+ - y'_+) = \mathscr{W}_R(x_+ - y_+) + (1/4\pi)\{\log(1 + b_-(x_+ + i\epsilon)) + \log(1 + b_-(y_+ - i\epsilon))\}. \qquad (4.22)$$

Thus one can easily verify that the two-point function $\mathscr{W}_R$ is invariant under the following transformation of the positive and negative energy parts of the field operators:

$$\alpha_b(\varphi_R^\pm(x_+)) = \varphi_R^\pm(x'_+) + \tfrac{1}{2}\log(1 + b_-(x_+ \pm i\epsilon))\varphi_R^\pm(v_0). \qquad (4.23)$$

[A transformation property of the form (4.23) has also appeared in the literature;[4,33] but its justification in terms of $\tau_{\mathscr{W}}$ continuity is missing.] Obviously, the same argument applies also to the positive and negative energy parts of the field $\varphi_L^\pm$. From (4.23) we see that the transformation acts differently on the positive and negative energy parts of the fields. The structure of the automorphism of the extended field algebra $\tilde{\mathscr{A}}_{\rm ext}$ defined by the transformation (4.23) becomes clear by writing it in the form

$$\begin{aligned}\alpha_b(\varphi_R(x_+)) &= \varphi_R(x'_+) + \tfrac{1}{2}[\log b_- + \log(x_+ + 1/b_- + i\epsilon)]\varphi_R^+(v_0) \\ &\quad + \tfrac{1}{2}[\log b_- + \log(x_+ + 1/b_- - i\epsilon)]\varphi_R^-(v_0) \\ &= \varphi_R(x'_+) + \tfrac{1}{2}(\log b_- + \log|x_+ + 1/b_-|)\varphi_R(v_0) \\ &\quad + \Theta(-(x_+ + 1/b_-))Q_R. \end{aligned} \qquad (4.24)$$

From (4.24) we see that the automorphism α_b involves both the infrared operators $\varphi_R(v_0)$, Q_R [and $\varphi_L(v_0)$, Q_L for the action on the fields φ_L] defined in Sec. III C.

It is not difficult to check that the term proportional to Q_R (Q_L), which does not commute with the field φ_R (φ_L), is needed to compensate for the violation of the local commutativity coming from the fact that the transformation (4.20) is able to change spacelike into timelike vectors.[4]

[1] L. K. Hadjiivanov, JINR communication, E2-80-445, Dubna, 1980.
[2] G. Morchio and F. Strocchi, Ann. Inst. H. Poincaré A 33, 251 (1980); Scuola Normale Superiore 9/79, Pisa, 1979.
[3] N. Nakanishi, Progr. Theor. Phys. 57, 269, 580, 1025 (1977).
[4] B. Schroer and J. A. Swieca, Phys. Rev. D 10, 480 (1974).
[5] A. Z. Capri and R. Ferrari, Nuovo Cimento 62 (1981); J. Math. Phys. 25, 1 (1983).
[6] G. Morchio and F. Strocchi, Nucl. Phys. B 211, 471 (1983); *Fundamental Problems of Gauge Field Theory. Erice Lectures 1985*, edited by G. Velo and A. S. Wightman (Plenum, New York, 1986).
[7] H. Grundlig, Commun. Math. Phys. 114, 69 (1988).
[8] D. Buchholz, Commun. Math. Phys. 85, 49 (1982).
[9] It is not, in general, appreciated that in the nowadays popular BRST quantization of covariant gauge theories (and even of string theories), the identification of the physical states requires solving the BRST subsidiary condition in a suitable closure of the local states.
[10] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).
[11] G. Morchio, D. Pierotti, and F. Strocchi, International School for Advanced Studies, Trieste 84/87/EP, 1987.
[12] An example of nonmaximal Hilbert structure for the massless scalar field in two dimensions is obtained in terms of a single Sobolev-type seminorm[2,11] p_1 on $\mathscr{S}(\mathbf{R}^2)$, which majorizes the Wightman functions in the sense that

$$|\mathscr{W}(f_n^* \times g_m)| < p_n(f_n)p_m(g_m),$$
$$p_n(f_1 \cdots f_n) = 2^n n! p_1(f_1)p_1(f_2)\cdots p_1(f_n).$$

The corresponding Hilbert space of states associated to the Wightman functions is, however, too narrow to allow the introduction of the (delocalized) operators and states widely used in the literature (without a rigorous discussion of them!); for a detailed discussion, see Ref. 11.
[13] J. Bognar, *Indefinite Inner Product Spaces* (Springer, Berlin, 1974).
[14] A. S. Wightman, in *Cargèse Lectures in Theoretical Physics 1964*, edited by M. Levy (Gordon & Breach, New York, 1967); J. Schwinger, Phys. Rev. 128, 2425 (1962).
[15] L. Gårding and A. S. Wightman, Ark. Fys. 28, 129 (1964).
[16] D. T. Stoyanov, JINR communication, E2-11884, Dubna, 1978.
[17] L. K. Hadjiivanov and D. T. Stoyanov, Rep. Math. Phys. 15, 361 (1979).
[18] L. K. Hadjiivanov, S. G. Mikhov, and D. T. Stoyanov, J. Phys. A 12, 119 (1979).
[19] J. H. Lowenstein and J. A. Swieca, Ann. Phys. (NY) 68, 172 (1971).
[20] A. Casher, J. Kogut, and L. Susskind, Phys. Rev. Lett. 31, 792 (1973).
[21] S. Coleman, R. Jackiw, and L. Susskind, Ann. Phys. (NY) 93, 267 (1975).
[22] J. Kogut and L. Susskind, Phys. Rev. D 11, 3594 (1975).
[23] J. A. Swieca, Fortschr. Phys. 25, 303 (1977).
[24] A. K. Raina and G. Wanders, Ann. Phys. (NY) 132, 404 (1981).
[25] G. Morchio, D. Pierotti, and F. Strocchi, Ann. Phys. (NY) 188, 127 (1988).
[26] J. Fröhlich, G. Morchio, and F. Strocchi, Ann. Phys. (NY) 119, 241 (1979).
[27] F. Strocchi and A. S. Wightman, J. Math. Phys. 15, 2198 (1974).
[28] R. Ferrari, L. E. Picasso, and F. Strocchi, Commun. Math. Phys. 35, 25 (1974); Nuovo Cimento A 39, 1 (1977).
[29] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. 1 (Academic, New York, 1972).
[30] G. Morchio and F. Strocchi, Commun. Math. Phys. 99, 153 (1985).
[31] About this point, a certain confusion seems to be present in the literature,[5,16–18] where the removal of such arbitrariness is not discussed in a convincing way, and the various choices look rather *ad hoc*.
[32] A transformation property of the form (4.4) has also appeared in a treatment[1] of the massless scalar field, which heavily uses the general results and ideas of Ref. 2. The general framework and strategy of Ref. 1 is, however, substantially different from the one adopted here, where the emphasis is on the definition of the theory in terms of Wightman functions without *a priori* assumptions on the existence of creation and annihilation operators (Segal-type approach), and where the construction of the field $\tilde{\varphi}$ and its transformation properties are *derived*, by using $\tau_{\mathscr{W}}$ continuity, without making simplifying and somewhat *ad hoc* assumptions.
[33] L. K. Hadjiivanov, S. G. Mikhov, and D. T. Stoyanov, Rep. Math. Phys. 16, 89 (1979).

The Dirac equation in a non-Riemannian manifold: II. An analysis using an internal local n-dimensional space of the Yang–Mills type

S. Marques-Bonham
NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, M.S. 209, P.O. Box 500, Batavia, Illinois 60510

(Received 22 April 1988; accepted for publication 25 October 1989)

The geometrical properties of a flat tangent space-time local to the generalized manifold of the Einstein–Schrödinger nonsymmetric theory, with an internal n-dimensional space with the SU(n) symmetry group, is developed here. As an application of the theory, a generalized Dirac equation, where the electromagnetic and the Yang–Mills fields are included in a more complex field equation, is then obtained. When the two-dimensional case is considered, the theory can be immediately interpreted through the algebra of quaternions, which, through the Hurwitz theorem, presupposes a generalization of the theory using the algebra of octonions.

I. INTRODUCTION

A geometrical treatment of a gauge theory built to describe particles in the presence of gravitation, electromagnetism, and Yang–Mills fields has been developed by some authors since Einstein's attempt to unify gravitation and electromagnetism in his (complex) nonsymmetric theory, the so-called Einstein–Schrödinger[1] (ES) theory. The Bonnor–Moffat–Boal[2] (BMB) theory was successful in obtaining a correct limit to the Einstein–Maxwell theory and the Borchsenius[3] theory used the same principle to include the Yang–Mills field. Even though these theories have been criticized[4] and the "physical limit" has not been convincing, they are attractive from the point of view of a geometrical treatment of gravitation plus gauge theory. Also, these theories permit the extension to an octonionic theory through a theorem of Hurwitz.[5,6] However, given the present status of actual unified theories, the use of such a theory is not yet clear, but at a minimum, it constitutes an attempt in making useful some mathematical tools such as algebra and symmetry properties in a (geometrical) unified theory on the curved space-time.

The main goal of the present work is to obtain the Dirac equation for a spin-1/2 particle placed locally to a curved space-time and in the presence of gravitation, electromagnetism, and Yang–Mills fields, using the ES nonsymmetrical theory (see Ref. 7). To achieve this it is necessary to introduce an n-dimensional internal space to the (complex) space-time manifold of the ES theory (the notation used in this work is about the same as used in Refs. 6 and 7). Since we are interested in working with Yang–Mills fields, here we use the internal space of the $n\times n$ matrices, with SU(n) as the internal symmetry group, as in the Borchsenius theory.[3] Every object in this internal space can be expanded in terms of n^2 linearly independent matrices $\{\tau_0,\tau_i,$ $i=1,2,\ldots,(n^2-1)\}$, where $\tau_0\equiv\mathbf{1}_{n\times n}$ and $\tau_i^\dagger=\tau$. The line element is defined on this extended manifold as

$$ds^2=(1/n)\mathrm{Tr}(G_{\mu\nu}\,dx^\mu\,dx^\nu),\tag{1.1}$$

where

$$G_{\mu\nu}=(G^{\,a}_{\mu\nu b}(x)),\quad a,b=1,\ldots,n\tag{1.2}$$

is a matrix in the internal space such that

$$(1/n)\mathrm{Tr}\,G_{\mu\nu}=g_{\mu\nu},\tag{1.3}$$

with $g_{\mu\nu}$ being the metric of the ES nonsymmetric theory. It is also imposed that

$$G^{\dagger}_{\mu\nu}=G_{\nu\mu},\tag{1.4}$$

where the Hermitian conjugation operates on the internal matrix indices. There exists an inverse $G^{\mu\nu}$ such that

$$G_{\mu\alpha}G^{\mu\nu}=G^{\nu\mu}G_{\alpha\mu}=\delta^\nu_\alpha\mathbf{1},\tag{1.5}$$

where the order of factors is important and where Eqs. (1.3) and (1.4) are used. The metric $G_{\mu\nu}$ can be written as

$$G_{\mu\nu}=q_{\mu\nu0}\tau_0+q_{\mu\nu i}\tau_i,\quad i=1,2,\ldots,(n^2-1),\tag{1.6}$$

where, following conditions (1.3) and (1.4), $q_{\mu\nu0}$ is the metric on the manifold of the ES or BMB theory, which includes the electromagnetism through the Maxwell tensor $F_{\mu\nu}$:

$$q_{\mu\nu0}=g_{\mu\nu}=g_{\underline{\mu\nu}}+ipF_{\underset{\vee}{\mu\nu}};\tag{1.7}$$

and $q_{\mu\nu_i}$ should be of the Yang–Mills type

$$q_{\mu\nu_i}=\tfrac{1}{2}(i\epsilon p^2/\hbar)F_{\underset{\vee}{\mu\nu}i},\tag{1.8}$$

where ϵ is the elementary isotopic charge when $n=2$. The constant p is defined such that in the limit $p\to0$, the field equations and geometrical properties of the Einstein–Maxwell–Yang–Mills theory are obtained (see Refs. 2 and 3): Its value is given as $p=-2\hbar/e$, $|p|=3.8\times10^{-32}$ cm ($c=G=1$).

The properties of covariant derivatives on the manifold for the ES nonsymmetrical manifold state that the space-time connection is such that

$$\Omega^\mu{}_{\alpha\nu}=\Omega^{*\mu}{}_{\nu\alpha}=\Omega^\mu{}_{\alpha\nu}+iK^\mu{}_{\underset{\vee}{\alpha\nu}}.\tag{1.9}$$

To obtain the field equations through a minimal action principle, we also have to define the Schrödinger connection

$$\theta^\rho{}_{\mu\nu}=\Omega^\rho{}_{\mu\nu}-(2/ip)\delta^\rho_\mu A_\nu,\tag{1.10}$$

where A_ν is the electromagnetic vector potential and can be written in terms of the vector torsion $\Omega^\rho{}_{\underset{\vee}{\rho\nu}}$ as

$$A_\nu=-\tfrac{1}{3}(ip)\Omega^\rho{}_{\underset{\vee}{\rho\nu}}.\tag{1.11}$$

Taking an internal vector $\psi^a=\psi^a(x)$, $a=1,\ldots,n$, the internal covariant derivative is given by

$$\psi^a{}_{\|\mu}=\psi^a{}_{,\mu}+\Gamma_\mu{}^a{}_b\psi^b,\tag{1.12}$$

where Γ_μ is the internal connection. Here Γ_μ is taken to be of the Yang–Mills form:

$$\Gamma_\mu = -(i\epsilon/\hbar)\mathbf{b}_\mu\cdot\boldsymbol{\tau}. \tag{1.13}$$

The internal curvature is then obtained through the difference

$$\psi^a{}_{\|\mu\nu} - \psi^a{}_{\|\nu\mu} = P_{\mu\nu}{}^a{}_b\psi^b, \tag{1.14}$$

where $P_{\mu\nu}$ is the internal curvature given by

$$P_{\mu\nu} = \Gamma_{\mu,\nu} - \Gamma_{\nu,\mu} - [\Gamma_\mu,\Gamma_\nu]. \tag{1.15}$$

An object $K = (K^a{}_b)$ with two internal matrix indices then transforms as

$$K' = UKU^\dagger, \tag{1.16}$$

where, since the symmetry group is SU(n), the transformation matrices U are unimodular matrices: $U^\dagger = U^{T*} = U^{-1}$, det $U = 1$.

A total covariant derivative of a space-time vector $V^\mu(x)$ can be obtained through the parallel transport of this vector on the extended space as

$$V^\mu{}_{|\overset{\alpha}{+}}(x) = V^\mu{}_{,\alpha} + \Omega^\mu{}_{\rho\alpha}V^\rho + [\Gamma_\alpha,V^\mu]. \tag{1.17}$$

A "total curvature" is then obtained through the difference

$$V^\mu{}_{|\overset{\alpha}{+}\overset{\rho}{+}} - V^\mu{}_{|\overset{\alpha}{+}\overset{\rho}{+}} = \mathbf{R}^\mu_{\lambda\alpha\rho}V^\lambda - V^\mu P_{\alpha\rho} - 2V^{\overset{\mu}{+}}{}_{|\lambda}\Omega^\lambda_{\underset{\vee}{\alpha\rho}}, \tag{1.18}$$

where $V^\mu(x)$ can be written in terms of internal components as

$$V^\mu(x) = v_0^\mu(x)\tau_0 + v_i^\mu(x)\tau_i, \quad i = 1,2,3.$$

The total curvature $\mathbf{R}^\mu{}_{\lambda\alpha\rho}$ gives the mixture of the space-time and internal curvatures:

$$\begin{aligned}\mathbf{R}^\mu{}_{\lambda\alpha\rho} &= (\boldsymbol{\Gamma}^\mu{}_{\lambda\alpha,\rho} + \boldsymbol{\Gamma}^\mu{}_{\nu\rho}\boldsymbol{\Gamma}^\nu{}_{\lambda\alpha}) - (\boldsymbol{\Gamma}^\mu{}_{\lambda\rho,\alpha} + \boldsymbol{\Gamma}^\mu{}_{\nu\alpha}\boldsymbol{\Gamma}^\nu{}_{\lambda\rho})\\ &= R^\mu{}_{\lambda\alpha\rho} + \delta^\mu_\lambda P_{\alpha\rho},\end{aligned} \tag{1.19}$$

with

$$\boldsymbol{\Gamma}^\rho{}_{\nu\alpha} = \Omega^\rho{}_{\nu\alpha}\tau_0 + \delta^\rho_\nu\Gamma_\alpha.$$

To obtain the field equations for the extended theory, we use the Palatini variational method. The action is

$$\mathscr{A} = \int \mathscr{L}\, d^4x,$$

where the Lagrangian $\mathscr{L}$ is taken as

$$\mathscr{L} = \mathrm{Tr}\{\mathscr{G}^{\mu\nu}\mathbf{R}_{\mu\nu} + [1/(ip)^2]\mathscr{G}^{\underset{\vee}{\mu\nu}}G_{\underset{\vee}{\mu\nu}}\}, \tag{1.20}$$

where $\mathbf{R}_{\mu\nu} = \mathbf{R}^\rho{}_{\mu\nu\rho}$ by (1.19), and

$$\mathbf{R}_{\mu\nu} = \mathbf{R}^\rho{}_{\mu\nu\rho} \tag{1.21}$$

The field equations obtained on this extended manifold are then (the notation used in the following equation for the covariant derivative of $\mathscr{G}^{\mu\nu}$ is the usual when it is given in terms of the Schrödinger connection $\theta^\rho{}_{\mu\nu}$):

$$\mathscr{G}^{\overset{\mu}{+}\overset{\nu}{-}}{}_{|\alpha} = (\boldsymbol{\Gamma}^\mu{}_{\lambda\alpha,\rho} + \boldsymbol{\Gamma}^\mu{}_{\nu\rho}\boldsymbol{\Gamma}^\nu{}_{\lambda\alpha}) - (\boldsymbol{\Gamma}^\mu{}_{\lambda\rho,\alpha} + \boldsymbol{\Gamma}^\mu{}_{\nu\alpha}\boldsymbol{\Gamma}^\nu{}_{\lambda\rho})$$

$$\mathscr{G}^{\overset{\mu}{+}\overset{\nu}{-}}{}_{|\alpha} \tag{1.22}$$

$$\mathscr{G}^{\underset{\vee}{\mu\nu}}{}_{,\alpha} = 0, \tag{1.23}$$

$$^*\mathbf{R}_{\mu\nu}(\theta) = 0, \tag{1.24}$$

$$^*\mathbf{R}_{\underset{\vee}{\mu\nu},\sigma}(\theta) + {}^*\mathbf{R}_{\underset{\vee}{\sigma\mu},\nu}(\theta) + {}^*\mathbf{R}_{\underset{\vee}{\nu\sigma},\mu}(\theta) = 0. \tag{1.25}$$

Equation (1.25) is a consequence of the fact that

$$^*\mathbf{R}_{\underset{\vee}{\mu\nu}}(\theta) = \tfrac{2}{3}(\Omega_{\mu,\nu} - \Omega_{\nu,\mu}) + \Gamma_{\nu,\mu} - \Gamma_{\mu,\nu} + [\Gamma_\mu,\Gamma_\nu]. \tag{1.26}$$

In Eqs. (1.24)–(1.26) the argument θ in $\mathbf{R}_{\mu\nu}(\theta)$ means that the expression for the generalized Ricci tensor is written in terms of the Schrödinger connection. Also, in (1.24) to (1.26) we have

$$^*\mathbf{R}_{\mu\nu}(\theta) = \mathbf{R}_{\mu\nu}(\theta) + I_{\mu\nu}, \tag{1.27}$$

where

$$I_{\mu\nu} = [1/(ip)^2](G_{\mu\sigma}G^{\underset{\vee}{\sigma\rho}}G_{\rho\nu} + \tfrac{1}{2}G_{\mu\nu}G_{\sigma\rho}G^{\underset{\vee}{\sigma\rho}} + G_{\underset{\vee}{\mu\nu}}). \tag{1.28}$$

We now proceed to Sec. II, where the properties of a tangent space on this extended manifold will be presented.

II. THE n-DIMENSIONAL COMPLEX TANGENT SPACE

A local tangent space can be defined on this extended space-time manifold through a generalized correspondence principle.[8] It is also supposed that this tangent space has attached to it the same n-dimensional internal space.

Define $n\times n$ matrix vierbeins $E^a_\mu(x)$ such that

$$G_{\mu\nu} = E^{\dagger a}_\nu E^b_\mu \eta_{ab}. \tag{2.1}$$

Then, according to the correspondence principle[9] generalized to this case, the line element can be written in both spaces as

$$\begin{aligned}ds^2 &= (1/n)\mathrm{Tr}(G_{\mu\nu}\,dx^\mu\,dx^\nu)\\ &= (1/n)\mathrm{Tr}(\eta_{ab}\,d\mathbf{x}^{\dagger a}\,d\mathbf{x}^b),\end{aligned} \tag{2.2}$$

$$d\mathbf{x}^a = E^a_\mu\,dx^\mu = , \quad d\mathbf{x}^{\dagger a} = E^{\dagger a}_\mu\,dx^\mu,$$

where, aiming toward a physical interpretation, the metric on the tangent space is taken with the structure of the Minkowski metric η_{ab}.

Since there exists an inverse $G^{\mu\nu}$ such that (1.5) is true, we must have

$$G^{\mu\nu} = E^{\dagger\mu}_a E^\nu_b \eta^{ab}. \tag{2.3}$$

From (2.3) we obtain the corresponding orthogonality relations for the matrix vierbeins:

$$E^b_\mu E^{\dagger\mu}_c = E^\mu_c E^{\dagger b}_\mu = \delta^b_c\tau_0, \quad E^{\dagger a}_\alpha E^\nu_a = E^{\dagger\nu}_a E^a_\alpha = \delta^\nu_\alpha\tau_0. \tag{2.4}$$

The vierbeins can be developed through the internal basis, for example, by taking $E^a_\mu(x)$ as

$$E^a_\mu = k^a_{\mu 0}(x)\tau_0 + k^a_{\mu i}(x)\tau_i$$

and

$$E^{\dagger a}_\mu = k^{*a}_{\mu 0}(x)\tau_0 + k^{*a}_{\mu i}(x)\tau_i, \tag{2.5}$$

since $\tau_i^\dagger = \tau_i$.

The transformation law for vectors on the tangent space is defined, as usual, through the Lorentzian rotation matrices $L^a{}_b$ such that $L^T\eta L = \eta$. Therefore, a more general transformation law for the matrix tangent vectors $E^a_\mu(x)$ shall now be

$$E'^{a}_{\mu}(x) = L^{a}{}_{b}(x)(U(n)E^{b}_{\mu}(x)U^{\dagger}(n)). \quad (2.6)$$

On this matrix tangent space we can now define the operation of covariant differentiation, for example, on the vector $E = (E^{\mu}_{a})$,

$$E^{\mu}_{a|\nu} = E^{\mu}_{a,\nu} + \Omega^{\mu}_{\rho\nu}E^{\rho}_{a} - \Lambda_{\nu}{}^{c}{}_{a}E^{\mu}_{c} + [\Gamma_{\nu}, E^{\nu}_{a}]. \quad (2.7)$$

It is important to remember that the space-time connection $\Omega^{\mu}{}_{\rho\nu}$ may include an (internal) complex connection related to the electromagnetic potential vector A_{ν} through relations (1.9) and (1.10): Using the notation of Ref. 6, here it will be called C_{ν}, which by (1.10) is given by

$$C_{\nu} = (2/ip)A_{\nu} = (ie/\hbar)A_{\nu}.$$

Therefore, the expressions corresponding to the field equation $G_{\underset{+}{\mu}\underset{-}{\nu}|\alpha} = 0$ (and its inverse $G^{\overset{\mu}{+}\overset{\nu}{-}}{}_{|\alpha} = 0$), for the matrix vierbeins, are as follows:

$$G_{\underset{+}{\mu}\underset{-}{\nu}|\alpha} = 0 \leftrightarrow E^{\dagger a}_{\underset{-}{\mu}|\alpha} = (E^{a}_{\underset{+}{\mu}|\alpha})^{\dagger} = 0,$$

$$E^{a}_{\underset{+}{\mu}|\alpha} = E^{a}_{\mu,\alpha} - E^{a}_{\rho}\Gamma^{\rho}{}_{\mu\alpha} + \Lambda_{\alpha}{}^{a}{}_{c}E^{c}_{\mu} = 0,$$

$$\Gamma^{\rho}{}_{\mu\alpha} = \theta^{\rho}{}_{\mu\alpha}\tau_0 + \delta^{\rho}_{\mu}\Gamma_{\alpha}, \quad \Gamma_{\alpha} = -(i\epsilon/\hbar)\mathbf{b}_{\alpha}\cdot\boldsymbol{\tau} = -\Gamma^{\dagger}_{\alpha},$$

$$\Lambda_{\alpha}{}^{a}{}_{c} = (\Lambda_{\alpha}{}^{a}{}_{c} + \delta^{a}_{c}C_{\alpha})\tau_0 + \delta^{a}_{c}\Gamma_{\alpha}; \quad (2.8)$$

$$G^{\overset{\mu}{+}\overset{\nu}{-}}{}_{|\alpha} = 0 \leftrightarrow E^{\dagger\overset{\mu}{+}}_{a|\alpha} = (E^{\overset{\mu}{-}}_{a|\alpha})^{\dagger} = 0,$$

$$E^{\overset{\mu}{-}}_{a|\alpha} = E^{\mu}_{a,\alpha} + E^{\rho}_{a}\Gamma^{\dagger\mu}{}_{\rho\alpha} - \Lambda^{\dagger}{}_{\alpha}{}^{c}{}_{a}E^{\mu}_{c} = 0,$$

$$\Gamma^{\dagger\mu}{}_{\rho\alpha} = \theta^{\mu}{}_{\alpha\rho}\tau_0 - \delta^{\mu}_{\rho}\Gamma_{\alpha}, \quad \text{because } \theta^{*\mu}{}_{\rho\alpha} = \theta^{\mu}{}_{\alpha\rho},$$

$$\Lambda^{\dagger}{}_{\alpha}{}^{a}{}_{b} = (\Lambda_{\alpha}{}^{a}{}_{b} - \delta^{a}_{b}C_{\alpha})\tau_0 - \delta^{a}_{b}\Gamma_{\alpha},$$

$$\text{because } C_{\alpha} = -(ie/\hbar)A_{\alpha} = -C^{*}_{\alpha}. \quad (2.9)$$

From (2.8) and (2.9) we can obtain a new expression for Λ_{γ} in terms of the matrix vierbeins

$$\Lambda_{\gamma}{}^{a}{}_{b} = E^{a}_{\mu}E^{\dagger\mu}_{b,\gamma} + E^{a}_{\mu}\Gamma^{\mu}{}_{\rho\gamma}E^{\dagger\rho}_{b}$$

$$= E^{a}_{\mu}E^{\dagger\overset{\mu}{+}}_{b;\gamma} + E^{a}_{\mu}\Gamma_{\gamma}E^{\dagger\mu}_{b} \quad (2.10)$$

and

$$\Lambda_{\gamma}{}^{a}{}_{b} = -E^{a}_{\mu,\gamma}E^{\dagger\mu}_{b} + E^{a}_{\rho}\Gamma^{\rho}{}_{\mu\gamma}E^{\dagger\mu}_{b}$$

$$= -E^{a}_{\underset{+}{\mu};\gamma}\,E^{\dagger\mu}_{b} + E^{a}_{\rho}\Gamma^{\rho}{}_{\mu\gamma}E^{\dagger\mu}_{b}. \quad (2.11)$$

The tangent space-time connection Λ_{γ} can then be written in this theory as

$$\Lambda_{\gamma}{}^{a}{}_{b} = \mathrm{Re}\{(1/n)\mathrm{Tr}[E^{a}_{\mu}E^{\dagger\overset{\mu}{+}}_{b;\gamma} + E^{a}_{\mu}\Gamma_{\gamma}E^{\dagger\mu}_{b}]\} \quad (2.12)$$

or

$$\Lambda_{\gamma}{}^{a}{}_{b} = \mathrm{Re}\{(1/n)\mathrm{Tr}[-E^{a}_{\underset{+}{\mu};\gamma}\,E^{\dagger\mu}_{b} + E^{a}_{\mu}\Gamma_{\gamma}E^{\dagger\mu}_{b}]\}. \quad (2.13)$$

The expression that relates the curvatures in the curved and tangent spaces is now

$$E^{a}_{\rho}\mathbf{R}^{\rho}{}_{\mu\nu\gamma} - \mathbf{S}_{\nu\gamma}{}^{a}{}_{c}E^{c}_{\mu} = 0, \quad (2.14)$$

where $\mathbf{R}^{\rho}_{\mu\nu\gamma}$ is the total curvature (1.19) written with the "connections" $\Gamma^{\rho}{}_{\mu\nu}$, and $\mathbf{S}_{\mu\nu}$ is the total curvature on the tangent space written with the "connections" Λ_{ν}:

$$\mathbf{S}_{\nu\gamma}{}^{a}{}_{c} = (\Lambda_{\nu,\gamma} - \Lambda_{\gamma,\nu} - [\Lambda_{\nu},\Lambda_{\gamma}])^{a}{}_{c}$$

$$= [S_{\nu\gamma}{}^{a}{}_{c} + \delta^{a}_{c}(C_{\nu,\gamma} - C_{\gamma,\nu})]\tau_0 + \delta^{a}_{c}P_{\nu\gamma}, \quad (2.15)$$

where $S_{\nu\gamma}$ is the curvature written with the tangent connection Λ_{γ} and $P_{\nu\gamma}$ is the internal curvature written for the internal connection Γ_{ν}. Also, the quanity $(C_{\nu,\gamma} - C_{\gamma,\nu})$ corresponds to the curvature of an internal (complex) space; here it is related to the electromagnetic tensor $F_{\nu\gamma}$.

III. THE GENERALIZATION OF THE FOCK–IVANENKO COEFFICIENTS

We can obtain a new generalized set of Dirac equations when we extend the treatment from the curved space-time of general relativity to the generalized matrix manifold. The anticommutation relations for the Dirac constant γ matrices[10] are

$$\{\gamma_a,\gamma_b\} = 2\eta_{ab}\mathbf{1}_4, \quad (3.1)$$

$$\{\gamma^a,\gamma^b\} = 2\eta^{ab}\mathbf{1}_4. \quad (3.2)$$

Multiplying (3.1) by E^{*a}_{ν} and E^{b}_{μ} and using (2.1), we obtain

$$\mathrm{Tr}\{\gamma_{\mu},\dot{\gamma}_{\nu}\} = 2\,\mathrm{Tr}(G_{\mu\nu})\mathbf{1}_4 = 2ng_{\mu\nu}\mathbf{1}_4, \quad (3.3)$$

where the Tr is taken on the n-dimensional matrix internal space and

$$E^{a}_{\mu}\gamma_a = \gamma_{\mu}, \quad E^{\dagger a}_{\mu}\gamma_a = \dot{\gamma}_{\mu}. \quad (3.4)$$

In (3.3), $g_{\mu\nu}$ is the metric of the ES nonsymmetric theory, by (1.3).

Analogously, multiplying (3.2) by $E^{\dagger\mu}_{a}$ and E^{ν}_{b} and taking the Tr over the internal n-dimensional matrices, we obtain

$$\mathrm{Tr}\{\dot{\gamma}^{\mu},\gamma^{\nu}\} = 2\,\mathrm{Tr}(G^{\mu\nu})\mathbf{1}_4 = 2ng^{\mu\nu}\mathbf{1}_4, \quad (3.5)$$

where

$$E^{\mu}_{a}\gamma^a = \gamma^{\mu}, \quad E^{\dagger\mu}_{a}\gamma^a = \dot{\gamma}^{\mu}, \quad (3.6)$$

and (2.3) was used. Considering the non-Riemannian manifold of the ES theory, the total covariant derivative of the new γ_{μ} is given by

$$\gamma_{\underset{+}{\mu}|\nu} = \gamma_{\mu,\nu} - \Omega^{\rho}{}_{\mu\nu}\gamma_{\rho} + [\Delta_{\nu},\gamma_{\mu}] + [\Gamma_{\nu},\gamma_{\mu}], \quad (3.7)$$

where Δ_{μ} is the internal connection corresponding to the space of the generalized γ matrices (or, also, the Dirac wavefunctions' space). Then taking (3.4) and (2.8), we have that

$$\gamma_{\underset{+}{\mu}|\nu} = (E^{a}_{\underset{+}{\mu}}\gamma_a)_{|\nu} = (E^{a}_{\underset{+}{\mu}|\nu})\gamma_a = 0, \quad (3.8)$$

since γ_a is a constant matrix. In the same way, we obtain

$$\dot{\gamma}_{\underset{-}{\mu}|\nu} = (E^{\dagger a}_{\mu}\gamma_a)_{|\nu} = (E^{\dagger a}_{\underset{-}{\mu}|\nu})\gamma_a = 0. \quad (3.9)$$

Expanding (3.8) and (3.9) we arrive at

$$\gamma_{\underset{+}{\mu}|\nu} = \gamma_{\mu,\nu} - \theta^{\rho}{}_{\mu\nu}\gamma_{\rho} + C_{\nu}\gamma_{\mu} + [\Delta_{\nu},\gamma_{\mu}] + [\Gamma_{\nu},\gamma_{\mu}] = 0 \quad (3.10)$$

and

$$\gamma_{\underset{-}{\mu}|\nu} = \dot{\gamma}_{\mu,\nu} - \theta^{\rho}{}_{\nu\mu}\dot{\gamma}_{\rho} - C_{\nu}\dot{\gamma}_{\mu} + [\Delta_{\nu},\dot{\gamma}_{\mu}] + [\Gamma_{\nu},\dot{\gamma}_{\mu}] = 0. \quad (3.11)$$

Therefore, we can obtain an expression for Δ_{ν}:

$$\Delta_{\nu} = (1/4i)\Lambda^{ab}_{\nu}\sigma_{ab}, \quad (3.12)$$

where Λ_{ν} is given in (2.12) or (2.13). Equation (3.12) is similar to the corresponding equation in general relativity.[11]

We can then use a minimal action principle to obtain field equations for a spin-$\frac{1}{2}$ particle of mass m, where the wavefunction is $\psi(x)$, placed in a non-Riemannian manifold of the ES theory and also under the influence of a (n-dimensional) Yang–Mills field. The action for this situation is

$$A=\int \mathscr{L}\, d^4x, \tag{3.13}$$

where the Lagrangian is given by

$$\mathscr{L}=\sqrt{-w}\{\bar{\psi}\gamma^\mu[\vec{\partial}_\mu+\Delta_\mu+C_\mu+\Gamma_\mu]\psi + \bar{\psi}[\overleftarrow{\partial}_\mu+\Delta_\mu-C_\mu-\Gamma_\mu]\psi\gamma^\mu-\mu\bar{\psi}\psi\}, \tag{3.14}$$

where μ is the mass term and the Tr is taken on the internal n-dimensional space. The function $\psi(x)$ is a complex object which locally transforms under the representation of the Lorentz group ($U(L)$), but also transforms under the (internal) $SU(n)$ group. The field equations obtained are

$$\gamma^\mu[\vec{\partial}_\mu+\Delta_\mu+C_\mu+\Gamma_\mu]\psi-\mu\psi=0 \tag{3.15}$$

and

$$-\bar{\psi}[\overleftarrow{\partial}_\mu+\Delta_\mu-C_\mu-\Gamma_\mu]\gamma^\mu-\mu\bar{\psi}=0, \tag{3.16}$$

where again, $\bar{\psi}(x)=\psi^\dagger(x)\gamma_0$. [It is important to note that the Dirac equations derived here are formally similar to those obtained by Borchsenius[3] when they are rewritten in terms of γ-Dirac matrices. However, the similarity ends at this point since Borchsenius did not define a more general $\sigma_p^{k\nu}$ which includes the new "internal" SU(2) degrees of freedom of the theory. In fact, that would be achieved in an expression similar to (5.10) of Ref. 3 with an additional symmetry property such as (1.4) given in this work.] Also, we can find an expression for the "charge conjugate" wavefunction ψ^c, which is

$$\gamma^\mu[\vec{\partial}_\mu+\Delta_\mu-C_\mu-\Gamma_\mu]\psi^c-\mu\psi^c=0, \tag{3.17}$$

where $\psi^c=C\bar{\psi}^T$ and C is the charge conjugation matrix.

IV. INCLUSION OF INTERNAL MASS TERMS

We are now going to analyze the case of an extended mass term, where we suppose there is nonzero mass on the internal space, i.e., we will suppose that for each internal axis we have a different mass term.

The n^2-dimensional vierbein $E^\mu_a(x)$ can be written as in (2.5):

$$E^\mu_a(x)=k^\mu_{a0}(x)\tau_0+k^\mu_{ai}(x)\tau_i,\quad i=1,\ldots,n^2-1. \tag{4.1}$$

Suppose that the mass term μ is a matrixlike term:

$$\mu=\mu_0\tau_0+\mu_i\tau_i. \tag{4.2}$$

We also are going to assume here that

$$k^\mu_{a0}=k^\mu_{a0R}+ik^\mu_{a0I}, \tag{4.3}$$

$$\mu_0=\mu_{0R}+i\mu_{0I}, \tag{4.4}$$

and k^μ_{ai} and μ_i are pure imaginary numbers. These hypotheses are consistent with the form of the metric defined in (1.6) and (1.7) and the definition of the matrix vierbeins in (2.5).

Define

$$k^\mu_{a0I}=p\lambda n^\mu_{a0},\quad \mu_{0I}=p\lambda m_0, \tag{4.5}$$

$$k^\mu_{ai}=i(p\lambda)^2 n^\mu_{ai},\quad \mu_i=i(p\lambda)^2 m_i, \tag{4.6}$$

where now p is being considered a parameter and λ is a constant with the value of the $1/p$, as in Ref 7:

$$\lambda\sim e/2\hbar=2.58\times10^{32}\ \mathrm{cm}^{-1},$$

where the maximal value for $|p|$ was taken, $|p|=|-2\hbar/e|=3.8\times10^{-32}$ cm, in the normalization used in Ref. 3.

Placing the above quantities into the Dirac equation (2.29), we can expand it as

$$\tau_0[k^\mu_{a0R}\gamma^a\nabla_\mu\psi-\mu_{0R}\psi]+ip\lambda\tau_0[\eta^\mu_a 0\gamma^a\nabla_\mu\psi-m_0\psi] + i(p\lambda)^2\tau_i[\eta^\mu_{ai}\gamma^a\nabla_\mu\psi-m_i\psi]=0, \tag{4.7}$$

where $\nabla_\mu=\partial_\mu+\Delta_\mu+C_\mu+\Gamma_\mu$. In the limit of the parameter $p\to0$, we should obtain the standard Dirac equation in the presence of gravitation, electromagnetism, and Yang–Mills fields. Consequently, we can obtain n^2+1 other sets of Dirac equations when we take $n^\mu_{a0}=n^\mu_{ai}=k^\mu_{a0R}\sim h^\mu_a$ and $m_0=m_i=\mu_{0R}$ for each i and where h^μ_a and μ_{0R} are taken as the vierbeins and mass term in the general relativity theory.

Therefore, the above analysis results in some sort of "projections" of the Dirac equation on the internal space which are due to the definition of more general vierbeins through (2.1). The value of the parameter p will then determine the amplitude of those projections through (4.5) and (4.6).

V. CONCLUSION

By taking the complex manifold of the ES nonsymmetrical theory and adding to it an n^2-dimensional internal space, it is possible to develop a generalized theory that in the case chosen here, where we used the $SU(n)$ symmetry group, permitted us to include the $SU(n)$ Yang–Mills field. It is also possible to obtain the tangent space local to the extended manifold. Then the corresponding generalized Dirac theory, as well as the generalized Dirac field equation, were developed. In the case of an extended mass term, where we assume there is nonzero mass on each internal axis, and defining the internal components of the vierbeins, as well as the internal components of the mass term, as being proportional to the parameter p, we obtained n^2 other sets of Dirac equations which are some sort of projections of the standard Dirac equation on the internal space. In this theory the value of the parameter p determines the amplitude of these projections through (4.5) and (4.6). In the limit of the parameter $p\to0$, the standard situation of the general relativity theory is obtained. If the Yang–Mills field is not present, this theory reduces to the "complex theory," which is presented in Ref. 7 as $\mathbf{f}_{\mu\nu}=0$ in (1.6) and (1.8) and $\mathbf{b}_\mu=0$ in (3.15) and (3.17). Also, in this case, the components of the wavefunction on the internal space generated through the the internal symmetry group ($SU(n)$) are null.

A question arises at this point: Where would a theory like this be consistent with the real world? We could just say that this should happen in regions of the space-time with high intensity fields (gravitation, electromagnetism, or Yang–Mills fields) and at distances of the order of the Planck length, where it would be reasonable to think of a nonzero p and the consequences of a more complex theory such as the one used in this work.

The present theory can be easily interpreted through a quaternionic theory in the case of $n = 2$. This will enable us to extend it to an octonionic theory, which would be convenient in this case since we are using a complex nonsymmetrical manifold. This is permitted by the theorem of Hurwitz. Thinking from this point of view, the gauge on the Dirac equation in a real manifold would just be one corresponding to the gravitational gauge. The electromagnetic gauge on the Dirac equation would be included when we consider the space-time manifold extended to a complex manifold. The Yang–Mills gauge would then be included when we extend the manifold to the matrix manifold, which is equivalent to the quaternionic manifold for the SU(2) symmetry group. The next step would then be to extend the quaternionic theory to the octonionic one and determine to which gauge it corresponds. This is the goal we will propose in the third part of this work for the analysis of the Dirac equation in a non-Riemannian manifold.

ACKNOWLEDGMENTS

This work was supported by CAPES-MEC, Brazil. This work was also supported in part by the DOE and by NASA under Grant No. NAGW-1340 at Fermilab.

[1]A. Einstein, Rev. Mod. Phys. **20**, 35 (1948); A. Einstein, *The Meaning of Relativity* (Princeton U.P., Princeton, NJ, 1955), Appendix 2; E. Schrödinger, *Space-Time Structures* (Cambridge U.P., London, 1954); M. A. Tonellating, *Les Théories Unitaires de l'Électromagnétisme et de la Gravitation* (Gauthier Villars, Paris, 1965).

[2]J. W. Moffat and D. H. Boal, Phys. Rev. D **11**, 1375 (1975); W. B. Bonnor, Proc. R. Soc. London Ser. A **226**, 336 (1954).

[3]K. Borchsenius, Phys. Rev. D **13**, 2707 (1976); C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954); see, also, J. W. Maluf and S. Okubo, Phys. Rev. D **31**, 1327 (1975).

[4]G. Kunstater, P. Leivo, and P. Savaria, Quant. Grav. **1**, 7 (1984); R. B. Mann and J. W.. Moffat, Phys. Rev. D **26**, 1858 (1982); C. R. Johnson, Phys. Rev. D **12**, 3831 (1975).

[5]See K. A. Zhevlakov, A. M. Slin'ko, I. P. Shestakov, and A. I. Shirshov, *Rings that are Nearly Associative* (Academic, New York, 1982), Chap. 2; see, also, M. Gunaydin and F. Gursey, J. Math. Phys. **14**, 1651 (1973) and S. Marques and C. G. Oliveira, J. Math. Phys. **26**, 3131 (1985).

[6]S. Marques and C. G. Oliveira (Ref. 5); S. Marques and C. G. Oliveira, Phys. Rev. D **36**, 1716 (1987).

[7]S. Marques, J. Math. Phys. **29**, 2127 (1988).

[8]D. R. Brill and J. A. Wheeler, Rev. Phys. **29**, 465 (1957); D. R. Brill and J. M. Cohen, J. Math. Phys. **7**, 238 (1966).

[9]D. W. Sciama, J. Math. Phys. **2**, 472 (1961); P. D. P. Smith, Nuovo Cimento B **24**, 255 (1974).

[10]P. A. M. Dirac, Proc. R. Soc. London Ser. A **117**, 610 (1928).

[11]V. Fock and D. Yvanenko, C. R. Acad. Sci. **188**, 1470 (1928); V. Fock, Z. Phys. **57**, 261 (1929); see, also, J. L. Anderson, *Principles of Relativity* (Academic, New York, 1967), p. 360.

Exact solution of the Dirac equation in a reducible Einstein space

Víctor M. Villalba[a)]
Centro de Física, Instituto Venezolano de Investigaciones Científicas (I.V.I.C.), Apdo. 21827, Caracas 1020A, Venezuela

(Received 25 October 1988; accepted for publication 24 January 1990)

In this paper, an exact solution of the Dirac equation in a static reducible Einstein space is presented. The asymptotic behavior of the spinor solution is analyzed.

I. INTRODUCTION

The study of the Dirac equation in external gravitational fields has been the object of detailed analysis, and recently some exact solutions have been obtained using different techniques and methods.[1–3] One of the most effective and powerful tools in solving systems of partial differential equations and in particular the Dirac equation is the method of separation of variables,[4–6] which allows us to reduce the problem to solving a system of ordinary differential equations. Recently, the diagonal metrics for which the Dirac equation admits separation of variables have been classified[5] and some exact solutions in cosmological universes have been obtained using this classification.[7–9]

The study of the behavior of relativistic particles obeying the Dirac equation in curved spaces, in particular in expanding universes, is of considerable importance in astrophysics and cosmology.[10,11] Such investigations enable us to quantize the relativistic spin-$\frac{1}{2}$ electron field in curved backgrounds and study the effect of gravity in atomic spectra.[12] The presence of the cosmological constant in Einstein equations allows us to consider, as background fields, the so-called Einstein spaces (such spaces are symmetric spaces where the Ricci tensor is proportional to the metric tensor). Among such spaces with pseudo-Riemannian signature we find the reducible Einstein spaces, where the metric tensor can be diagonalized as a sum of two independent two-dimensional metrics.[13] This space, after the substitution $t' = it$, becomes a four-dimensional Euclidean Einstein manifold, which might be expected to be important in the Euclidean path integral formulation of gravity.[14,15] In the present paper, we solve the Dirac equation in this reducible space-time via the separation of variables.

This paper is organized as follows: In Sec. II, the explicit form of the symmetric reducible Einstein space is computed. In Sec. III, we solve the Dirac equation in the metric obtained in Sec. II via the separation of variables. In Sec. IV, we present an analysis of the asymptotic behavior of the Dirac spinor.

II. COMPUTATION OF THE METRIC

Let us consider the interval

$$ds^2 = dx^2 + a^2(x)dy^2 + dz^2 - b^2(z)dt^2, \qquad (2.1)$$

where the spatial variables are x, y, z and t is the time.

It is easy to see from expression (2.1) that the metric tensor can be written as the sum of two two-dimensional metrics. To calculate the Ricci tensor, we consider the following basis one-forms:

$$\theta^1 = dx, \quad \theta^2 = a(x)dy, \quad \theta^3 = dz, \quad \theta^0 = b(z)dt, \qquad (2.2)$$

so that

$$\eta_{\mu\nu} = \text{diag}\{-1, +1, +1, +1\}. \qquad (2.3)$$

Taking the exterior differentials of (2.2) and using the first equation of structure

$$d\theta^\alpha + \omega^\alpha_\beta \wedge \theta^\beta = 0, \qquad (2.4)$$

we obtain the following nonzero connection one-forms:

$$\omega^1_2 = -\omega^2_1 = -a_{,x}/a\theta^2, \quad \omega^0_3 = \omega^3_0 = b_{,z}/b\theta^2, \qquad (2.5)$$

where a comma indicates ordinary differentiation. We obtain the curvature two-forms from the second equation of structure

$$\Omega^\alpha_\beta = d\omega^\alpha_\beta + \omega^\alpha_\mu \wedge \omega^\mu_\beta. \qquad (2.6)$$

We obtain the components of the Riemann tensor, making the identification

$$\Omega^\alpha_\beta = \tfrac{1}{2}R^{\ \alpha}_{\beta\mu\nu}\theta^\mu \wedge \theta^\nu \qquad (2.7)$$

in Eq. (2.6)

The nonvanishing components of Ω^α_β and $R^{\ \alpha}_{\beta}$ are found to be

$$\Omega^1_2 = -a_{,xx}/a\theta^1 \wedge \theta^2, \qquad (2.8)$$

$$\Omega^0_3 = b_{,zz}/b\theta^0 \wedge \theta^3, \qquad (2.9)$$

$$R_{00} = -R_{33} = b_{,zz}/b, \qquad (2.10)$$

$$R_{11} = R_{22} = -a_{,xx}/a. \qquad (2.11)$$

Einstein equations with the cosmological constant λ read as

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R + \lambda g_{\mu\nu} = 0. \qquad (2.12)$$

The solutions of (2.12) are the so-called Einstein spaces, which are chracterized by the property that the Ricci tensor is proportional to the metric tensor, that is,

$$R_{\mu\nu} = \chi g_{\mu\nu}. \qquad (2.13)$$

Substituting (2.13) into (2.12), we determine that $\lambda = \chi$ and

$$R_{,\mu} = 0. \qquad (2.14)$$

Relation (2.14) defines a symmetric space. Substituting (2.1) into (2.9), we obtain

$$b_{,zz} + \lambda b = 0, \qquad (2.15)$$

$$a_{,xx} + \lambda a = 0. \qquad (2.16)$$

The solution of Eqs. (2.15) and (2.16) is

$$a = \cos\sqrt{\lambda}\,x, \qquad (2.17)$$

$$b = \cos\sqrt{\lambda}\,z, \qquad (2.18)$$

[a)] Postal address: Carmelitas 4282, Caracas 1010A, Venezuela.

where we have imposed the condition $a(0) = b(0) = 1$. Then the interval (2.1), solution of (2.12), takes the form

$$ds^2 = dx^2 + \cos^2(\sqrt{\lambda}\, x)dy^2 + dz^2 - \cos^2(\sqrt{\lambda}\, z)dz^2 . \tag{2.19}$$

The metric (2.19) can be rewritten in a more familiar form if we make the change of variables

$$x' = \sqrt{\lambda}\, x + \pi/2, \quad y' = \sqrt{\lambda}\, y, \tag{2.20}$$

and

$$r = (1/\sqrt{\lambda})\sin\sqrt{\lambda}\, z . \tag{2.21}$$

Then, in the new coordinates x', y', r, t, the interval (2.19) reads

$$ds^2 = \frac{1}{\sqrt{\lambda}}(dx'^2 + \sin^2(x')dy'^2) + \frac{dr^2}{1-\lambda r^2} - (1-\lambda r^2)dt^2 . \tag{2.22}$$

Notice that if we make the substitution $i\tau = t$ in expression (2.22), we obtain an $S^2 \times S^2$ metric product of two two-spheres, each with radius $\lambda^{-1/2}$ and area $4\pi\lambda^{-1}$. It can be regarded as a limiting case of the Schwarzschild–de Sitter solution representing a gravitational instanton.[16,17]

In order to carry out the separation of variables in a more simple way, it is convenient to redefine the coordinates t and z as follows:

$$\sqrt{\lambda}\, z - \pi/2 = z', \quad \sqrt{\lambda}\, t = t' . \tag{2.23}$$

Then the interval (2.19) takes the form

$$ds^2 = (1/\sqrt{\lambda})[dx'^2 + [\sin^2 x']dy'^2 + dz'^2 - [\sin^2 z']dt'^2] \tag{2.24}$$

and the radial variable r [(2.21)] reads as

$$r = (1/\sqrt{\lambda})\cos z' . \tag{2.25}$$

III. SEPARATION OF VARIABLES IN THE DIRAC EQUATION

In this section, we shall solve the Dirac equation in the metric (2.24) via the separation of variables.

The equation describing a relativistic electron in curved space-time is given by[10]

$$[\bar{\gamma}^\alpha(\partial_\alpha - \Gamma_\alpha) + m]\psi = 0, \tag{3.1}$$

where Γ_α are the spin connections

$$\Gamma_\alpha = -\tfrac{1}{4} g_{\alpha\beta}\Gamma^\alpha_{\nu\mu} S^{\beta\nu} \tag{3.2}$$

with

$$S^{\beta\nu} = \tfrac{1}{2}(\bar{\gamma}^\beta\bar{\gamma}^\nu - \bar{\gamma}^\nu\bar{\gamma}^\beta) . \tag{3.3}$$

The $\bar{\gamma}^\beta$ are generalized Dirac matrices and are related to the flat space-time gammas as follows:

$$\bar{\gamma}^1 = \gamma^1\sqrt{\lambda}, \quad \bar{\gamma}^2 = (\sqrt{\lambda}/c)\gamma^2, \quad \bar{\gamma}^3 = \gamma^3\sqrt{\lambda}, \quad \bar{\gamma}^0 = (\sqrt{\lambda}/d)\gamma^0, \tag{3.4}$$

with $c = \sin x'$, $d = \sin z'$, and

$$\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\eta^{\alpha\beta}, \quad \eta^{\alpha\beta} = \mathrm{diag}(-1,1,1,1) . \tag{3.5}$$

Substituting the metric (2.19) into the equation for the spin connections (3.2) and using the Dirac matrices (3.4), we obtain

$$\Gamma_0 = \tfrac{1}{2} d_{,z}\gamma^0\gamma^3, \quad \Gamma_1 = 0, \quad \Gamma_2 = \tfrac{1}{2} c_{,x}\gamma^1\gamma^2, \quad \Gamma_3 = 0 . \tag{3.6}$$

Substituting results (3.6) into (3.1), we obtain

$$((\gamma^0/d)\partial_0 + \gamma^1\partial_1 + (\gamma^2/c)\partial_2 + \gamma^3\partial_3 + m)\Phi = 0, \tag{3.7}$$

where we have made the identification

$$\Psi = (c\ d)^{-1/2}\Phi . \tag{3.8}$$

Since the metric associated to the interval (2.24) is a function of x' and z' only, we can set

$$\Phi = \Phi(x,t)\cdot\exp i(k_y y' - E't') . \tag{3.9}$$

The remaining variables in (3.7) can be separated by writing Eq. (3.7) as a sum of two first-order differential operators, commuting between them as follows:

$$K_1 = -i(\gamma^3\partial_3 - i\gamma^0 E/d + m)\gamma^0\gamma^3, \tag{3.10}$$

$$K_2 = -i(\gamma^1\partial_1 + i\gamma^2 k_y/c)\gamma^0\gamma^0, \tag{3.11}$$

with

$$K_1\Phi = -K_2\Phi = k\Phi, \quad [K_1,K_2] = 0, \quad \bar{\Phi} = i\gamma^0\gamma^3\Phi, \tag{3.12}$$

where k is a constant of separation.

Adopting the conventions of Jauch and Rohrlich[18] for the Dirac matrices, Eq. (3.11) becomes

$$[\sigma^2\partial_1 - ik_y/c\sigma^1 + ik]\Phi_1 = 0, \tag{3.13}$$

$$[-\sigma^2\partial_1 + ik_y/c\sigma^1 + ik]\Phi_2 = 0, \tag{3.14}$$

where

$$\bar{\Phi} = \begin{pmatrix}\Phi_1\\ \Phi_2\end{pmatrix} .$$

From Eqs. (3.13) and (3.14), it is clear that Φ_1 and Φ_2 are related as follows:

$$\Phi_2 = f(x^3)\sigma^3\Phi_1 . \tag{3.15}$$

Equation (3.10) takes a more symmetric form if we put $i\gamma^1$, γ^2 instead of γ^0, γ^3 respectively: This is possible after the transformation

$$\gamma^\mu \to S^{-1}\gamma^\mu S, \quad \Phi \to S^{-1}\Phi, \tag{3.16}$$

where the matrix S is given by

$$S = \tfrac{1}{2}(1 + \gamma^3\gamma^2)(1 + i\gamma^0\gamma^1) . \tag{3.17}$$

Using expressions (3.13), (3.14), and (3.17) and the Dirac matrices in the representation,[18] we find that the spinor $\bar{\Phi}$ has the following structure:

$$\bar{\Phi} = \begin{pmatrix}(\alpha - i\beta)\xi\\ (\alpha - i\beta)\zeta\\ i(\alpha + i\beta)\xi\\ -i(\alpha + i\beta)\zeta\end{pmatrix} \exp i(k_y - E't'), \tag{3.18}$$

where the functions α, β, ζ, ξ satisfy two systems of coupled differential equations given by

$$\beta_{,3} - \frac{iE}{d}\beta = -(m + ik)\alpha, \tag{3.19}$$

$$\alpha_{,3} + \frac{iE}{d}\alpha = -(m - ik)\beta, \tag{3.20}$$

$$\zeta_{,1} + k_y/c\zeta - k\xi = 0, \tag{3.21}$$

$$\xi_{,1} - k_y/c\xi + k\zeta = 0 . \tag{3.22}$$

Here, it is enough to consider the solution to the above system of equations (3.19)–(3.22) when k_y and k are positive because the other three cases can be obtained by interchanging the roles of ξ and ζ. The ansatz

$$\xi = \sin^{k_y}(x')\cos(x'/2)f(x'),\quad \zeta = \sin^{k_y}(x')\sin(x'/2)g(x') \tag{3.23}$$

lead to

$$\left[(q+1)\frac{d}{dq} + \left(\frac{1}{2}+k_y\right)\right]f = kg, \tag{3.24}$$

$$\left[(q-1)\frac{d}{dq} + \left(\frac{1}{2}+k_y\right)\right]g = kf, \tag{3.25}$$

where

$$q = \cos x'.$$

Eliminating f from (3.24) and (3.25), we obtain

$$\left\{(1-q^2)\frac{d^2}{dq^2} + \left(-1-2q(1+k_y)\right)\frac{d}{dq} - \left[\left(\frac{1}{2}+k_y\right)^2 - k^2\right]\right\}g = 0. \tag{3.26}$$

The solution of (3.26), taking into account regularity at $q=0$, is given in terms of the Jacobi polynomials $P_n^{(\alpha,\beta)}$ (Refs. 19–21):

$$g = cP_n^{(k_y+1/2,k_y-1/2)}(q), \tag{3.27}$$

where c is a constant of normalization and n is given by

$$n = k - k_y - \tfrac{1}{2}. \tag{3.28}$$

Substituting (3.27) into (3.25) and using the recurrence relation for the Jacobi polynomials[19] we obtain the expression for f. Therefore, the functions ξ and ζ are

$$\xi = c\sin^{k_y}(x')\cos(x'/2)P_n^{(k_y+1/2,k_y-1/2)}(\cos x'), \tag{3.29}$$
$$\zeta = c\sin^{k_y}(x')\sin(x'/2)P_n^{(k_y-1/2,k_y+1/2)}(\cos x').$$

The block structure of the spinor $\overline{\Phi}$, given by (3.18), allows us to consider the following condition of normalization for the contribution of the variables y' and z' of the solution to the Dirac equation:

$$2\pi\int_0^\pi [\xi_k^*\xi_{k'} + \zeta_k^*\zeta_{k'}]d\theta = \delta_{k,k'}. \tag{3.30}$$

Substituting (3.29) into (3.30) and considering the relation[19]

$$\frac{2^{\alpha+\beta+1}(n+\alpha)!(n+\beta)!}{n!(2n+\alpha+\beta+1)!(n+\alpha+\beta)\}!} = \int_{-1}^{+1}(P_n^{(\alpha,\beta)}(z))^2(1-z)^\alpha(1+z)^\beta dz, \tag{3.31}$$

we determine that the constant c is given by

$$c = [(k-k_y-\tfrac{1}{2})!(k-k_y+\tfrac{1}{2})!]^{1/2}/(k-1)!2^{k_y}(2\pi)^{1/2}. \tag{3.32}$$

The second system of coupled equations (3.19) and (3.20) can be solved in a way similar to that used to obtain (3.29). Then let us consider the following expressions for α and β:

$$\alpha = \sin^{iE'}(z')\cos(z'/2)I(z'),\quad \beta = \sin^{iE'}(z')\sin(z'/2)J(z'), \tag{3.33}$$

where the functions I and J satisfy the system of equations

$$(q+1)\frac{dI}{dq} + (iE'+\tfrac{1}{2})I = (m-ik)J,\quad (q-1)\frac{dJ}{dq} + (iE'+\tfrac{1}{2})J = -(m+ik)I, \tag{3.34}$$

with

$$q = \cos z'. \tag{3.35}$$

The solution of system (3.34) and (3.35) can be expressed in terms of the hypergeometric functions $F(a,b,\gamma,z)$ as follows:

$$\begin{aligned} I &= [(m-ik)/(iE'+\tfrac{1}{2})]sF(a,b,iE'+\tfrac{3}{2},w) + nw^{-(iE'+1/2)}F(a-iE'-\tfrac{1}{2},b-iE'-\tfrac{1}{2},\tfrac{1}{2}-iE',w), \\ J &= sF(a,b,iE'+\tfrac{1}{2},w) + [(m+ik)/(\tfrac{1}{2}-iE')] \times nw^{(1/2-iE')}F(a-iE'+\tfrac{1}{2},b-iE'+\tfrac{1}{2},\tfrac{3}{2}-iE',w), \end{aligned} \tag{3.36}$$

where

$$a = iE'+\tfrac{1}{2}+i\sqrt{m^2+k^2},\quad b = iE'+\tfrac{1}{2}-i\sqrt{m^2+k^2}, \tag{3.37}$$

and w is given by

$$w = (q+1)/2 = \cos^2(z'/2). \tag{3.38}$$

Then, substituting (3.36) into (3.33), we obtain the solution of the system (3.19) and (3.20). It should be noted that the functions I and J do not reduce to Jacobi polynomials by virtue of the specific values of the arguments α and β in (3.37).

Finally, we can write the solution of (3.7) by substituting into (3.18) the explicit expressions of ζ, ξ and α, β given by (3.29) and (3.33), respectively. By virtue of the matrix transformation (3.12) relating Φ and $\overline{\Phi}$ and relation (3.8), we obtain

$$\psi = (\sin x'\sin z')^{-1/2}\begin{pmatrix} i(\alpha+i\beta)\xi \\ i(\alpha+i\beta)\zeta \\ -(\alpha-i\beta)\xi \\ (\alpha-i\beta)\zeta \end{pmatrix}\exp i(k_y - E't'). \tag{3.39}$$

IV. ANALYSIS OF THE SOLUTION

In order to analyze the behavior of the spinor solution Ψ in the background field obtained in Sec. II, it is convenient to express the metric in the form (2.22), where the presence of a singularity at the value $r_h = 1/\sqrt{\lambda}$ is evident. It is the purpose of the present section to carry out the asymptotic study of the radial functions α and β at the singularity r_h, where the coordinate r is defined by Eq. (2.21).

In order to find the asymptotic form of the spinor Ψ[(3.39)], we must express the argument w in terms of the radial variable r. Then we have

$$w = (1+\sqrt{\lambda}r)/2. \tag{4.1}$$

Notice that solutions (3.36) are given in terms of hypergeometric functions defined over a domain that lies inside the unit circle and w goes to 1 when $r \to r_h$. Then, taking into account the value of the limit of the function $F(a,b,\gamma,z)$ as $z \to 1^-$, we have

$$\lim_{z\to 1} F(a,b,\gamma,z) = \Gamma(\gamma)\Gamma(\gamma-a-b)/\Gamma(\gamma-a)\Gamma(\gamma-b) . \quad (4.2)$$

From (2.35), it is clear that the following relation takes place:

$$\lim_{z'\to 0} \sin^{iE'}(z') = \lim_{r\to r_h} (1-\lambda r^2)^{iE'/2} = 2^{iE'} e^{-iE' r^* \sqrt{\lambda}} , \quad (4.3)$$

where we have defined the new "turtle" coordinate r^* as

$$r^* = (1/2\sqrt{\lambda}) \log((1+\sqrt{\lambda}\, r)/(1-\sqrt{\lambda}\, r)) , \quad (4.4)$$

From expression (4.4), it is clear that $r^* \to +\infty$ when r approaches the singularity at r_h. Then, substituting into (3.36) the limits (4.2) and (4.3), we have that the components α and β of the spinor Ψ present the following asymptotic behavior at $r = r_h$:

$$\alpha \to e^{-iE' r^*}\left((m-ik)s \frac{\Gamma(iE'+\frac{1}{2})\Gamma(\frac{1}{2}-iE')}{\Gamma(1-i\sqrt{m^2+k^2})\Gamma(1+i\sqrt{m^2+k^2})} + n \frac{(\Gamma(\frac{1}{2}-iE'))^2}{\Gamma(\frac{1}{2}-iE'-i\sqrt{m^2+k^2})\Gamma(\frac{1}{2}-iE'+i\sqrt{m^2+k^2})}\right), \quad (4.5)$$

$$\beta \to e^{iE' r^*}\left(n(m+ik) \frac{\Gamma(iE'+\frac{1}{2})\Gamma(\frac{1}{2}-iE')}{\Gamma(1-i\sqrt{m^2+k^2})\Gamma(1+i\sqrt{m^2+k^2})} + s \frac{(\Gamma(\frac{1}{2}+iE'))^2}{\Gamma(\frac{1}{2}+iE'-i\sqrt{m^2+k^2})\Gamma(\frac{1}{2}+iE'+i\sqrt{m^2+k^2})}\right). \quad (4.6)$$

Finally, it should be mentioned that, from expressions (4.5), (4.6), and (3.39), it is clear that each component of the spinor solution Ψ to the Dirac equation (3.1) behaves, at singularity $r = r_h$, as a superposition of an incoming and an outgoing wave.

ACKNOWLEDGMENT

I am indebted to Umberto Percoco for his careful reading of the final draft and his enlightening suggestions.

[1] A. O. Barut and I. H. Duru, Phys. Rev. D **36**, 3705 (1987).
[2] M. A. Castagnino, C. D. El Hasi, F. D. Mazzitelli, and J. P. Paz, Phys. Lett. A **128**, 25 (1988).
[3] L. P. Chimento and M. S. Mollerah, Phys. Rev. D **34**, 3689 (1986).
[4] E. G. Kalnins, W. Miller, Jr., and G. C. Williams, J. Math. Phys. **27**, 1893 (1986).
[5] I. E. Andrushkevich and G. V. Shishkin, Theor. Math. Phys. **70**, 289 (1987).
[6] G. V. Shishkin and V. M. Villalba, J. Math. Phys. **30**, 2132 (1989).
[7] G. V. Shishkin and I. E. Andrushkevich, Phys. Lett. A **110**, 84 (1985).
[8] G. V. Shishkin and I. E. Andrushkevich, Vestsi. AN. BSSR. Ser. Fiz. Mat. Navuk **4**, 109 (1986).
[9] V. M. Villalba and U. Percoco, J. Math. Phys. **31**, 715 (1990).
[10] D. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).
[11] W. Unruh, Phys. Rev. Lett. **31**, 1265 (1973).
[12] L. Parker, Phys. Rev. Lett. **44**, 1559 (1980).
[13] A. Z. Petrov, *Einstein Spaces* (Pergamon, London, 1969).
[14] G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **66**, 267 (1979).
[15] G. W. Gibbons and S. W. Hawking, Commun. Math. Phys. **66**, 291 (1979).
[16] G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **61**, 239 (1978).
[17] T. Eguchi, P. G. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).
[18] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976).
[19] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[20] M. M. Nieto, Phys. Rev. A **17**, 1273 (1978).
[21] I. H. Duru, Phys. Rev. D **30**, 2121 (1984).

Covariant quantization of gauge theories in the framework of extended BRST symmetry

I. A. Batalin
Lebedev Physical Institute, Moscow 117924, U.S.S.R.
P. M. Lavrov
Tomsk State Pedagogical Institute, Tomsk 634041, U.S.S.R.
I. V. Tyutin
Moscow Institute of Electronic Engineering, Moscow 109028, U.S.S.R.

(Received 3 November 1989; accepted for publication 3 January 1990)

The quantization rules for gauge theories in the Lagrangian formalism are formulated on the basis of the requirement of an extended BRST symmetry. The independence of the S matrix to the choice of a gauge is proved. The Ward identities are derived, and the existence theorem for the solutions of the generating equations within the given formalism is proved. Rank 1 gauge theories are considered as an example.

I. INTRODUCTION

The modern quantization method for gauge theories in the Lagrangian formalism[1,2] is based on the idea of a special type of global supersymmetry, the so-called BRST symmetry.[3,4] BRST symmetry implies invariance of the resultant action of a gauge theory under global nilpotent transformations of dynamical variables and results from gauge invariance of the initial action of the classical theory.

The requirement of BRST symmetry in gauge theories can be appreciably strengthened by the requirement that the resultant action be invariant not only under BRST transformations but also under the so-called anti-BRST transformations introduced in Refs. 5 and 6. According to the conventional terminology, BRST and anti-BRST symmetries together are referred to as "the extended BRST symmetry." A fairly large number of papers is devoted to the various aspects connected with the extended BRST symmetry in gauge theories.[7–16] The general formalism of the Hamiltonian BFV (Batalin, Fradkin, Vilkovisky) approach with extended BRST symmetry is developed in Ref. 17. But a consistent formulation of quantization of general gauge theories in the Langrangian formalism based on the extended BRST symmetry principle has not yet been found. In the present paper we suggest (Sec. II) the version of covariant quantization of general gauge theories in the framework of extended BRST symmetry, prove gauge invariance of the S matrix, derive Ward identities, prove the existence theorem for solutions of the generating equations (Sec. III), and consider gauge theories of rank 1 (Sec. IV).

We have used the condensed notations[18]; if not otherwise specified, the derivatives with respect to fields are understood as right and those with respect to sources as left. Left derivatives with respect to fields are labeled by the subscript "l," for example, $\delta_l/\delta\phi$ denote the left derivative with respect to field.

II. EXTENDED BRST QUANTIZATION IN LAGRANGIAN FORMALISM

Let us consider the theory of fields A^i, $i=1,2,..,n$, $\varepsilon(A^i)=\varepsilon_i$ for which the initial classical action $\mathcal{S}(A)$ is invariant under the gauge transformations $\delta A^i = R^i_\alpha(A)\xi^\alpha$:

$$\mathcal{S}_{,i}(A)R^i_\alpha(A)=0,$$
$$\alpha=1,2,\dots,m,\quad 0\leqslant m\leqslant n,\quad \varepsilon(\xi^\alpha)=\varepsilon_\alpha, \tag{1}$$

where the ξ^α are arbitrary functions, and the $R^i_\alpha(A)$ are generators of gauge transformations. We propose that the set $R^i_\alpha(A)$ be linearly independent and complete. One can say[19] that as a consequence of the condition of completeness the algebra of generators has the following general form:

$$R^i_{\alpha,j}(A)R^j_\beta(A)-(-1)^{\varepsilon_\alpha\varepsilon_\beta}R^i_{\beta,j}(A)R^j_\alpha(A)$$
$$=-R^i_\gamma(A)F^\gamma_{\alpha\beta}(A)-\mathcal{S}_{,j}(A)M^{ij}_{\alpha\beta}(A), \tag{2}$$

where $M^{ij}_{\alpha\beta}$ satisfies the conditions

$$M^{ij}_{\alpha\beta}=-(-1)^{\varepsilon_i\varepsilon_j}M^{ji}_{\alpha\beta}=-(-1)^{\varepsilon_\alpha\varepsilon_\beta}M^{ij}_{\beta\alpha}.$$

In the literature, the gauge theories whose generators satisfy Eqs. (2) are called general gauge theories. As has already been mentioned, covariant quantization of such theories in the framework of a standard BRST symmetry has been proposed in Refs. 1 and 2.

Our main aim here is the construction of a consistent formulation of Lagrangian quantization of the general gauge theories (1) and (2) based on the extended BRST symmetry principle. To do this, we shall first define the total configuration space ϕ^A of the theory in question:

$$\phi^A=(A^i,B^\alpha,C^{\alpha a}),\quad a=1,2,\quad \varepsilon(\phi^A)=\varepsilon_A. \tag{3}$$

We introduce additional fields B^α,

$$\varepsilon(B^\alpha)=\varepsilon_\alpha,\quad \mathrm{gh}(B^\alpha)=0,$$

and the Sp(2) doublet of the ghost fields $C^{\alpha a}$,

$$\varepsilon(C^{\alpha a})=\varepsilon_\alpha+1,\quad \mathrm{gh}(C^{\alpha a})=-(-1)^a.$$

We also introduce the sets of "antifields" ϕ^*_{Aa} and $\bar\phi_A$:

$$\phi^*_{Aa}=(A^*_{ia},B^*_{\alpha a},C^*_{\alpha ab}),\quad \bar\phi_A=(\bar A_i,\bar B_\alpha,\bar C_{\alpha a}),$$
$$\varepsilon(\phi^*_{Aa})=\varepsilon_A+1,\quad \mathrm{gh}(\phi^*_{Aa})=(-1)^a-\mathrm{gh}(\phi^A), \tag{4}$$
$$\varepsilon(\bar\phi_A)=\varepsilon_A,\quad \mathrm{gh}(\bar\phi_A)=-\mathrm{gh}(\phi^A).$$

 0022-2488/90/061487-07$03.00

The antifields ϕ^*_{Aa} and $\bar{\phi}_A$ can be treated as sources of BRST, anti-BRST, and mixed transformations.

The basic object of the extended BRST quantization in the Lagrangian formalism is the boson functional $S=S(\phi,\phi^*_a,\bar{\phi})$. We require that S be a solution to the generating equation

$$\tfrac{1}{2}(S,S)^a + V^a S = i\hbar\Delta^a S, \tag{5}$$

with the boundary condition

$$S\,|_{\phi^*_a=\bar{\phi}=0} = \mathscr{S}(A). \tag{6}$$

In writing (5) we have introduced an "extended antibracket"

$$(F,G)^a \equiv \frac{\delta F}{\delta\phi^A}\frac{\delta G}{\delta\phi^*_{Aa}} - \frac{\delta G}{\delta\phi^A}\frac{\delta F}{\delta\phi^*_{Aa}}(-1)^{(\varepsilon(F)+1)(\varepsilon(G)+1)} \tag{7}$$

whose properties are analyzed in Appendix A. In (5), $\hbar$ is the Planck constant, and the notation

$$V^a \equiv \varepsilon^{ab}\phi^*_{Ab}\frac{\delta}{\delta\bar{\phi}_A}, \quad \varepsilon^{ab} = -\varepsilon^{ba},$$
$$\Delta^a \equiv (-1)^{\varepsilon_A}\frac{\delta_l}{\delta\phi^A}\frac{\delta}{\delta\phi^*_{Aa}} \tag{8}$$

is used. Equation (5) arises in the Yang–Mills theory if to the Yang–Mills action we add terms with the sources ϕ^*_{A1} to the BRST transformation, those with the sources ϕ^*_{A2} to the anti-BRST transformation, and those with the sources $\bar{\phi}_A$ to the commutator of BRST and the anti-BRST transformations (the right-hand side in the equation for S will be equal to zero). An equation for S, of the form (5), also follows from the Hamiltonian formulation[17]; as in the case of a nonextended, standard BRST symmetry, the Hamiltonian formalism implies the master equation for the effective action.[20] It can be readily estabished that the algebra of operators (8) has the form

$$V^{\{a}V^{b\}}=0, \quad \Delta^{\{a}V^{b\}}=0, \quad \Delta^{\{a}V^{b\}}+V^{\{a}\Delta^{b\}}=0, \tag{9}$$

where, for the quantities carrying the index of the group Sp(2), we have used the notation

$$F^{\{a}G^{b\}} \equiv F^aG^b + F^bG^a.$$

It should be immediately noted that Eqs. (4) are compatible. The simplest way to establish this fact is to rewrite Eqs. (5) in an equivalent form of linear differential equations:

$$\bar{\Delta}^a\exp\{(i/\hbar)S\}=0, \quad \bar{\Delta}^a=\Delta^a+(i/\hbar)V^a. \tag{10}$$

From (9) it follows that the operators $\bar{\Delta}^a$ in (10) possess the properties

$$\bar{\Delta}^{\{a}\bar{\Delta}^{b\}}=0. \tag{11}$$

By virtue of Eqs. (11), we immediately establish compatibility of Eqs. (10) and therefore (5).

The action S is still degenerate. To lift the degeneracy, we should introduce a gauge. We denote the action modified using a gauge by $S_{\rm ext}=S_{\rm ext}(\phi,\phi^*_a,\bar{\phi})$. The gauge should be introduced so as, first, to lift the degeneracy in ϕ and, second, to retain Eq. (5), which provides invariant properties of the theory, for $S_{\rm ext}$. To meet these conditions, the gauge is introduced as

$$\exp\{(i/\hbar)S_{\rm ext}\} = \hat{U}\exp\{(i/\hbar)S\}, \tag{12}$$

where

$$\hat{U} = \exp\left\{\frac{\delta F}{\delta\phi^A}\frac{\delta}{\delta\bar{\phi}_A} + \frac{i\hbar}{2}\varepsilon_{ab}\frac{\delta}{\delta\phi^*_{Aa}}\frac{\delta^2 F}{\delta\phi^A\,\delta\phi^B}\frac{\delta}{\delta\phi^*_{Bb}}\right\}. \tag{13}$$

A direct verification shows that the operator $\hat{U}$ commutes with the operators $\bar{\Delta}^a$ in Eq. (10):

$$\hat{U}\bar{\Delta}^a = \bar{\Delta}^a\hat{U}. \tag{14}$$

Consequently,

$$\bar{\Delta}^a\exp\{(i/\hbar)S_{\rm ext}\}=0 \tag{15}$$

and $S_{\rm ext}$ satisfies Eq. (5).

We define the generating functional of the Green's functions $Z(\mathscr{T})$ as

$$Z(\mathscr{T}) = \int d\phi\,\exp\left\{\frac{i}{\hbar}\left[S_{\rm eff}(\phi) + \mathscr{T}_A\phi^A\right]\right\}, \tag{16}$$

where

$$S_{\rm eff}(\phi) = S_{\rm ext}(\phi,\phi^*_a,\bar{\phi})\,|_{\phi^*_a=\bar{\phi}=0}. \tag{17}$$

It can be represented in the form

$$Z(\mathscr{T}) = \int d\phi\,d\phi^*_a\,d\bar{\phi}\,d\pi^a\,d\lambda\,\exp\left\{\frac{i}{\hbar}\left[S(\phi,\phi^*_a,\bar{\phi}) + \phi^*_{Aa}\pi^{Aa} + \left(\bar{\phi}_A - \frac{\delta F}{\delta\phi^A}\right)\lambda^A - \frac{1}{2}\varepsilon_{ab}\pi^{Aa}\frac{\delta^2 F}{\delta\phi^B\,\delta\phi^A}\pi^{Bb} + \mathscr{T}_A\phi^A\right]\right\}. \tag{18}$$

Here the $\mathscr{T}_A$ are the usual sources to the fields ϕ^A $[\varepsilon(\mathscr{T}_A)=\varepsilon_A,\ {\rm gh}(\mathscr{T}_A) = -{\rm gh}(\phi^A)]$, and $F=F(\phi)$ is the boson gauge functional. We have introduced the sets of auxiliary fields π^{Aa} and λ^A,

$$\varepsilon(\pi^{Aa}) = \varepsilon_A + 1, \quad {\rm gh}(\pi^{Aa}) = -{\rm gh}(\phi^*_{Aa}),$$
$$\varepsilon(\lambda^A) = \varepsilon_A, \quad {\rm gh}(\lambda^A) = -{\rm gh}(\bar{\phi}_A);$$

and $S(\phi,\phi^*_a,\bar{\phi})$ satisfies Eqs. (5) and (6).

An important property of the integrand in Eq. (18) for $\mathscr{T}_A=0$ is its invariance under the following global transformations [which is, in turn, a consequence of the validity of Eq. (5) for $S_{\rm ext}$]:

$$\delta\phi^A = \pi^{Aa}\mu_a, \quad \delta\phi^*_{Aa} = \mu_a\frac{\delta S}{\delta\phi^A},$$
$$\delta\bar{\phi}_A = \varepsilon^{ab}\mu_a\phi^*_{Ab}, \quad \delta\pi^{Aa} = \delta\lambda^A = 0, \tag{19}$$

where μ_a is a Sp(2) doublet of the constant anticommuting Grassmann parameters. The transformations (19) realize the extended BRST transformation in the space of the variables ϕ, ϕ^*_a, $\bar{\phi}$, π^a, and λ.

The symmetry of the vacuum functional $Z(0)$ under the transformations (19) permits establishing the independence of the S matrix on the choice of a gauge within the proposed extended BRST quantization scheme (5) and (18). Indeed, suppose $Z_F \equiv Z(0)$. We shall change the gauge $F\to F+\Delta F$. In the functional integral for $Z_{F+\Delta F}$ we make the change of variables (13), choosing, for the parameters μ_a,

$$\mu_a = -\frac{i}{\hbar}\varepsilon_{ab}\frac{\delta(\Delta F)}{\delta\phi^A}\pi^{Ab}. \tag{20}$$

After simple algebraic transformations we find that

$$Z_{F+\Delta F} = Z_F, \tag{21}$$

and therefore the S matrix is gauge invariant.

Finally, we proceed to the derivation of the Ward identities, which are the consequence of the fact that the boson functional $S(\phi,\phi^*_a,\bar\phi)$ satisfies Eqs. (5). Consider the extended generating functional of the Green's functions:

$$Z(\mathscr{T},\phi^*_a,\bar\phi) = \int d\phi \exp\left\{\frac{i}{\hbar}[S_{\rm ext}(\phi,\phi^*_a,\bar\phi) + \mathscr{T}_A\phi^A]\right\}. \tag{22}$$

Note that from definition (22), with allowance made for Eqs. (16) and (17), it follows that

$$Z(\mathscr{T},\phi^*_a,\bar\phi)\big|_{\phi^*_a=\bar\phi=0} = Z(\mathscr{T}), \tag{23}$$

where $Z(\mathscr{T})$ is described by Eq. (18). Next, we multiply Eq. (15) from the left by $\exp\{(i/\hbar)\mathscr{T}_A\phi^A\}$ and integrate it over ϕ^A:

$$\int d\phi \exp\left\{\frac{i}{\hbar}\mathscr{T}_A\phi^A\right\}\bar\Delta^a \exp\left\{\frac{1}{\hbar}S_{\rm ext}(\phi,\phi^*_a,\bar\phi)\right\} = 0. \tag{24}$$

Integrating in Eq. (24) by parts and assuming the integrated expression vanishes, we can rewrite equality (24), with allowance made for definition (22), as

$$\left(\mathscr{T}_A\frac{\delta}{\delta\phi^*_{Aa}} - \varepsilon^{ab}\phi^*_{Ab}\frac{\delta}{\delta\bar\phi_A}\right)Z(\mathscr{T},\phi^*_a,\bar\phi) = 0, \tag{25}$$

which are the Ward identities written for the generating functional of the Green's functions. Introducing in a standard manner the generating functional of the vertex functions,

$$\Gamma(\phi,\phi^*_a,\bar\phi) = (\hbar/i)\ln Z(\mathscr{T},\phi^*_a,\bar\phi) - \mathscr{T}_A\phi^A,$$
$$\phi^A = \frac{\hbar}{i}\frac{\delta \ln Z(\mathscr{T},\phi^*_a,\bar\phi)}{\delta\mathscr{T}_A}, \tag{26}$$

we obtain

$$\tfrac{1}{2}(\Gamma,\Gamma)^a + V^a\Gamma = 0. \tag{27}$$

The identities (27) for Γ were derived earlier in Ref. 17 in the framework of the BFV Hamiltonian formulation with extended BRST symmetry. In Yang–Mills-type theories (see also Sec. IV) identities of the form (27) were obtained in Ref. 11.

III. THE EXISTENCE THEOREM FOR GENERATING EQUATIONS

The question of existence of solutions of Eqs. (5) satisfying the boundary condition (6) is essential in the construction of extended BRST quantization within the Lagrangian formalism. Here we restrict ourselves to the proof of existence of solutions to the equations

$$\tfrac{1}{2}(S,S)^a + V^aS = 0, \tag{28}$$

with the boundary condition (6). Note that, for local S, $\Delta^aS\sim\delta(0)$, and, using the dimensional regularization $[\delta(0)=0]$, Eqs. (5) become (28).

The solution of Eqs. (28) will be sought in the form of a power series of the fields B^α and $C^{\alpha a}$. In this connection it turns out to be convenient to ascribe a so-called "new ghost number," ngh, to all the quantities. For the variables ϕ^A, ϕ^*_{Aa}, and $\bar\phi_A$ the new ghost number will be defined by the rule

$$\mathrm{ngh}(A^i)=0,\quad \mathrm{ngh}(C^{\alpha a})=1,\quad \mathrm{ngh}(B^\alpha)=2,$$
$$\mathrm{ngh}(\phi^*_{Aa}) = -1-\mathrm{ngh}(\phi^A),\quad \mathrm{ngh}(\bar\phi_A) = -2-\mathrm{ngh}(\phi^A). \tag{29}$$

The solution of Eqs. (28) will be sought in the class of boson functionals $S=S(\phi,\phi^*_a,\bar\phi)$ with $\mathrm{ngh}(S)=0$ in the form of the expansion

$$S=\mathscr{S}(A) + \sum_{n=1} S_n,\quad \mathrm{ngh}(S_n)=0,\quad S_n\sim C^{n-m}B^m. \tag{30}$$

Let us consider the first approximation S_1. The most general form of the functional S_1 meeting the above-mentioned requirements is

$$S_1 = A^*_{ia}\Lambda^{ia}_{\alpha b}(A)C^{\alpha b} + C^*_{\alpha ab}\Lambda^{\alpha ab}_\beta(A)B^\beta$$
$$+\bar A_i\Lambda^i_\alpha(A)B^\alpha + \tfrac{1}{2}A^*_{ia}A^*_{jb}\Lambda^{(ia)(jb)}_\alpha(A)B^\alpha. \tag{31}$$

Here $\Lambda^{(ia)}_{\alpha b}$, $\Lambda^{\alpha ab}_\beta$, Λ^i_α, and $\Lambda^{(ia)(jb)}_\alpha$ are some unknown matrices depending on the fields A^i, where

$$\Lambda^{(ia)(jb)}_\alpha = (-1)^{(\varepsilon_i+1)(\varepsilon_j+1)}\Lambda^{(jb)(ia)}_\alpha. \tag{32}$$

Next, we require that the functional $\mathscr{S}(A)+S_1$ satisfy Eqs. (28) to first order. This leads to the following equations for $\Lambda^{ia}_{\alpha b}$, $\Lambda^{\alpha ab}_\beta$, Λ^i_α, and $\Lambda^{(ia)(jb)}_\alpha$:

$$\mathscr{S}_{,i}(A)\Lambda^{ia}_{\alpha b}(A) = 0, \tag{33}$$

$$\mathscr{S}_{,j}(A)\Lambda^{(ia)(jb)}_\alpha(-1)^{\varepsilon_i} + \Lambda^{ib}_{\beta d}\Lambda^{\beta ad}_\alpha + \varepsilon^{ab}\Lambda^i_\alpha = 0. \tag{34}$$

From Eqs. (33) it follows that $\Lambda^{ia}_{\alpha b}$ can be identified with the generators of the gauge transformations

$$\Lambda^{ia}_{\alpha b}(A) = R^i_{\ \alpha}(A)\delta^a_b. \tag{35}$$

Next, we set

$$\Lambda^{(ia)(jb)}_\alpha \equiv 0. \tag{36}$$

Then Eq. (34) implies

$$\Lambda^{\beta ab}_\alpha = \Lambda^\beta_\alpha\varepsilon^{ab}. \tag{37}$$

Considering Eq. (31), we come to the conclusion that redefining the field B^α, one can assume without loss of generality that

$$\Lambda^\beta_\alpha = -\delta^\beta_\alpha. \tag{38}$$

Turning again to Eq. (34) we obtain

$$\Lambda^i_\alpha(A) = R^i_{\ \alpha}(A). \tag{39}$$

This result for the first approximation corresponds to imposing, in addition to condition (6), the boundary conditions

$$\left.\frac{\delta^2 S}{\delta A^*_{ia}\,\delta C^{\alpha b}}\right|_{C=B=0} = R^i_{\ \alpha}(A)\delta^a_b,$$
$$\left.\frac{\delta^2 S}{\delta \bar A_i\,\delta B^\alpha}\right|_{C=B=0} = R^i_{\ \alpha}(A),$$

$$\left.\frac{\delta^2 S}{\delta C^*_{\alpha ab}\,\delta B^\beta}\right|_{C=B=0} = -\varepsilon^{ab}\delta^\alpha_\beta, \quad (40)$$

$$\left.\frac{\delta^2 S}{\delta A^*_{ia}\,\delta B^\alpha}\right|_{C=B=0} = 0.$$

Thus as the first approximation we finally choose the functional

$$S_1 = A^*_{ia} R^i_\alpha(A) C^{\alpha a} + \bar{A}_i R^i_\alpha(A) B^\alpha - \varepsilon^{ab} C^*_{\alpha ab} B^\alpha. \quad (41)$$

Suppose now that we have constructed the functional $[S]_n$, where

$$[S]_n \equiv \mathscr{S}(A) + \sum_{k=1}^{n} S_k, \quad (42)$$

which satisfies (28) to within nth order:

$$\tfrac{1}{2}([S]_n,[S]_n)^a_k + V^a S_k = 0, \quad k = 1,2,\ldots,n. \quad (43)$$

In Eq. (43) and hereafter $(\ ,\ )^a_k$ denotes the kth order in powers of the fields B^α and $C^{\alpha a}$ of the extended antibracket $(\ ,\ )^a$. For the $(n+1)$th approximation S_{n+1} of the solution of Eq. (28), we have, taking Eq. (41) into account,

$$W^a S_{n+1} = F^a_{n+1}. \quad (44)$$

The operators W^a in Eqs. (44) have the form

$$W^a = W^a_0 + V^a,$$

$$W^a_0 = \mathscr{S}_{,i}\frac{\delta}{\delta A^*_{ia}} + A^*_{ib} R^i_\alpha \frac{\delta}{\delta C^*_{\alpha ab}} + (\bar{A}_i R^i_\alpha - \varepsilon^{bc} C^*_{\alpha bc})\frac{\delta}{\delta B^*_{\alpha a}} + (-1)^{\varepsilon_\alpha}\varepsilon^{ab} B^\alpha \frac{\delta_l}{\delta C^{\alpha b}} \quad (45)$$

and possess the following important properties

$$W^{\{a}W^{b\}} = 0. \quad (46)$$

The functionals F^a_{n+1} in Eqs. (44) are constructed from S_k, $k \leqslant n$, by the rule

$$F^a_{n+1} = -\tfrac{1}{2}([S]_n,[S]_n)^a_{n+1}. \quad (47)$$

From Eqs. (46) it follows that for Eqs. (44) to be compatible it is necessary that the relations

$$W^{\{a}F^{b\}}_{n+1} = 0 \quad (48)$$

hold. We shall show that relations (48) do hold. To this end we consider the identities (A4) and rewrite them as

$$(\tfrac{1}{2}(S,S)^{\{a} + V^{\{a}S,S)^{b\}} - (V^{\{a}S,S)^{b\}} \equiv 0. \quad (49)$$

The properties (9) and (A7) of the operators V^a enable Eq. (49) to be identically written as

$$(S,\tfrac{1}{2}(S,S)^{\{a} + V^{\{a}S)^{b\}} + V^{\{a}[\tfrac{1}{2}(S,S)^{b\}} + V^{b\}}S\,] \equiv 0. \quad (50)$$

We now consider the identities (50) in the $(n+1)$th approximation. We take into account that by virtue of Eqs. (43) and the lowest approximation for the expression $\tfrac{1}{2}(S,S,)^a + V^a S$ is $(n+1)$th order, which is equal to $W^a S_{n+1} - F^a_{n+1}$. Then in the $(n+1)$th approximation the identities (50) become

$$W^{\{a}_0(W^{b\}}S_{n+1} - F^{b\}}_{n+1}) + V^{\{a}(W^{b\}}S_{n+1} - F^{b\}}_{n+1}) = 0. \quad (51)$$

By virtue of Eqs. (45) and (46), from Eq. (51) the relations (48), which provide compatibility of Eqs. (44), follow.

Further proof of the existence theorem rests on the following lemma.

Lemma: Any regular solution of the equations

$$W^a X = 0. \quad (52)$$

$$W^{\{a}X^{b\}} = 0, \quad (53)$$

vanishing for $\mathscr{S}_{,i} = \phi^*_{Aa} = \bar{\phi}_A = 0$, has the form

$$X = \tfrac{1}{2}\varepsilon_{ab} W^a W^b Y, \quad (54)$$

$$X^a = W^a Z, \quad (55)$$

with some functionals Y and Z, respectively. Moreover, if X and X^a are Sp(2) covariant, then Y and Z can be chosen as Sp(2) scalars. The proof of the lemma is given in Ref. 17. It is based on the existence of the operators Γ_a, $\varepsilon(\Gamma_a) = 1$, "conjugate" to the operators W^a and such that

$$W^a\Gamma_b + \Gamma_b W^a = \delta^a_b N, \quad \Gamma_{\{a}\Gamma_{b\}} = 0, \quad (56)$$

where N is a scalar operator under the group Sp(2). In Appendix B we prove the existence of the operators Γ_a and show that the operator N possesses the properties

$$W^a N = N W^a, \quad \Gamma_a N = N\Gamma_a, \quad (57)$$

and, with the solutions dealt with in the lemma, N is positive definite.

We now return to the solution of Eqs. (44). Since $\mathrm{ngh}(F^a_{n+1}) = 1$ and $n \geqslant 1$, it follows that $F^a_{n+1} = 0$ for $\mathscr{S}_{,i} = \phi^*_{Aa} = \bar{\phi}_A = 0$, and therefore, by virtue of the lemma, the solution of (48) is representable in the form

$$F^a_{n+1} = W^a X_{n+1}. \quad (58)$$

Choosing $S_{n+1} = X_{n+1}$, we find that Eqs. (28) are already satisfied to within $(n+1)$th-order terms. Then by induction we conclude the proof of the existence of solutions of Eqs. (28). Note that for the S_{n+1} we could take the functional

$$S_{n+1} = X_{n+1} + \tfrac{1}{2}\varepsilon_{ab} W^a W^b Y_{n+1}, \quad (59)$$

and, as before, Eqs. (28) would be satisfied to within terms of order $(n+1)$. On the basis of the lemma it is not difficult to show that, boundary conditions (6) and (40) being given, the arbitrariness (59) in the choice of the $(n+1)$th approximation is unique.

IV. GAUGE THEORIES WITH A CLOSED ALGEBRA

To illustrate the formalism of the extended BRST quantization developed here, we consider irreducible gauge theories of rank 1 with a closed algebra. Such theories are characterized by the fact that in the algebra of generators, Eqs. (2), $M^{ij}_{\alpha\beta} = 0$, and the solution of any equation of the form $R^i_\alpha X^\alpha = 0$ is $X^\alpha = 0$. The majority of the theories discussed in the literature belong to the indicated class (Yang–Mills, gravity, supergravity theories with auxiliary fields, etc.). From the viewpoint of extended BRST quantization, typical of all these theories is that the solution of Eqs. (28) exists as a linear functional in the antifields ϕ^*_{Aa} and $\bar{\phi}_A$:

$$S = \mathscr{S}(A) + \phi^*_{Aa} X^{Aa} + \bar{\phi}_A Y^A. \tag{60}$$

Here X^{Aa} and Y^A are functionals of the fields ϕ^A:

$$\varepsilon(X^{Aa}) = \varepsilon(\phi^*_{Aa}), \quad \mathrm{ngh}(X^{Aa}) = -\mathrm{ngh}(\phi^*_{Aa}),$$
$$\varepsilon(Y^A) = \varepsilon(\bar{\phi}_A), \quad \mathrm{ngh}(Y^A) = -\mathrm{ngh}(\bar{\phi}_A). \tag{61}$$

Substituting Eqs. (66) into Eqs. (28), we obtain the system of equations for finding X^{Aa} and Y^A transformable to the form

$$\frac{\delta \mathscr{S}(A)}{\delta \phi^A} X^{Aa} = 0, \tag{62}$$

$$\frac{\delta X^{Aa}}{\delta \phi^B} X^{Bb} + \frac{\delta X^{Ab}}{\delta \phi^B} X^{Ba} = 0, \tag{63}$$

$$Y^A = \frac{1}{2} \varepsilon_{ab} \frac{\delta X^{Aa}}{\delta \phi^B} X^{Bb}, \tag{64}$$

$$\frac{\delta Y^A}{\delta \phi^B} X^{Ba} = 0. \tag{65}$$

Let

$$X^{Aa} = (X^{ia}_1, X^{\alpha a}_2, X^{\alpha ab}_3), \quad Y^A = (Y^i_1, Y^\alpha_2, Y^{\alpha a}_3). \tag{66}$$

Equations (62) are of the form

$$\mathscr{S}_{,i} X^{ia}_1 = 0 \tag{67}$$

and X^{ia}_1 may, therefore, be thought of as the generator of gauge transformations [see Eqs. (1)]. Taking into account that $\mathrm{ngh}(X^{ia}_1) = 1$ and $\mathrm{ngh}(C^{\alpha a}) = 1$, we choose X^{ia}_1 in the form

$$X^{ia}_1 = R^i_\alpha C^{\alpha a}. \tag{68}$$

With allowance made for Eqs. (68), we find, from the solution of Eqs. (63), that

$$X^{\alpha a}_2 = -\tfrac{1}{2} F^\alpha_{\gamma\beta} B^\beta C^{\gamma a} - \tfrac{1}{12}(-1)^{\varepsilon_\beta}(2F^\alpha_{\gamma\beta,j} R^j_\rho$$
$$+ F^\alpha_{\gamma\sigma} F^\sigma_{\beta\rho}) C^{\rho b} C^{\beta a} C^{\gamma c} \varepsilon_{cb}, \tag{69}$$

$$X^{\alpha ab}_\beta = -\varepsilon^{ab} B^\alpha - \tfrac{1}{2}(-1)^{\varepsilon_\beta} F^\alpha_{\beta\gamma} C^{\gamma b} C^{\beta a}. \tag{70}$$

Substituting Eqs. (68)–(70) into Eqs. (64), we are led to the following expression for the functionals:

$$Y^i_1 = R^i_\alpha B^\alpha + \tfrac{1}{2}(-1)^{\varepsilon_\alpha} R^i_{\alpha,j} R^j_\beta C^{\beta b} C^{\alpha a} \varepsilon_{ab}, \tag{71}$$

$$Y^\alpha_2 = 0, \quad Y^{\alpha a}_3 = -2X^{\alpha a}_3. \tag{72}$$

A direct verification shows us that Eqs. (65) with the functionals (68)–(72) hold identically. In the solution of the system of equations (62)–(65) we have intensely used the Jacobi identity

$$(-1)^{\varepsilon_\beta \varepsilon_\rho}(F^\alpha_{\beta\gamma,i} R^i_\rho + F^\alpha_{\beta\sigma} F^\sigma_{\gamma\rho}) + \mathrm{cycle}(\beta,\gamma,\rho) \equiv 0. \tag{73}$$

The relations (68)–(72) specify the transformations of extended BRST symmetry for theories with closed algebra. In a particular case, where all the gauge parameters ξ^α are boson functions, i.e., $\varepsilon_\alpha = 0$, the extended BRST symmetry transformations are also obtained in Ref. 16.

We shall show that in the class of gauges $F(\phi)$ depending only on the initial fields A^i,

$$F(\phi) = F(A), \tag{74}$$

the generating functional (18) is reduced to the standard Faddeev–Popov result.[21] Indeed, by virtue of Eqs. (60) and (74), the integration over the variables $\bar{\phi}_A$, ϕ^*_{Aa}, λ^A, and π^{Aa} is trivial and yields

$$Z(\mathscr{T}) = \int d\phi \exp\left\{\frac{i}{\hbar}\left[\mathscr{S}(A) - \frac{1}{2}\varepsilon_{ab} X^{ia}_1 \frac{\delta^2 F}{\delta A^j \delta A^i} X^{jb}_1 + \frac{\delta F}{\delta A^i} Y^i_1 + \mathscr{T}_A \phi^A\right]\right\}. \tag{75}$$

Taking into account Eqs. (68) and (71), we come to

$$\frac{\delta F}{\delta A^i} Y^i_1 - \frac{1}{2}\varepsilon_{ab} X^{ia}_1 \frac{\delta^2 F}{\delta A^j \delta A^i} X^{jb}_1$$
$$= \frac{\delta F}{\delta A^i} R^i_\alpha B^\alpha + \frac{1}{2}(-1)^{\varepsilon_\alpha}\left(\frac{\delta F}{\delta A^i} R^i_{\alpha,j} R^j_\beta + \frac{\delta^2 F}{\delta A^j \delta A^i} R^j_\alpha R^i_\beta (-1)^{\varepsilon_i(\varepsilon_j + \varepsilon_\alpha)}\right) C^{\beta b} C^{\alpha a} \varepsilon_{ab}. \tag{76}$$

If we introduce the function

$$\chi_\alpha = \frac{\delta F}{\delta A^i} R^i_\alpha \tag{77}$$

and identify $C^{\alpha 1} \equiv C^\alpha$ and $C^{\alpha 2} \equiv \bar{C}^\alpha$, then, taking account of Eq. (76), the functional integral (75) can finally be written as

$$Z(\mathscr{T}) = \int d\phi \exp\left\{\frac{i}{\hbar}[\mathscr{S}(A) + \bar{C}^\alpha \chi_{\alpha,i} R^i_\beta C^\beta + \chi_\alpha \beta^\alpha + \mathscr{T}_A \phi^A]\right\}. \tag{78}$$

This is the standard Faddeev–Popov result for gauge theories with a closed algebra when the gauge is introduced by means of the function χ_α

It is noteworthy that if in any theory we go over to quadratic approximation, the algebra of the gauge transformations becomes Abelian and the action S_{eff} acquires the form of the action of a theory with closed algebra. Then the consideration presented in this section shows that the method of gauge fixing proposed in Sec. II will actually lift the degeneracy of the classical gauge-invariant action.

Concluding, we note that we have not considered the important questions of the description of the general solution of Eq. (5) and the establishment of equivalence between the formalism with extended BRST symmetry and the standard Lagrangian BRST formalism. These questions are apparently closely connected with each other.

APPENDIX A: THE PROPERTIES OF THE EXTENDED ANTIBRACKET

From definition (7),

$$\varepsilon((F,G)^a) = \varepsilon(F) + \varepsilon(G) + 1,$$

$$\mathrm{gh}((F,G)^a) = -(-1)^a + \mathrm{gh}(F) + \mathrm{gh}(G), \quad a = 1,2,$$

and

$$(F,G)^a = -(G,F)^a(-1)^{(\varepsilon(F)+1)(\varepsilon(G)+1)}, \tag{A1}$$

$$(F,GH)^a = (F,G)^a H + (F,H)^a G(-1)^{\varepsilon(G)\varepsilon(H)}, \tag{A2}$$

In particular, for any fermion functional F, $\varepsilon(F) = 1$. Eqs. (A1) imply $(F,F)^a \equiv 0$. By means of simple but cumbersome calculations, one can establish the Jacobi identity for the extended antibracket:

$$((F,G)^{\{a},H)^{b\}}(-1)^{(\varepsilon(F)+1)(\varepsilon(H)+1)} + \text{cycle}(F,G,H) \equiv 0. \tag{A3}$$

For any boson functional S, $\varepsilon(S)=0$, from Eqs. (A3) it follows that

$$((S,S)^{\{a},S)^{b\}} \equiv 0. \tag{A4}$$

The action of the operators V^a, Eqs. (8), upon the extended antibracket is given by the relations

$$V^a(F,G)^b = (V^aF,G)^b - (-1)^{\varepsilon(F)}(F,V^aG)^b - \varepsilon^{ab}\left(\frac{\delta F}{\delta\phi^A}\frac{\delta G}{\delta\bar\phi_A} - \frac{\delta G}{\delta\phi^A}\frac{\delta F}{\delta\bar\phi_A} \times (-1)^{\varepsilon(F)(\varepsilon(G)+1)}\right). \tag{A5}$$

Therefore

$$V^{\{a}(F,G)^{b\}} = (V^{\{a}F,G)^{b\}} - (-1)^{\varepsilon(F)}(F,V^{\{a}G)^{b\}}. \tag{A6}$$

For any boson functional S, Eq. (A6) implies

$$\tfrac{1}{2}V^{\{a}(S,S,)^{b\}} = (V^{\{a}S,S)^{b\}}. \tag{A7}$$

APPENDIX B: CONSTRUCTION OF OPERATORS Γ_a

In this appendix we prove the existence of the operators Γ_a "conjugate" to W^a in the sense of Eqs. (56) and establish the validity of relations (57). The crucial point of our consideration is the possibility of reducing the operators W^a to the "standard" form, i.e., to that of the operators $G_i\,\delta/\delta P_i$, where both the set of G_i and the set of P_i are functionally independent.

Reduction of the operators W^a to the standard form is realized in several steps. First, from the initial variables A^i we go over, using a nonsingular change, to the variables A'^i:

$$A^i = A^i(A') \leftrightarrow A'^i = A'^i(A) = (\varphi^m,\xi^\alpha). \tag{B1}$$

Here the initial classical action does not depend on the gauge fields ξ^α explicitly:

$$\mathscr{S}(A) = \mathscr{S}(A(A')) = \mathscr{S}'(A') = \mathscr{S}'(\varphi). \tag{B2}$$

Given this, the gauge invariance condition (1) becomes

$$\mathscr{S}_{,i}(A)R^i_{\ \alpha}(A) = \mathscr{S}'_{,i}N^i_jR^j_{\ \alpha} = \mathscr{S}'_{,i}R'^i_{\ \alpha}(A') = 0, \tag{B3}$$

where

$$R'^i_{\ \alpha}(A') = N^i_j(A(A'))R^j_{\ \alpha}(A(A')), \quad N^i_j(A) = \frac{\delta A'^i(A)}{\delta A^j}. \tag{B4}$$

With allowance made for Eq. (B2), the identity (B3) can now be rewritten as

$$\mathscr{S}'_{,i}R'^i_{\ \alpha}(A') = \mathscr{S}'_{,m}R'^m_{\ \alpha}(A') = 0. \tag{B5}$$

From Eq. (B5) we conclude that $R'^m_{\ \alpha}(A')$ can be only trivial generators for the action $\mathscr{S}'(\varphi)$:

$$R'^n_{\ \alpha}(A') = \mathscr{S}'_{,m}\Lambda^{nm}_\alpha(A'), \quad \Lambda^{nm}_\alpha = -(-1)^{\varepsilon_m\varepsilon_n}\Lambda^{mn}_\alpha. \tag{B6}$$

The generators $R'^i_{\ \alpha}(A')$ are representable in the form

$$R'^i_{\ \alpha} = (\mathscr{S}'_{,n}\Lambda^{mn}_\alpha,\tilde R^\beta_{\ \alpha}), \tag{B7}$$

where $\tilde R^\beta_{\ \alpha}$ is the nondegenerate matrix.

In addition to the changes (B1), we also make the following antifield transformations:

$$\begin{aligned} &A^{*\prime}_{ia} = A^*_{ja}M^j_i, \quad \bar A'_i = \bar A_jM^j_i,\\ &C^{*\prime}_{\alpha ab} = C^*_{\beta ab}(\tilde R^{-1})^\beta_\alpha, \quad B^{*\prime}_{\alpha a} = B^*_{\beta a}(\tilde R^{-1})^\beta_\alpha,\\ &\bar B'_\alpha = \bar B_\beta(\tilde R^{-1})^\beta_\alpha, \quad \bar C'_{\alpha a} = \bar C_{\beta a}(\tilde R^{-1})^\beta_\alpha, \end{aligned} \tag{B8}$$

where we have introduced the notation

$$M^i_j \equiv \frac{\delta A^i(A')}{\delta A'^j}, \quad M^i_jN^j_k = \delta^i_k. \tag{B9}$$

As a result of the changes (B1) and (B8), the operators $W^a \to W'^a$,

$$\begin{aligned} W'^a = {}& \mathscr{T}_m\frac{\delta}{\delta A^{*\prime}_{ab}} + (A^{*\prime}_{\alpha b} + A^{*\prime}_{mb}\mathscr{T}_n\Lambda^{mn}_\beta(\tilde R^{-1})^\beta_\alpha)\frac{\delta}{\delta C^{*\prime}_{\alpha ab}}\\ &+ (\bar A'_\alpha + \bar A'_m\mathscr{T}_n\Lambda^{mn}_\beta(\tilde R^{-1})^\beta_\alpha - \varepsilon^{bc}C^{*\prime}_{\alpha bc})\frac{\delta}{\delta B^{*\prime}_{\alpha a}}\\ &+ \varepsilon^{ab}\left(A^{*\prime}_{mb}\frac{\delta}{\delta\bar A'_m} + A^{*\prime}_{\alpha b}\frac{\delta}{\delta\bar A'_\alpha} + B^{*\prime}_{\alpha b}\frac{\delta}{\delta\bar B'_\alpha}\right.\\ &\left. + C^{*\prime}_{\alpha bc}\frac{\delta}{\delta\bar C'_{\alpha a}} + (-1)^{\varepsilon_\alpha}B^\alpha\frac{\delta_b}{\delta C^{\alpha b}}\right),\\ \mathscr{T}_m \equiv {}& \mathscr{S}'_{,m}. \end{aligned} \tag{B10}$$

In the operators W'^a [(B10)], we make the change

$$\begin{aligned} &A^{*\prime\prime}_{\alpha a} = A^{*\prime}_{\alpha a} + A^{*\prime}_{ma}\mathscr{T}_n\Lambda^{mn}_\beta(\tilde R^{-1})^\beta_\alpha, \quad A^{*\prime\prime}_{ma} = A^{*\prime}_{ma},\\ &\bar A''_\alpha = \bar A'_\alpha + \bar A'_m\mathscr{T}_n\Lambda^{mn}_\beta(\tilde R^{-1})^\beta_\alpha - \varepsilon^{bc}C^{*\prime}_{\alpha bc},\\ &\bar A''_m = \bar A'_m, \quad C^{*\prime\prime}_{\alpha bc} = C^{*\prime}_{\alpha bc}, \end{aligned} \tag{B11}$$

and $W'^a \to W''^a$, where

$$\begin{aligned} W''^a = {}& \mathscr{T}_m\frac{\delta}{\delta A^{*\prime\prime}_{ma}} + \varepsilon^{ab}A^{*\prime\prime}_{mb}\frac{\delta}{\delta\bar A''_m} + A^{*\prime\prime}_{\alpha b}\frac{\delta}{\delta C^{*\prime\prime}_{\alpha ab}}\\ &+ \varepsilon^{ab}C^{*\prime\prime}_{\alpha bc}\frac{\delta}{\delta\bar C'_{\alpha c}} + \bar A''_\alpha\frac{\delta}{\delta B^{*\prime}_{\alpha a}}\\ &+ \varepsilon^{ab}B^{*\prime}_{\alpha b}\frac{\delta}{\delta\bar B'_\alpha} + \varepsilon^{ab}(-1)^{\varepsilon_\alpha}B^\alpha\frac{\delta_l}{\delta C^{\alpha b}}. \end{aligned} \tag{B12}$$

The operators W''^a are already of the "standard" form. We shall now construct operators Γ''_a such that

$$W''^a\Gamma''_b + \Gamma''_bW''^a = \delta^a_bN'', \quad \Gamma''_{\{a}\Gamma''_{b\}} = 0. \tag{B13}$$

The solution of Eqs. (B13) does exist. For example, for the operators Γ''_a one can choose

$$\begin{aligned} \Gamma''_a = {}& A^{*\prime\prime}_{ma}\frac{\delta}{\delta\mathscr{T}_m} - \varepsilon_{ab}\bar A''_m\frac{\delta}{\delta A^{*\prime\prime}_{mb}} + C^{*\prime\prime}_{\alpha ab}\frac{\delta}{\delta A^{*\prime\prime}_{\alpha b}}\\ &- \varepsilon_{ab}\bar C'_{\alpha c}\frac{\delta}{\delta C^{*\prime\prime}_{\alpha bc}} + B^{*\prime}_{\alpha a}\frac{\delta}{\delta\bar A''_\alpha} - \varepsilon_{ab}\bar B'_\alpha\frac{\delta}{\delta B^{*\prime}_{\alpha b}}. \end{aligned} \tag{B14}$$

Then for operators N'' in Eqs. (B13) we deduce

$$N'' = \mathscr{T}_m \frac{\delta}{\delta \mathscr{T}_m} + A^{*''}_{ma} \frac{\delta}{\delta A^{*''}_{ma}} + \bar{A}''_m \frac{\delta}{\delta A^{*''}_m}$$
$$+ C^{*''}_{aab} \frac{\delta}{\delta C^{*''}_{aab}} + A^{*''}_{aa} \frac{\delta}{\delta A^{*''}_{aa}}$$
$$+ \bar{C}'_{aa} \frac{\delta}{\delta \bar{C}'_{aa}} + B^{*'}_{aa} \frac{\delta}{\delta B^{*'}_{aa}}$$
$$+ \bar{A}''_\alpha \frac{\delta}{\delta \bar{A}''_\alpha} + \bar{B}'_\alpha \frac{\delta}{\delta \bar{B}'_\alpha}. \quad \text{(B15)}$$

Through a direct verification we make sure that the equalities

$$W''^a N'' = N'' W''^a, \quad \Gamma''_a N'' = N'' \Gamma''_a \quad \text{(B16)}$$

hold. In Eqs. (B13)–(B16) we now make transformations inverse to (B1), (B8), and (B11) to obtain

$$W^a \Gamma_b + \Gamma_b W^a = \delta^a_b N, \quad \Gamma_{\{a} \Gamma_{b\}} = 0,$$
$$W^a N = N W^a, \quad \Gamma_a N = N \Gamma_a, \quad \text{(B17)}$$

where the operators W^a are given by the expression (45), and the operators Γ_a and N have the form

$$\Gamma_a = A^*_{ia} Q^i_j \frac{\delta}{\delta \mathscr{S}_{,j}} - \varepsilon_{ab} \bar{A}_i Q^i_l P^l_j \frac{\delta}{\delta A^*_{jb}} + C^*_{aab} L^\alpha_i \frac{\delta}{\delta A^*_{ib}}$$
$$- \varepsilon_{ab} \bar{C}_{\alpha c} \frac{\delta}{\delta C^*_{abc}} - \bar{C}_{\alpha a} L^\alpha_i \frac{\delta}{\delta \bar{A}_i}$$
$$+ B^*_{\alpha a} L^\alpha_i \frac{\delta}{\delta \bar{A}_i} - \varepsilon_{ab} \bar{B}_\alpha \frac{\delta}{\delta B^*_{\alpha b}}, \quad \text{(B18)}$$

$$N = \mathscr{S}_{,i} Q^i_j \frac{\delta}{\delta \mathscr{S}_j} + \phi^*_{Aa} \frac{\delta}{\delta \phi^*_{Aa}} + \bar{\phi}_A \frac{\delta}{\delta \bar{\phi}_A}. \quad \text{(B19)}$$

In Eqs. (B18) and (B19) we have used the notation

$$Q^i_j \equiv M^i_m N^m_j, \quad L^\alpha_i = (\tilde{R}^{-1})^\alpha_\beta N^\beta_i,$$
$$P^i_j = \delta^i_j - R^i_\alpha L^\alpha_j. \quad \text{(B20)}$$

A direct verification shows that the matrices Q^i_j, L^α_i, and P^i_j possess the properties

$$P^i_l P^l_j = P^i_j, \quad Q^i_l Q^l_j = Q^i_j,$$
$$P^i_l Q^l_j = Q^i_j, \quad Q^i_l P^l_j R^j_\alpha = 0, \quad \text{(B21)}$$
$$L^\alpha_j Q^j_i = 0, \quad L^\alpha_i P^i_j = 0.$$

In the derivation of (B21) we have used the relation

$$\tilde{R}^\alpha_\beta = N^\alpha_j R^j_\beta. \quad \text{(B22)}$$

[1]B. de Wit and J. W. van Holten, Phys. Lett. B **79**, 389 (1978).
[2]I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **102**, 27 (1981).
[3]C. Becchi, A. Rouet and R. Stora, Commun. Math. Phys. **42**, 127 (1975).
[4]I. V. Tyutin, Preprint Lebedev Inst. No. 39, 1975.
[5]G. Curci and R. Ferrari, Phys. Lett. B **63**, 91 (1976).
[6]I. Ojima, Prog. Theor. Phys. Suppl. **64**, 625 (1979).
[7]N. Nakanishi and I. Ojima, Z. Phys. C **6**, 155 (1980).
[8]F. R. Ore and P. van Nieuwenhuizen, Nucl. Phys. B **204**, 317 (1982).
[9]L. Bonora and M. Tonin, Phys. Lett. B **98**, 48 (1981).
[10]L. Bonora, P. Pasti, and M. Tonin, J. Math. Phys. **23**, 839 (1982).
[11]L. Alvarez-Gaume and L. Bauleiu, Nucl. Phys. B **212**, 255 (1983).
[12]L. Baulieu and J. Thierry-Mieg, Nucl. Phys. B **197**, 477 (1982).
[13]N. K. Falck and A. C. Hirshfeld, Phys. Lett. B **138**, 52 (1984).
[14]F. A. Kaempffer, J. Math. Phys. **26**, 1077 (1985).
[15]W. Siegel and B. Zwiebach, Nucl. Phys. B **288**, 332 (1987).
[16]V. P. Spiridonov, Nucl. Phys. B **308**, 527 (1988).
[17]I. A. Batalin, P. M. Lavrov, and I. V. Tyutin, J. Math. Phys. **31**, 6 (1990).
[18]B. S. De Witt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965).
[19]I. A. Batalin and G. A. Vilkovisky, J. Math. Phys. **26**, 172 (1985).
[20]I. A. Batalin and E. S. Fradkin, Ann. Inst. H. Poincaré (Phys. Theor.) **49**, 145 (1988).
[21]L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 29 (1967).

Explicit construction of nontrivial elements for homotopy groups of classical Lie groups

Albert T. Lundell
Department of Mathematics, University of Colorado, Campus Box 426, Boulder, Colorado 80309
Yasunari Tosa
Department of Physics, University of Colorado, Campus Box 390, Boulder, Colorado 80309

(Received 3 August 1989; accepted for publication 31 January 1990)

Nontrivial elements of homotopy groups for unitary, orthogonal, and symplectic groups are given *explicitly*. In particular, (a) representatives of generators of nontrivial homotopy groups of stable special unitary, orthogonal, and symplectic groups are constructed using Clifford algebras; (b) the values for "winding numbers" for stable SU, SO, and Sp are calculated for generators of homotopy groups; and (c) representatives of generators of homotopy groups $\Pi_{n-2}(O(n-1))$, $\Pi_{2n-2}(U(n-1))$, $\Pi_{4n-2}(\mathrm{Sp}(n-1))$ are given.

I. INTRODUCTION

Homotopy groups of compact connected simple Lie groups and their coset spaces are beginning to be used in various aspects of high-energy physics.[1,2] If D denotes the dimension of space (or space-time) and M the target manifold of a configuration, the following are examples associated with $\Pi_{D-1}(M)$.

(i) A vortex solution in a $D=2$ static, Euclidean, Abelian, Yang–Mills–Higgs system associated with $\Pi_1(U(1))$.

(ii) A magnetic monopole solution in a $D=3$ static, Euclidean, Yang–Mills–Higgs system associated with $\Pi_2(S^2)$.

(iii) An instanton solution in a $D=4$ Euclidean, Yang–Mills system associated with Π_3 (SU(2)). [Note that a Yang–Mills or Yang–Mills–Higgs system cannot have a finite energy or an action solution for $D>4$ (Ref. 2).]

(iv) A soliton in current algebra associated with $\Pi_{D-1}(G)$. An example associated with $\Pi_D(M)$ is:

(v) a global gauge anomaly[3] that is caused by a nontrivial element of $\Pi_D(H)$ where H is the Yang–Mills gauge group.[3–11] An example associated with $\Pi_{D+1}(M)$ is:

(vi) the Wess–Zumino effective Lagrangian of current algebra associated with $\Pi_5(\mathrm{SU}(3))$ (Ref. 3).

The techniques of algebraic topology permit most calculations to be done without knowing explicit functional forms representing nontrivial elements of homotopy groups. Such functional forms are well known only for the cases of a vortex, a magnetic monopole, and an instanton.[1,2] [These two cases are related to Hopf fiberings, $S^1=U(1)\to S^3\to S^2$ and $S^3=\mathrm{SU}(2)\to S^7\to S^4$ (Ref. 1).] Some functional forms in higher dimensions are scattered in the mathematical literature. In this paper, we treat them in a uniform way without the need of detailed knowledge of algebraic topology and provide *explicit* formulas representing nontrivial elements of homotopy groups. By providing these explicit forms, physicists may gain more insight into global chiral Yang–Mills gauge anomalies and solitons in higher dimensions.

In Sec. II, generators of stable homotopy groups of classical simple Lie groups are constructed in terms of Clifford algebras. We first give a generator of $\Pi_{2n+1}(\mathrm{SU})$, where SU is the stable special unitary group [i.e., $\Pi_k(\mathrm{SU}) = \Pi_k(\mathrm{SU}(m))$ for $m\geqslant(k+1)/2$]. This is important because it is used to fix the normalization of the anomaly of example 5. The normalization of the anomaly is always referred to Bott and Seeley[12] who give no explicit functional form. Next, we give formulas for representation of generators of $\Pi_{4n-1}(\mathrm{SO})$ and $\Pi_{4n-1}(\mathrm{Sp})$, where SO and Sp denote stable groups. [Note that $\Pi_k(\mathrm{SO}) = \Pi_k(\mathrm{SO}(m))$ for $m\geqslant k+2$, and $\Pi_k(\mathrm{Sp}) = \Pi_k(\mathrm{Sp}(m))$ for $m\geqslant(k-1)/4$.] Using these representations, we proceed to calculate their "winding numbers" for fundamental representations.

In Sec. III, we give formulas representing generators of the first nonstable homotopy groups of the classical Lie groups, i.e., the groups $\Pi_{n-2}(O(n-1))$, $\Pi_{2n-2}(U(n-1))$, or $\Pi_{4n-2}(\mathrm{Sp}(n-1))$. For unitary and symplectic groups, this completely describes the homotopy generators, but for orthogonal groups, our generator is only one of the generators in dimensions $n=8m+1$, $8m+2$, $8m+3$, and $8m+5$.

II. GENERATORS OF STABLE GROUPS AND WINDING NUMBERS

In various situations, we need to calculate the following "winding number" integral:

$$\int_{S^{2n+1}} \gamma^G_{2n+1}(f), \tag{2.1}$$

where the $(2n+1)$-form $\gamma^G_{2n+1}(f)$ is given by

$$(i/2\pi)^{n+1}[n!/(2n+1)!]\,\mathrm{Tr}(f^{-1}\,df)^{2n+1}. \tag{2.2}$$

The function f represents a nontrivial element in $\Pi_{2n+1}(G)$. If G is a classical group and f represents a homotopy element in the stable range, the value of this integral is independent of the rank of G. The integral is homotopy invariant and defines

 0022-2488/90/061494-09$03.00

a homomorphism $\Pi_{2n+1}(G)\to\mathbb{R}$ (real numbers). Thus we see that:

(i) If $\Pi_{2n+1}(G)$ is finite the integral vanishes.[5,8] [More generally, the torsion subgroup of $\Pi_{2n+1}(G)$ does not contribute to the integral.] Thus for stable SO and Sp groups, the integral is nonvanishing only for n odd.

(ii) The smallest absolute value of the integral is attained by the generator of an infinite cyclic summand of $\Pi_{2n+1}(G)$.

The value of the integral (2.1) must be an integer. We will prove this later using the Bott periodicity theorem.

The representation dependence of the integral (2.1) is given by

$$\int_{S^{2n+1}}\gamma^G_{2n+1}(f(\rho))=Q^G_{n+1}(\rho)\int_{S^{2n+1}}\gamma^G_{2n+1}(f(\Box)),$$

where $Q^G_{n+1}(\rho)$ is the Dynkin index[13] of a representation ρ. This fact follows from Appendix B and the identity[13]

$$\operatorname{Tr}F^{n+1}(\rho)=Q_{n+1}(\rho)\operatorname{Tr}F^{n+1}(\Box)+\sum a_{2m}(\rho)X_{2m}(\Box)X_{2n+2-2m}(\Box),$$

where $\Box$ denotes the fundamental representation, $X_{2\ell}$ denotes the invariant polynomial made of traces of a curvature 2 form $F(\Box)$, and $a_{2m}(\rho)$ denotes a numerical coefficient depending only on ρ. Using the fact that for $1\leqslant m<n+1$,

$$X_{2m}(\Box)=d\Omega^0_{2m-1}(\Box),$$

since the sphere had nontrivial de Rham cohomology $H^k(S^{2n+2})$ only for $k=0$ and $k=2n+2$, and that

$$dX_{2m}(\Box)=0,$$

we derive the desired result. Since the Dynkin index $Q_{n+1}(\rho)$ is an integer, if it is nonzero, the smallest absolute value of the integral (2.1) is attained by the fundamental representation.

In summary, if $f\colon S^{2n+1}\to G$ in the fundamental representation yields the value one for the integral (2.1), f is a representative of a generator of an infinite cyclic summand of $\Pi_{2n+1}(G)$.

As examples of this integral (2.1), we cite:

(i) Dirac's magnetic monopole, $f=e^{im\phi}$, $n=0$ and the integral is the monopole charge;[1]

(ii) f is a chiral soliton field and the integral above is the baryon number of this soliton;[1]

(iii) the value of the Wess–Zumino functional is given by Eq. (2.1) for a rotated soliton in dimension $(D+1)$;[14]

(iv) the global chiral Yang–Mills anomaly.[3–11]

To calculate the global anomaly of G with respect to a representation ω, we choose a group G' with representation $\tilde\omega$ satisfying the condition $G\subset G'$, $\Pi_D(G')=0$ and the representation $\tilde\omega$ reduces to ω plus singlets of G. For a connected simple classical Lie group G, one may choose G' as follows:[5–8]

(i) for $G=\mathrm{Su}(p)$, take $G'=\mathrm{SU}$;

(ii) for $G=\mathrm{Sp}(q)$, take $G'=\mathrm{Sp}$ if $D\equiv2 \pmod 4$ and $G'=\mathrm{SU}$ if $D\equiv0 \pmod 4$;

(iii) for $G=\mathrm{SO}(r)$, take $G'=\mathrm{SO}$ if $D\equiv2 \pmod 4$ and $G'=\mathrm{SU}$ if $D\equiv0 \pmod 4$.

For global anomalies of exceptional groups, see Ref. 5.

A. Clifford algebra construction

The method of constructing maps representing generators of the stable homotopy of the classical groups is to first construct a map $S^{n-1}\to\mathrm{Spin}(n)$ using a real or complex Clifford algebra and then take a representation $\mathrm{Spin}(n)\to G$, where G is a classical group. Consequently, we review the results we need on Clifford algebras and Spin(n). For a more complete account see Ref. 15.

Let $\mathbb{F}$ be one of the fields $\mathbb{R}$ (real numbers), $\mathbb{C}$ (complex numbers), or $\mathbb{H}$ (quaternions).

A Clifford algebra $\mathrm{Cl}_\pm(n)$ of a positive (+) or negative (−) definite quadratic form is a unital algebra that is generated multiplicatively by elements $\{\gamma_j^\pm\}$ satisfying the identities

$$\gamma_i^\pm\gamma_j^\pm+\gamma_j^\pm\gamma_i^\pm=\pm2\delta_{ij}1,\quad i,j=1,\dots,n. \tag{2.3}$$

Note that $\pm(\gamma_j^\pm)^2=1$. In general we will write $\gamma_j^\pm=\gamma_j$ in a situation independent of sign.

If $x=(x_1,x_2,\dots,x_n)\in\mathbb{R}^n$, we write

$$X=X(x)=\sum_{j=1}^n x_j\gamma_j\in\mathrm{Cl}_\pm(n),$$

so that $\mathrm{Cl}_\pm(n)$ consists of products $X_1X_2\cdots X_r$, where $X_k=\Sigma\, x_{k,j}\gamma_j$.

The Lie group Spin(n) is defined by

$$\mathrm{Spin}(n)=\left\{X_1X_2\cdots X_{2m}\,\middle|\,X_k=\sum_{j=1}^n x_{k,j}\gamma_j^- \text{ and } \sum_{j=1}^n x_{k,j}^2=1,\quad\text{for } 1\leqslant k\leqslant 2m\right\}.$$

Note that Spin(n) is a multiplicative subgroup of $\mathrm{Cl}_-(n)$.

Now let $\mathbb{F}(n)$ denote the matrix algebra of $n\times n$ matrices over $\mathbb{F}$. We recall that

$$\begin{aligned}&\mathrm{Cl}_-(8m)\cong\mathrm{Cl}_+(8m)\cong\mathbb{R}(2^{4m}),\\&\mathrm{Cl}_-(8m+4)\cong\mathrm{Cl}_+(8m+4)\cong\mathbb{H}(2^{4m+1}),\end{aligned}\tag{2.4}$$

and

$$\mathrm{Cl}_c(2m)\cong\mathrm{Cl}_\pm(2m)\otimes\mathbb{C}\cong\mathbb{C}(2^m). \tag{2.5}$$

This concludes our review and introduction of notation for Clifford algebras. For details see Ref. 15.

From now on we will restrict our attention to the universal Clifford algebras $\mathrm{Cl}_\pm(2n+2)$ and $\mathrm{Cl}_c(2n+2)$. We use $\mathrm{Cl}(2n+2)$ to collectively denote any of these algebras and regard them as matrix algebras as given above, except that we represent the quaternionic algebras $\mathbb{H}(n)\subset\mathbb{C}(2n)$ via a representation $\mathbb{H}\to\mathbb{C}(2)$.

For a numerical coefficient c, define

$$\gamma_{2n+3}=c\gamma_1\gamma_2\cdots\gamma_{2n+2}, \tag{2.6}$$

which satisfies

$$\begin{aligned}&(\gamma_{2n+3})^2=c^2(-1)^{(n+1)(2n+1)}\cdot1,\\&\gamma_j\gamma_{2n+3}+\gamma_{2n+3}\gamma_j=0,\quad\text{for } 1\leqslant j\leqslant 2n+2.\end{aligned}\tag{2.7}$$

For a complex Clifford algebra, or for a real Clifford algebra where n is odd, we can diagonalize γ_{2n+3}.

Define a map $g\colon S^{2n+1}\to\mathrm{Cl}(2n+2)\subset\mathbb{F}(2^{n+1})$ ($\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$) by[16]

$$g(x) = \gamma_{2n+2} \sum_{j=1}^{2n+2} \gamma_j x_j, \tag{2.8}$$

for $x = (x_1,...,x_{2n+2}) \in S^{2n+1}$.

In formula (2.8) any one of the γ_j can be substituted for γ_{2n+2}, or can be multiplied from the right. The final results are independent of these rearrangements.

Note that g: $S^{2n+1} \to \mathrm{Cl}_-(2n+2)$ has its image in Spin$(n+1)$ and

$$\hat{g}(x) = \sum_{j=1}^{2n+2} x_j \gamma_j \gamma_{2n+2} \tag{2.9}$$

is an inverse for $g(x)$, so that g: $S^{2n+1} \to \mathrm{GL}(2^{n+1}, \mathbb{F}) \subset \mathbb{F}(2^{n+1})$.

Also note that if $\gamma_j^{-1} = \bar{\gamma}_j^t$, the relation (2.8) implies that

$$g(x)^{-1} = \sum x_j \gamma_j \gamma_{2n+2} = \sum x_j \bar{\gamma}_j^t \, \overline{\gamma_{2n+2}}^t$$

$$= \sum x_j (\overline{\gamma_{2n+2}\gamma_j})^t = \overline{g(x)}^t.$$

An element of GL$(2^r, \mathbb{F})$ for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ is in a unitary or orthogonal group if and only if its inverse is its conjugate transpose. Since we will see later that we can choose our generators γ_j in a unitary or orthogonal group, we may choose $g(x)$ to be unitary or orthogonal.

From Eq. (2.7), we have

$$g(x)\gamma_{2n+3} = \gamma_{2n+3} g(x), \tag{2.10}$$

so that $g(x)$ and γ_{2n+3} preserve each other's eigenspaces. The eigenspaces V_+ of $+ci^{n+1}$ (the so-called *chiral* subspace) and V_- of $-ci^{n+1}$ have dimension 2^n because $\mathrm{Tr}\, \gamma_{2n+3} = 0$, and $g(x)$ is block diagonal with respect to V_+ and V_-. Define

$$g|V_+: S^{2n+1} \to \mathrm{GL}(2^n, \mathbb{F}), \tag{2.11}$$

by $g|V_+(x) = g(x)|V_+$. Then $g|V_+(x): V_+ \to V_+$ and

$$\mathrm{Tr}^{V_+} \gamma_{2n+3} = 2^n ci^{n+1}. \tag{2.12}$$

We show $g|V_+$: $S^{2n+1} \to \mathrm{SL}(2^n, \mathbb{F})$. From Eqs. (2.3) and (2.8) it follows that

$$g(x) + g(x)^{-1} = \pm 2x_{2n+2} 1_{2^{n+1}},$$

and the eigenvalues of $g(x)$ are

$$\pm (x_{2n+2} \pm i\sqrt{1 - x_{2n+2}^2}).$$

The eigenvalues of $g|V_+$ are among those of g, so $g|V_+$ has eigenvalues $\pm (x_{2n+2} \pm i\sqrt{1 - x_{2n+2}^2})$. Also,

$$\mathrm{Tr}\, g|V_+(x) = \sum x_j \, \mathrm{Tr}(\gamma_{2n+2}\gamma_j)|V_+$$

$$= \sum x_j \frac{1}{2} \mathrm{Tr}(\gamma_{2n+2}\gamma_j + \gamma_j\gamma_{2n+2})|V_+$$

$$= \pm \sum x_j \delta_{(2n+2)j} \, \mathrm{Tr}\, 1|V_+ = \pm 2^n x_{2n+2},$$

which is real. Thus the eigenvalues of $g|V_+(x)$ occur in conjugate pairs, and $\det g|V_+(x) = 1$.

We investigate the value for the integral (2.1) given by $g|V_+$. In order to evaluate the "winding number," we need to compute the form (2.2). We utilize the fact that

$$\mathrm{Tr}((g|V_+)^{-1} d(g|V_+))^{2n+1} = \mathrm{Tr}(g^{-1}\,dg)^{2n+1}|V_+$$
$$= \mathrm{Tr}^{V_+}(g^{-1}\,dg)^{2n+1}.$$

First note that

$$g^{-1}\,dg = \pm \sum x_p \, dx_q \, \gamma_p \gamma_q,$$

$$(g^{-1}\,dg)^2 = \mp \sum dx_l \, dx_n \, \gamma_l \gamma_n.$$

Then

$$(g^{-1}\,dg)^{2n+1} = (\mp 1)^{n+1} \sum \gamma_{j_1}\gamma_{j_2} \cdots \gamma_{j_{2n+1}} dx_{j_1} \times dx_{j_2} \cdots dx_{j_{2n+1}} \gamma_{j_{2n+2}} x_{j_{2n+2}},$$

where all j_k are all distinct, either because of the exterior multiplication of the forms dx_j or because of the relations $\Sigma x_k \, dx_k = 0$ and $\Sigma x_k^2 = 1$. Rearranging and collecting terms, we obtain

$$(g^{-1}\,dg)^{2n+1}$$
$$= (\mp 1)^{n+1}(2n+1)! \sum (-1)^j x_j \, dx_1 \, dx_2 \cdots \times d\hat{x}_j \cdots dx_{2n+2} \, \gamma_1\gamma_2 \cdots \gamma_{2n+2}$$
$$= \frac{(\mp 1)^{n+1}(2n+1)!}{c} \sum (-1)^j x_j \, dx_1 \, dx_2 \cdots \times d\hat{x}_j \cdots dx_{2n+2} \, \gamma_{2n+3},$$

where the circumflex indicates dx_j is omitted. Finally, we have

$$\mathrm{Tr}^{V_+}(g^{-1}\,dg)^{2n+1}$$
$$= [(\mp 1)^{n+1}(2n+1)!/c] \, \mathrm{Tr}^{V_+} \gamma_{2n+3} \, dV$$
$$= (\mp i)^{n+1}(2n+1)! 2^n \, dV, \tag{2.13}$$

where dV is the volume of S^{2n+1}. Thus (2.1) has the value

$$\left(\frac{i}{2\pi}\right)^{n+1} \frac{n!}{(2n+1)!} \int_{S^{2n+1}} (\mp i)^{n+1}(2n+1)! 2^n \, dV$$
$$= (\pm 1)^{n+1} \frac{n!}{2\pi^{n+1}} \int_{S^{2n+1}} dV = (\pm 1)^{n+1}, \tag{2.14}$$

because the volume of S^{2n+1} is $2\pi^{n+1}/n!$. This result is independent of how we normalize γ_{2n+3}, since c cancels in the final expression. [If we choose V_- instead of V_+, then (2.1) has the value $-(\pm 1)^{n+1}$]. Thus, assuming (2.1) has integral values, $g|V_+$ represents a homotopy generator.

B. The special unitary group

We construct a representative for the generator of Π_{2n+1}(SU). A formula for a representative of a generator of Π_{2n+1}(SU$(n+1)$) has been constructed by one of us (A.L.),[17] but the calculation of the integral, using this formula, turned out to be formidable.

We first construct a suitable set of generators $\{\gamma_1,...,\gamma_{2n+2}\}$ for $\mathrm{Cl}_c(2n+2) \equiv \mathrm{Cl}_+(2n+2) \otimes \mathbb{C}$, give the maps $g_\mathbb{C}$ and $g_\mathbb{C}|V_+$, and use the Bott periodicity theorem to show *topologically* that $g_\mathbb{C}|V_+$ represents a homotopy generator of SU.

For $n = 0$, the Hermitian gamma matrices are Pauli matrices, i.e.,

$$\gamma_1^{(1)} = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma_2^{(1)} = \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

and

$$\gamma_3^{(1)} = -i\sigma_1\sigma_2 = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Suppose we have defined the matrices $\gamma_j^{(k)}$ for $k \geqslant 1$ and $j = 1,2,...,2k$. We define

$$\gamma_{2k+1}^{(k)} \equiv -i^k \gamma_1^{(k)} \cdots \gamma_{2k}^{(k)}, \tag{2.15}$$

which satisfies

$$\gamma_{2k+1}^{(k)}\gamma_j^{(k)} + \gamma_j^{(k)}\gamma_{2k+1}^{(k)} = 0.$$

Next we define

$$\begin{aligned} &\gamma_j^{(k+1)} = \gamma_j^{(k)} \otimes \sigma_2, \quad j = 1,...,2k+1, \\ &\gamma_{2k+2}^{(k+1)} = 1_2^{(k)} \otimes \sigma_1. \end{aligned} \tag{2.16}$$

with $1_2 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. It is easy to see that all these γ matrices are Hermitian, unitary, satisfy the relations (2.3), generate a complex universal Clifford algebra, and that $\gamma_j^{(n+1)}$ is a $2^{n+1} \times 2^{n+1}$ matrix. In particular,

$$\gamma_{2n+3}^{(n+1)} = 1_2^{(n)} \otimes \sigma_3,$$

and therefore $\gamma_{2n+3}^{(n+1)}$ has eigenvalues ± 1 and its square is the identity. The $+1$ eigenspace V_+ of $\gamma_{2n+3}^{(n+1)}$ is the *chiral* subspace and is of dimension 2^n.

Now we observe that

$$\begin{aligned} g(\square)(x) = g_{\mathrm{C}}(x) &= \sum_{j=1}^{2n+2} \gamma_{2n+2}^{(n+1)}\gamma_j^{(n+1)} x_j \\ &= x_{2n+2} 1_{2^n} \otimes 1_2 + \sum_{j=1}^{2n+1} i x_j \gamma_j^{(n)} \otimes \sigma_3, \end{aligned}$$

so that

$$g_{\mathrm{C}} | V_+ = g_{\mathrm{C}}(x) | V_+ = x_{2n+2} 1_{2^n} + \sum_{j=1}^{2n+1} x_j i \gamma_j^{(n)}. \tag{2.17}$$

In the case $n = 1$, we obtain a well-known generator of $\Pi_3(\mathrm{SU}(2))$,

$$\begin{aligned} g_{\mathrm{C}} | V_+(x) &= i x_1 \sigma_1 + i x_2 \sigma_2 + i x_3 \sigma_3 + x_4 1_2 \\ &= \begin{pmatrix} x_4 + i x_3 & x_2 + i x_1 \\ -x_2 + i x_1 & x_4 - i x_3 \end{pmatrix}. \end{aligned}$$

We show that our $g_{\mathrm{C}} | V_+$ is homotopic to the topological construction of a stable generator in the stable homotopy of SU.

We begin with an inductive description of generators of $\Pi_{2n+1}(\mathrm{SU})$. Let λ: $U(n) \to U(2n)$ be the embedding given in terms of matrices as

$$\lambda(M) = \begin{pmatrix} M & 0 \\ 0 & 1_n \end{pmatrix},$$

and for $(s,t) \in I^2 = \{(u,v) \in \mathbb{R}^2 | 0 \leqslant u \leqslant 1, 0 \leqslant v \leqslant 1\}$, let $W^{(n)}(s,t) \in \mathrm{SU}(2n)$ be the block matrix

$$W(s,t) = W^{(n)}(s,t) = \begin{pmatrix} \alpha(s,t) 1_n & \beta(s,t) 1_n \\ -\overline{\beta(s,t)} 1_n & \overline{\alpha(s,t)} 1_n \end{pmatrix}, \tag{2.18}$$

where $\alpha(s,t) = \cos \pi s + i \sin \pi s \cos \pi t$, and $\beta(s,t) = i \sin \pi s \sin \pi t$. Now from (2.17) if we abbreviate $g^{(n)} = g_{\mathrm{C}} | V_+ : S^{2n-1} \to \mathrm{SU}(2^{n-1})$, then

$$g^{(n)} = x_{2n} 1_{2^{n-1}} + \sum_{j=1}^{2n-1} i x_j \gamma_j^{(n-1)}$$

and

$$(g^{(n)})^{-1} = x_{2n} 1_{2^{n-1}} - \sum_{j=1}^{2n-1} i x_j \gamma_j^{(n-1)}.$$

We show that the conjugate $\lambda((g^{(n)}(x))^{-1}) \times W(s,t) \lambda(g^{(n)}(x))$ is the map $g^{(n+1)} = g_{\mathrm{C}} | V_+$: $S^{2n+1} \to \mathrm{SU}(2^n)$. We calculate

$$\begin{aligned} &\lambda((g^{(n)}(x))^{-1}) W^{(2^{n-1})}(s,t) \lambda(g^{(n)}(x)) \\ &= \begin{pmatrix} \alpha(s,t) 1_{2^{n-1}} & \beta(s,t) g^{(n)}(x)^{-1} \\ -\overline{\beta(s,t)} g^{(n)}(x) & \overline{\alpha(s,t)} 1_{2^{n-1}} \end{pmatrix} \\ &= \cos \pi s 1_{2^{n-1}} \otimes 1_2 + i \sin \pi s \cos \pi t\, 1_{2^{n-1}} \otimes \sigma_3 \\ &\quad + i x_{2n} \sin \pi s \sin \pi t\, 1_{2^{n-1}} \otimes \sigma_1 \\ &\quad + \sum_{j=1}^{2n-1} i x_j \sin \pi s \sin \pi t \gamma_j^{(n-1)} \otimes \sigma_2 \\ &= y_{2n+2} 1_{2^{n-1}} \otimes 1_2 + i y_{2n+1} 1_{2^{n-1}} \otimes \sigma_3 \\ &\quad + i y_{2n} 1_{2^{n-1}} \otimes \sigma_1 + \sum_{j=1}^{2n-1} i y_j \gamma_j^{(n-1)} \otimes \sigma_2 \\ &= g^{(n+1)}(y_1, y_2, ..., y_{2n+2}), \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} y_j &= x_j \sin \pi s \sin \pi t, \quad \text{for } 1 \leqslant j \leqslant 2n, \\ y_{2n+1} &= \sin \pi s \cos \pi t, \\ y_{2n+2} &= \cos \pi s. \end{aligned}$$

It is worth observing that if $x_1, x_2, ..., x_{2n}$ are spherical polar coordinates on S^{2n-1}, then $y_1, y_2, ..., y_{2n+2}$ are spherical polar coordinates on S^{2n+1}. If we start with $g^{(1)}$: $S^1 \to U(1)$ the identity map

$$g^{(1)}(\cos 2\pi t + i \sin 2\pi t) = (\cos 2\pi t + i \sin 2\pi t) 1_1,$$

we can construct $g^{(n)}$ by repeated conjugation with the matrices W.

To see the relationship with the Bott periodicity, we recall that one can define a homomorphism B_n: $\Pi_m(U(n)) \to \Pi_{m+2}(\mathrm{SU}(2n))$ as follows.[17–19] For $z = (0,0,...,0,1) \in S^m$ and f: $(S^m, z) \to (U(n), 1_n)$, define $\hat{f}$: $S^m \times I \to \mathrm{SU}(2n)$ by

$$\hat{f}(x,s,t) = \lambda(f(x)) V(s,t) \lambda(f(x)^{-1}) V(s,t)^{-1}. \tag{2.20}$$

The map $\hat{f}$ has the property that

$$\hat{f}(S^m \times \partial I^2 \cup z \times I^2) = 1_{2^n},$$

and hence defines a map $\hat{f}$: $(S^{n+2}, z') \to (\mathrm{SU}(2n), 1_{2n})$. The assignment

$$B_n([f]) = [\tilde{f}],$$

where the brackets denote homotopy class, is a homomorphism. The Bott periodicity theorem states that if $m \leqslant 2n-1$, then B_n: $\Pi_m(U(n)) \to \Pi_{m+2}(\mathrm{SU}(2n))$ is an isomorphism.

We show that $g^{(n)}$ constructed iteratively is homotopic to the Bott construction of the generator in the stable homotopy of SU. First of all, the matrices $V(s,t)$ as given in Ref. 17 differ from $W(s,t)$ by a minus sign on the coefficients $\beta(s,t)$. One easily corrects this by a homotopy [multiply the $\beta(s,t)$ by $e^{\pi iu}$ for $0 \leqslant u \leqslant 1$ to deform $\beta(s,t)$ to $-\beta(s,t)$]. Secondly, the Bott construction involves a commutator, whereas the construction we gave involves a conjugate. Again this is a minor matter involving the difference between a reduced and a nonreduced suspension, which in our case does not effect homotopy.

Thus we have a *topological* proof that the maps $g_{\mathbb{C}} | V_+$ represent generators in the stable homotopy of SU. This enables us to see that in order to have the winding number of a generator of stable homotopy ± 1, we must have the numerical factor $(i/2\pi)^{n+1} n!/(2+1)!$ in the form (2.2), and the integral (2.1) is always an integer in the unitary case. That the integral is an integer for the other classical groups follows from their embeddings in the unitary group.

C. The special orthogonal group

In order to give generators for the homotopy group of SO and compute their "winding numbers," we can attempt to parallel the calculations in the case of SU. The additional complexity of universal *real* Clifford algebras (called real Clifford algebra hereafter) force us to also compare the results for SU using the embeddings ι: SU→SO and κ: SO→SU.

1. Clifford algebra considerations

We recall from the constructions of Sec. II A that for $G = \mathrm{SO}$ we are interested in *real* Clifford algebras $\mathrm{Cl}_\pm(2n+2)$ satisfying the condition that the product of the generators γ_{2n+3} can be made diagonal. This restricts us further to the cases of the Clifford algebras $\mathrm{Cl}_\pm(4n)$. Looking at the table (2.4) we see $\mathrm{Cl}_\pm(8m)$ has a real representation in $\mathbb{R}(2^{4m})$, while $\mathrm{Cl}_\pm(8m+4)$ has a real representation in $\mathbb{R}(2^{4m+3})$. In the latter case, when we compute the form and integrate for the winding number, the effect of this doubling of the dimension of the representation space and will be to multiply the form and hence the winding number by two. As we will see, this is exactly what is forced by homotopy theory. Hence in both dimensions $8m-1$ and $8m+3$ maps constructed by real Clifford algebras represent generators of the homotopy groups of SO.

2. Generators of $\Pi_k(SO)$ with $k=8m+1$ and $8m+3$

We first consider the embedding ι: SU→SO. In terms of matrices this can be described as follows. If with respect to a choice of coordinates, $M = A + iB \in \mathrm{SU}(n)$, then

$$\iota(M) = 1_2 \otimes A + \rho_2 \otimes B.$$

Now $\iota \circ g_{\mathbb{C}}: S^{2n+1} \to \mathrm{SO}(2^{n+2})$ and $\iota \circ (g_{\mathbb{C}} | V_+)$: $S^{2n+1} \to \mathrm{SO}(2^{n+1})$. From Bott periodicity and the homotopy exact sequences,

$n \pmod 4$	$\Pi_{2n+2}(\mathrm{SO/SU})$	→	$\Pi_{2n+1}(\mathrm{SU})$	→	$\Pi_{2n+1}(\mathrm{SO})$	→	$\Pi_{2n+1}(\mathrm{SO/SU})$
-1	$\mathbb{Z}_2$		$\mathbb{Z}$	one-to-one	$\mathbb{Z}$		$\mathbb{Z}_2$
0	$\mathbb{Z}$		$\mathbb{Z}$	onto	$\mathbb{Z}_2$		0
1	0		$\mathbb{Z}$	≅	$\mathbb{Z}$		0
2	$\mathbb{Z}$		$\mathbb{Z}$	trivial	0		0 ,

we see that $\iota \circ (g_{\mathbb{C}} | V_+)$ represents a generator of $\Pi_{2n+1}(\mathrm{SO})$ for $n \equiv 0,1 \pmod 4$, i.e., in dimensions $8m+1$ and $8m+3$. In case $n \equiv -1 \pmod 4$, the map $\iota \circ (g_{\mathbb{C}} | V_+)$ represents *twice* a generator.

3. A generator of $\Pi_{8m-1}(SO)$

We construct a generating set for the real universal Clifford algebra $\mathrm{Cl}_\pm(8m) \cong \mathbb{R}(2^{4m})$. The generators γ_j will lie in $O(2^{4m})$.

Let

$$\rho_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \rho_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

let

$$\begin{aligned}
\tau_1^+ &= \rho_2 \otimes \rho_2 \otimes \rho_2 \otimes \rho_2, & \tau_1^- &= \rho_2 \otimes \rho_2 \otimes \rho_2 \otimes \rho_1,\\
\tau_2^+ &= 1_2 \otimes \rho_1 \otimes \rho_2 \otimes \rho_2, & \tau_2^- &= 1_2 \otimes \rho_1 \otimes \rho_2 \otimes \rho_1,\\
\tau_3^+ &= \rho_2 \otimes 1_2 \otimes \rho_1 \otimes \rho_2, & \tau_3^- &= \rho_2 \otimes 1_2 \otimes \rho_1 \otimes \rho_1,\\
\tau_4^+ &= 1_2 \otimes \rho_3 \otimes \rho_2 \otimes \rho_2, & \tau_4^- &= 1_2 \otimes \rho_3 \otimes \rho_2 \otimes \rho_1,\\
\tau_5^+ &= \rho_1 \otimes \rho_2 \otimes 1_2 \otimes \rho_2, & \tau_5^- &= \rho_1 \otimes \rho_2 \otimes 1_2 \otimes \rho_1,\\
\tau_6^+ &= \rho_3 \otimes \rho_2 \otimes 1_2 \otimes \rho_2, & \tau_6^- &= \rho_3 \otimes \rho_2 \otimes 1_2 \otimes \rho_1,\\
\tau_7^+ &= \rho_2 \otimes 1_2 \otimes \rho_3 \otimes \rho_2, & \tau_7^- &= \rho_2 \otimes 1_2 \otimes \rho_3 \otimes \rho_1,\\
\tau_8^+ &= 1_2 \otimes 1_2 \otimes 1_2 \otimes \rho_1, & \tau_8^- &= 1_2 \otimes 1_2 \otimes 1_2 \otimes \rho_2,
\end{aligned} \tag{2.21}$$

and set

$$\tau_9^\pm = \tau_9 = -\tau_1\tau_2 \cdots \tau_8 = 1_2 \otimes 1_2 \otimes 1_2 \otimes \rho_3.$$

Then, $\tau_1,...,\tau_8,\tau_9 \in O(2^4)$ satisfy the relations $\tau_i^\pm \tau_j^\pm + \tau_j^\pm \tau_i^\pm = \pm 2\delta_{ij}$ for $i,j = 1,...,8$ while $\tau_9 \tau_j^\pm + \tau_j^\pm \tau_9 = 0$ and $\tau_9^2 = 1_{16}$. Now define

$$\begin{aligned}
&\theta_j^{\pm(1)} = \tau_j^\pm, \quad \text{for } 1 \leqslant j \leqslant 9,\\
&\theta_j^{\pm(m+1)} = \theta_j^{+(m)} \otimes \tau_1^\pm, \quad \text{for } 1 \leqslant j \leqslant 8m+1,\\
&\theta_{8m+r}^{\pm(m+1)} = 1_{16}^{(m)} \otimes \tau_r^\pm, \quad \text{for } 2 \leqslant r \leqslant 9.
\end{aligned} \tag{2.22}$$

Then

$$\theta_{8m+1}^{(m)} = \theta_1^{(m)} \theta_2^{(m)} \cdots \theta_{8m}^{(m)} = 1_{16}^{(m-1)} \otimes 1_8 \otimes \rho_3.$$

The $\theta_j^{(m)}$ for $1 \leqslant j \leqslant 8m$ satisfy the relation $\theta_k^{(m)} \theta_l^{(m)} + \theta_l^{(m)} \theta_k^{(m)} = \pm 2\delta_{kl}$ and generate a real universal Clifford algebra $\mathrm{Cl}_\pm(8m)$. We also have $\theta_{8m+1}^{(m)} \theta_j^{(m)} + \theta_j^{(m)} \theta_{8m+1}^{(m)} = 0$ and $(\theta_{8m+1}^{(m)})^2 = 1_{16}^{(m)}$.

We now define $g_{\mathbb{R}} = g_{\mathbb{R}}(\square): S^{8m-1} \to O(2^{4m})$ by

$$g_{\mathbb{R}}(x_1,x_2,...,x_{8m}) = \sum_{j=1}^{8m} x_j \theta_j^{(m)} \theta_{8m}^{(m)}.$$

From Sec. II A, $\det g_{\mathbf{R}}(x)=1$, so that $g_{\mathbf{R}}: S^{8m-1} \to SO(2^{4m})$. As in the complex case we are interested in the restriction of $g_{\mathbf{R}}$ to the $+1$ eigenspace V_+ of $\theta^{(m)}_{8m+1}$. The argument of Sec. II A shows that $g_{\mathbf{R}}|V_+: S^{8m-1} \to SO(2^{4m-1})$, where

$$(g_{\mathbf{R}}|V_+)(x) = x_{8m} 1^{(m-1)}_{16} \otimes 1_8 + \sum_{r=2}^{7} x_{8(m-1)+r} 1^{(m-1)}_{16} \otimes \hat{\tau}_r + \sum_{j=1}^{8m-7} x_j \theta^{(m-1)}_j \otimes \hat{\tau}_1, \qquad (2.23)$$

where $\tau^{\pm}_j = \hat{\tau}_j \otimes \rho_k$.

If κ: SO→SU is the embedding given by considering a real matrix as a complex one, then for $\kappa\circ(g_{\mathbf{R}}|V_+)$: $S^{8m-1}\to SU(2^{4m-1})$, the form $\mathrm{Tr}^{V_+}(g_{\mathbf{R}}^{-1}\, dg_{\mathbf{R}})^{8m-1}$ is the same as Eq. (2.31). Thus $\kappa\circ(g_{\mathbf{R}}|V_+)$ represents a generator of $\Pi_{8m-1}(SU)$. Since the sequence $0=\Pi_{8m}(SU/SO)\to\Pi_{8m-1}(SO)\to^{\kappa_*}\Pi_{8m-1}(SU)\to\Pi_{8m-1}(SU/SO)=0$ is exact, $g_{\mathbf{R}}|V_+$ represents a generator of $\Pi_{8m-1}(SO)$.

For the case $n=1$, the map $g_{\mathbf{R}}|V_+: S^7\to SO(8)$ has the form

$$(g_{\mathbf{R}}|V_+)(x) = \begin{pmatrix} x_8 & x_7 & x_6 & x_5 & x_4 & x_3 & x_2 & x_1 \\ -x_7 & x_8 & x_5 & -x_6 & -x_3 & x_4 & -x_1 & x_2 \\ -x_6 & -x_5 & x_8 & x_7 & x_2 & -x_1 & -x_4 & x_3 \\ -x_5 & x_6 & -x_7 & x_8 & x_1 & x_2 & -x_3 & -x_4 \\ -x_4 & x_3 & -x_2 & -x_1 & x_8 & -x_7 & x_6 & x_5 \\ -x_3 & -x_4 & x_1 & -x_2 & x_7 & x_8 & x_5 & -x_6 \\ -x_2 & x_1 & x_4 & x_3 & -x_6 & -x_5 & x_8 & -x_7 \\ -x_1 & -x_2 & -x_3 & x_4 & -x_5 & x_6 & x_7 & x_8 \end{pmatrix}$$

and is homotopic to a cross section of the bundle p: $SO(8)\to S^7$ [p chooses the first row of a matrix in SO(8)], and will represent a stable generator when included in $SO(n)$ for $n \geqslant 9$.

4. Generators of Π_k(SO) with $k=8m$, $8m+1$

It is known from the work of Kervaire[20] that the generator of $\Pi_{8m}(SO)$ is represented by the composition $\beta\circ\eta_{8m-1}$ where η_{8m-1} is the class of an $(8m-3)$-fold suspension of the Hopf map $S^3\to S^2$, while the generator of $\Pi_{8m+1}(SO)$ is represented by the composition $\beta\circ\eta_{8m-1}\circ\eta_{8m}$. Observe that we now have two representations of the generator of $\Pi_{8m+1}(SO)$. Note that η_{8m-1} is the generator of $\Pi_{8m}(S^{8m-1})$ and $\eta_{8m-1}\circ\eta_{8m}$ is the generator of $\Pi_{8m+1}(S^{8m-1})$. Since we have decribed a representative $S^{8m-1}\to SO$ of a generator of a stable homotopy group, it will suffice to give a representative $h^{(m)}$ of $\eta_m \in \Pi_m(S^{m-1})$. An easy modification of a formula by Steenrod[21] gives the formula for $h^{(m)}$ (see Appendix A).

The generator for $\Pi_1(SO)=\mathbb{Z}_2$ is represented by the map f: $S^1\to SO(3)\to SO$, where

$$f(\cos 2\pi t, \sin 2\pi t) = \begin{pmatrix} \cos 2\pi t & \sin 2\pi t & 0 \\ -\sin 2\pi t & \cos 2\pi t & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with $0<t<1$.

We summarize representatives of generators of $\Pi_k(SO)$ with $k=8m-1$, $8m$, $8m+1$, $8m+3$ as follows:

$$g_{8m-1}(x) = \sum_{j=1}^{8m} x_j (\theta^{(m)}_j \theta^{(m)}_{8m})|V_+,$$

$$g_{8m}(x) = \sum_{j=1}^{8m} h^{(8m-1)}_j(x) (\theta^{(m)}_j \theta^{(m)}_{8m})|V_+,$$

$$g_{8m+1}(x) = \begin{cases} \sum_{j=1}^{8m} h^{(8m-1)}_j (h^{(8m)}(x)) (\theta^{(m)}_j \theta^{(m)}_{8m})|V_+, \\ \sum_{j=1}^{8m+2} x_j \iota((\gamma^{(4m+1)}_{8m+2} \gamma^{(4m+1)}_j)|V_+), \end{cases}$$

$$g_{8m+3}(x) = \sum_{j=1}^{8m+4} x_j \iota((\gamma^{(4m+2)}_{8m+4} \gamma^{(4m+2)}_j)|V_+). \qquad (2.24)$$

5. Winding numbers for SO

For the calculation of "winding numbers," we use the property that the inclusion map ι is a homomorphism and commutes with the exterior differentiation. Thus, if $\alpha\in\Pi_{4n+3}(SU)$ and $\beta\in\Pi_{4n+3}(SO)$ are generators, $\iota_*(\alpha)=\beta$ (n even) or 2β (n odd) and

$$\begin{aligned}(\iota^*\gamma^{SO})(g_{\mathbf{C}}|V_+) &= \gamma^{SO}(\iota\circ(g_{\mathbf{C}}|V_+)) \\ &= (\mp i)^{n+1}(2n+1)!\,\mathrm{Tr}\,\iota(\gamma^{(n+1)}_{2n+3}|V_+)dV \\ &= 2\gamma^{SU}(g_{\mathbf{C}}|V_+),\end{aligned}$$

and therefore

$$\gamma^{SO}(\iota_*\alpha) = (\iota^*\gamma^{SO})(\alpha) = 2\gamma^{SU}(\alpha).$$

Hence, we have

$$\int_{S^{4n+3}} \iota^*\gamma^{SO}(\alpha) = \int_{S^{4n+3}} \gamma^{SU}(\alpha) = \pm 2.$$

Thus, "winding numbers" for our choice of generators of stable homotopy groups of orthogonal groups are given by

$$\int_{S^{4n+3}} \gamma^{SO}_{4n+3}(\beta) = \begin{cases} \pm 2, & \text{for } n \text{ even}, \\ \pm 1, & \text{for } n \text{ odd}. \end{cases} \qquad (2.25)$$

For n odd, we have already obtained this result, using $\mathrm{Tr}^{V_+}(g_{\mathbf{R}}^{-1}\, dg_{\mathbf{R}})^{8m-1}$. For n even, we see that $g_{\mathbf{R}}|V_+$ constructed by the orthogonal Clifford algebra generators of $Cl_{\pm}(8m+4)$ also represents a generator of $\Pi_{8m+3}(SO)$.

D. The symplectic group

We give representations of homotopy generators and corresponding "winding numbers" for $G=$ Sp. Again, we exploit our calculations for SU and do our constructions in parallel with those for SO.

We first describe the group Sp(n) as the subgroup of SU($2n$) consisting of matrices M such that

$$(\rho_2\otimes 1_n)M(\rho_2\otimes 1_n)^{-1} = \overline{M}.$$

[Our Sp(n) is a so-called unitary symplectic group.]

From the classification of real Clifford algebras, it is possible to represent the algebras $Cl_{\pm}(8m+4)$ by *symplectic* matrices in $GL(2^{4m+2},\mathbb{C})$ so that the product of generators of $Cl_{\pm}(8m+4)$ is diagonal. If we follow the real representation of $Cl_{\pm}(8m)$ in $GL(2^{4m},\mathbb{R})$ by the inclusions $GL(2^{4m},\mathbb{R})\xrightarrow{\kappa}GL(2^{4m},\mathbb{C})\xrightarrow{\iota}GL(2^{4m+1},\mathbb{R})\xrightarrow{\kappa}GL\times(2^{4m+1},\mathbb{C})$ (i.e., $M\to 1_2\otimes M$), it turns out that we have a symplectic representation in $GL(2^{4m+1},\mathbb{C})$ for $Cl_{\pm}(8m)$. When we calculate the form (2.2) for this representation, the trace operation now has the effect of multiplying our original form by two. This multiplies the winding number by two.

1. Generators of $\Pi_k(Sp)$ with $k=8m-1$ and $8m+5$

For a choice of basis the inclusion map ι: $SU(n)\to Sp(n)$ may be given by

$$\iota(M)=1_1\otimes A+i\rho_3\otimes B,$$

where $M=A+iB\in SU(n)$.

Now $\iota\circ g_{\mathbb{C}}$: $S^{2n+1}\to Sp(2^{n+1})$ and $\iota\circ(g_{\mathbb{C}}|V_+)$: $S^{2n+1}\to Sp(2^n)$. From the Bott periodicity and the homotopy exact sequences,

$n \pmod 4$	$\Pi_{2n+2}(Sp/SU)$	$\to$	$\Pi_{2n+1}(SU)$	$\to$	$\Pi_{2n+1}(Sp)$	$\to$	$\Pi_{2n+1}(Sp/SU)$
-1	0		$\mathbb{Z}$	$\cong$	$\mathbb{Z}$		0
0	$\mathbb{Z}$		$\mathbb{Z}$	trivial	0		0
1	$\mathbb{Z}_2$		$\mathbb{Z}$	one-to-one	$\mathbb{Z}$		$\mathbb{Z}_2$
2	$\mathbb{Z}$		$\mathbb{Z}$	onto	$\mathbb{Z}_2$		0,

we see that $\iota\circ(g_{\mathbb{C}}|V_+)$ represents a generator of $\Pi_{2n+1}(Sp)$ for $n\equiv -1,2 \pmod 4$, i.e., in dimensions $8m-1$ and $8m+5$. In case $n\equiv 1 \pmod 4$, dimension $8m+3$, the map $\iota\circ(g|V_+)$ represents *twice* a generator of $\Pi_{8m+3}(Sp)$.

2. A generator of $\Pi_{8m+3}(Sp)$

It is possible to construct a symplectic generating set for the real Clifford algebra $Cl_{\pm}(8m+4)\cong\mathbb{H}(2)\otimes\mathbb{R}(16)^m$ by recalling that $\mathbb{H}$ may be represented in $\mathbb{C}(2)$ by representing a quaternion $x_1+ix_2+jx_3+kx_4$ as $x_1 1_2+x_2\rho_2+x_3 i\rho_1+x_4 i\rho_3$. Now define

$$\begin{aligned}
&\phi_1^{+(0)}=\rho_2\otimes\rho_2,\quad \phi_1^{-(0)}=\rho_2\otimes\rho_1,\\
&\phi_2^{+(0)}=i\rho_1\otimes\rho_2,\quad \phi_2^{-(0)}=i\rho_1\otimes\rho_1,\\
&\phi_3^{+(0)}=i\rho_3\otimes\rho_2,\quad \phi_3^{-(0)}=i\rho_3\otimes\rho_1,\\
&\phi_4^{+(0)}=1_2\otimes\rho_1,\quad \phi_4^{-(0)}=1_2\otimes\rho_2,
\end{aligned}\tag{2.26}$$

and for $m\geqslant 1$,

$$\begin{aligned}
&\phi_j^{+(m)}=\rho_2\otimes\theta_j^{+(m)}\otimes\rho_2,\quad \phi_j^{-(m)}=\rho_2\otimes\theta_j^{+(m)}\otimes\rho_1,\\
&\phi_{8m+2}^{+(m)}=i\rho_1\otimes 1_{16}^{(m)}\otimes\rho_2,\quad \phi_{8m+2}^{-(m)}=i\rho_1\otimes 1_{16}^{(m)}\otimes\rho_1,\\
&\phi_{8m+3}^{+(m)}=i\rho_3\otimes 1_{16}^{(m)}\otimes\rho_2,\quad \phi_{8m+3}^{-(m)}=i\rho_3\otimes 1_{16}^{(m)}\otimes\rho_1,\\
&\phi_{8m+4}^{+(m)}=1_2\otimes 1_{16}^{(m)}\otimes\rho_1,\quad \phi_{8m+4}^{-(m)}=1_2\otimes 1_{16}^{(m)}\otimes\rho_2,
\end{aligned}\tag{2.27}$$

for $1\leqslant j\leqslant 8m+1$.

Then

$$\phi_{8m+5}^{(m)}=\phi_1^{\pm(m)}\cdots\phi_{8m+4}^{\pm(m)}=1_2\otimes 1_{16}^{(m)}\otimes\rho_3,$$

which is diagonal.

The $\phi_k^{(m)}$ are evidently in $U(2^{4m+2})$, and since

$$(\rho_2\otimes 1_2^{(4m+1)})\phi_k^{(m)}(\rho_2\otimes 1_2^{(4m+1)})^{-1}=\overline{\phi_k^{(m)}},$$

we have

$$\phi_k^{(m)}\in Sp(2^{2m+1})\subset SU(2^{2m+2}).$$

Note that

$$\phi_k^{\pm(m)}\phi_l^{\pm(m)}+\phi_l^{\pm(m)}\phi_k^{\pm(m)}=\pm 2\delta_{kl},$$

for $1\leqslant k,l\leqslant 8m+4$ and the $\phi_k^{\pm(m)}$ generate a universal Clifford algebra $Cl_{\pm}(8m+4)$. Also note that $\phi_k^{\pm(m)}\phi_{8m+5}^{(m)}+\phi_{8m+5}^{(m)}\phi_k^{\pm(m)}=0$ and $(\phi_{8m+5}^{(m)})^2=1_{2^{4m+2}}$.

Now define $g_{\mathbb{H}}=g_{\mathbb{H}}(\square)$: $S^{8m+3}\to Sp(2^{4m+1})$ by

$$g_{\mathbb{H}}(x_1,\ldots,x_{8m+4})=\sum_{j=1}^{8m+4}x_j\phi_j^{(m)}\phi_{8m+4}^{(m)}.$$

Then, if V_+ is the $+1$ eigenspace of $\phi_{8m+5}^{(m)}$, we have

$$\begin{aligned}
(g_{\mathbb{H}}|V_+)(x)=&\,x_{8m+4}1_2\otimes 1_{16}^{(m)}+x_{8m+3}i\rho_3\otimes 1_{16}^{(m)}\\
&+x_{8m+2}i\rho_1\otimes 1_{16}^{(m)}+x_{8m+1}\rho_2\otimes 1_{16}^{(m-1)}\\
&\otimes 1_8\otimes\rho_3+\sum_{j=1}^{8m}x_j\rho_2\otimes\theta_j^{(m)},
\end{aligned}\tag{2.28}$$

which represents a generator of $\Pi_{8m+3}(Sp)$. For the case $m=0$ $g_{\mathbb{H}}|V_+$ is a well-known generator of $SU(2)=Sp(1)$.

3. Generators of $\Pi_k(Sp)$ with $k=8m+4$ and $8m+5$

As in the case of SO, the generator of $\Pi_{8m+4}(Sp)$ is given by the composition $\beta\circ\eta_{8m+3}$ where η_{8m+3} is the class of an $(8m+1)$-fold suspension of the Hopf map $S^3\to S^2$; and the generator of $\Pi_{8m+5}(Sp)$ is given by the composition $\beta\circ\eta_{8m+3}\circ\eta_{8m+4}$. Again, we have two representations of the generator of $\Pi_{8m+5}(Sp)$.

We summarize these formulas for a function g_k representing a generator $\beta_k\in\Pi_k(Sp)$ for $k=8m-1$, $8m+3$, $8m+4$, $8m+5$. Thus

$$\begin{aligned}
g_{8m-1}(x)&=\sum_{j=1}^{8m}x_j\iota((\gamma_{8m}^{(4m)}\gamma_j^{(4m)})|V_+),\\
g_{8m+3}(x)&=\sum_{j=1}^{8m+4}x_j(\phi_j^{(m)}\phi_{8m+4}^{(m)})|V_+,\\
g_{8m+4}(x)&=\sum_{j=1}^{8m+4}h_j^{(8m+3)}(x)(\phi_j^{(m)}\phi_{8m+4}^{(m)})|V_+,\\
g_{8m+5}(x)&=\begin{cases}\sum_{j=1}^{8m+6}h_j^{(8m+4)}(h^{(8m+4)}(x))(\phi_j^{(m)}\phi_{8m+4}^{(m)})|V_+,\\ \sum_{j=1}^{8m+6}x_j\iota((\gamma_{8m+4}^{(4m+2)}\gamma_j^{(4m+2)})|V_+).\end{cases}
\end{aligned}\tag{2.29}$$

4. Winding numbers for Sp

The inclusion map ι is a homomorphism and commutes with exterior differentiation. Thus, if $\alpha\in\Pi_{4n+3}(\mathrm{SU})$ and $\beta\in\Pi_{4n+3}(\mathrm{SP})$ are generators, then we have $\iota_*(\alpha)=\beta$ (n odd) and 2β (n even) and $\iota^*\gamma^{\mathrm{Sp}}(g_{\mathrm{C}}|V_+)=2\gamma^{\mathrm{SU}}(g_{\mathrm{C}}|V_+)$. Thus,

$$\gamma^{\mathrm{Sp}}(\iota_*\alpha)=\iota^*\gamma^{\mathrm{Sp}}(\alpha)=2\gamma^{\mathrm{SU}}(\alpha)$$

and

$$\int_{S^{4n+3}}\iota^*\gamma^{\mathrm{Sp}}(\alpha)=\pm 2.$$

We obtain

$$\int_{S^{4n+3}}\gamma^{\mathrm{Sp}}_{4n+3}(\beta)=\begin{cases}\pm 1, & \text{for } n \text{ even,}\\ \pm 2, & \text{for } n \text{ odd.}\end{cases}\tag{2.30}$$

The second equation also follows from the explicit construction of $g_{\mathrm{H}}(x)$. The first equation shows that $g_{\mathrm{H}}|V_+$ constructed by the symplectic Clifford algebra generators of $\mathrm{Cl}_\pm(8m)$ is also a generator of $\Pi_{8m-1}(\mathrm{Sp})$.

III. GENERATORS OF UNSTABLE HOMOTOPY GROUPS

If d is the dimension of $\mathbb{F}$ as an $\mathbb{R}$ algebra, then $d=1,2$, or 4 according as $\mathbb{F}$ is $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. The group $O_{n-1}=O_{n-1}(\mathbb{F})$ is the group of $(n-1)\times(n-1)$ matrices over $\mathbb{F}$ which satisfy the relation $\bar{A}^tA=1=A\bar{A}^t$, where the bar indicates conjugation in the appropriate field. Thus, O_{n-1} is the group $O(n-1)$, $U(n-1)$, or $\mathrm{Sp}(n-1)$, according to $\mathbf{F}=\mathbf{R}$, $\mathbf{C}$, or $\mathbb{H}$. We adopt the notation $O_{n-1}=O_{n-1}(\mathbf{F})$ to treat the three cases simultaneously.

The characteristic function T: $S^{dn-2}\to O_{n-1}$ of the principal fiber bundle $O_{n-1}\xrightarrow{\iota}O_n\xrightarrow{\pi}O_n/O_{n-1}=S^{dn-1}$ represents a generator of $\Pi_{dn-2}(O_{n-1})$, the first nonstable homotopy group. The general reference is to Steenrod.[21] For S^{dn-2}

$$=\left\{y=(x_1,x_2,\ldots,x_{n-1},r)\in\mathbb{F}^n\,\middle|\,\sum_{j=1}^{n-1}|x_j|^2+|r|^2=1,\ \text{and } r+\bar{r}=0\right\},$$

T is given by

$$T(y)=(\delta_{p,q}-2x_p(1+r)^{-2}\overline{x_q}).\tag{3.1}$$

One should compare these formulas to those of Steenrod[21] in the cases of $d=1,2$. In the case of $d=4$, one should follow $T(y)$ by the embedding in $\mathrm{SU}(2n-2)$.

For $d=1$ and 2,

$$\det T(y)=-(1+r)^{-2}(1-r)^2,$$

and it is easy to change this matrix to the special orthogonal or special unitary matrix. In the case $d=2$, we can define a characteristic map T': $S^{2n-2}\to\mathrm{SU}(n-1)$ by eliminating the phase factor from T. It is given by $T'(y)=AT(y)A$, where

$$A=\begin{pmatrix}1 & 0\\ 0 & i(1+r)(1-r)^{-1}\end{pmatrix}$$

or

$$T'(y)=\begin{pmatrix} & & -2i\dfrac{x_1x_{n-1}^*}{1-r^2}\\ & \delta_{p,q}-2\dfrac{x_px_q^*}{(1+r)^2} & \vdots\\ & & -2i\dfrac{x_{n-2}x_{n-1}^*}{1-r^2}\\ -2i\dfrac{x_{n-1}x_1^*}{1-r^2} & \cdots\ -2i\dfrac{x_{n-1}x_{n-2}^*}{1-r^2} & 1-2\dfrac{1-|x_{n-1}|^2+r^2}{(1-r)^2}\end{pmatrix}.\tag{3.2}$$

In the case of $d=1$, we can define a characteristic map T': $S^{n-2}\to\mathrm{SO}(n-1)$ by $T'(y)=BT(y)$, where

$$B=\begin{pmatrix}1_{n-2} & 0\\ 0 & -1\end{pmatrix}.$$

Since $\Pi_{2(n-1)}(U(n-1))\cong\mathbb{Z}_{(n-1)!}$ in the unitary case the characteristic map has homotopy class of order $(n-1)!$. Since $\Pi_{4n-2}(\mathrm{Sp}(n-1))\cong\mathbb{Z}_{(n,2)(2n-1)!}$ where $(n,2)$ is the greatest common divisor of n and 2, the characteristic map has order $(n,2)(2n-1)!$ The case of $\mathrm{SO}(n-1)$ is more complicated since $\Pi_{n-2}(\mathrm{SO}(n-1))$ need not be cyclic. However the homotopy class of T' is of order 2 for n even ($n\neq 2,4,8$) and has infinite order for n odd.

In the cases where T' has infinite order, i.e., T': $S^{2n-1}\to\mathrm{SO}(2n)$, the winding number fails to detect the homotopy nontriviality of T'. One easily calculates that $(T'^{-1}dT')^3$ is identically 0. Thus the form (2.2) is zero for $n\geqslant 2$, and for $n=1$ one easily computes $\mathrm{Tr}(T'^{-1}dT')=0$. Thus the winding number integral (2.1) has the value 0.

It is interesting to compute the effect of the covering projection p: $\mathrm{Spin}(n)\to\mathrm{SO}(n)$ on the map g: $S^{n-1}\to\mathrm{Spin}(n)$. The covering map is the restriction of the map defined by $p(Y)Z=YZ\bar{Y}$ for $Y,Z\in\mathrm{Cl}_-(n)$. Note that for $x\in S^{n-1}$, $p(X)$ corresponds to a reflection with respect to x, i.e., $p(X)X=-X$ and $p(X)Y=Y$ for y orthogonal to x. By writing $z=(0,0,\ldots,1)\in S^{n-1}$, g can be written as $g(x)=ZX\in\mathrm{Spin}(n)$. Therefore the representation of $p(g(x))=p(Z)p(X)$ is given by a matrix

$$B_n\cdot(\delta_{pq}-2x_px_q).\tag{3.3}$$

Thus $p \circ g$ is just the characteristic map $T: S^{n-1} \to SO(n)$ (Ref. 15).

ACKNOWLEDGMENT

This work is supported in part by the U.S. Department of Energy under contract No. DE-AC02-86ER-40253 (Y.T.).

APPENDIX A: A HOPF MAP

We give a formula for a Hopf map and its suspensions to complete formulas (2.24) and (2.29).

The Hopf map $h: S^3 \to S^2$ may be given by the formula

$$h(x_1,x_2,x_3,x_4) = (-2(x_1x_4 - x_2x_3), 2(x_1x_3 + x_2x_4), x_4^2 + x_3^2 - x_2^2 - x_1^2). \quad \text{(A1)}$$

The suspension $H^{(n+2)}: S^{n+3} \to S^{n+2}$, which generates the group $\pi_{n+3}(S^{n+2}) \cong \mathbf{Z}_2$, may be given as follows. If $x = (x_1,...,x_{n+4})$ and $c = (x_{n+1}^2 + \cdots + x_{n+4}^2)^{1/2}$, then

$$\begin{aligned} h^{(n+2)}(x) &= (x_1,...,x_n, -2(x_{n+1}x_{n+4} - x_{n+2}x_{n+3})/c, \\ &\quad 2(x_{n+1}x_{n+3} + x_{n+2}x_{n+4})/c, \\ &\quad (x_{n+4}^2 + x_{n+3}^2 - x_{n+2}^2 - x_{n+1}^2)/c), \\ &\quad \text{if } c \neq 0, \\ &= (x_1,...,x_n,0,0,0), \quad \text{if } c = 0. \end{aligned} \quad \text{(A2)}$$

APPENDIX B: THE WINDING NUMBER

We outline a differential geometric proof that the value of the integral (2.1) must be an integer.

Consider a principle fibre bundle (P, S^{2n+2}, π, G) with the connection one-form ω on P. We define the gauge potential $A_\pm = s_\pm^* \omega$ where $s_\pm$ are sections associated with the trivialization $(H_\pm, \phi_\pm)$ over the (slightly enlarged) hemi spheres of S^{2n+2} as the local coverings, and with $S^{2n+1} \subset H_+ \cap H_-$ a deformation retract. Define a local curvature two-form $F_\pm = dA_\pm + [A_\pm, A_\pm]$. We take the transformation function as f. Then the forms $A_\pm$ and $F_\pm$ are related by $A_+ = f^{-1}A_- f + f^{-1}\,df$ and $F_+ = f^{-1}F_- f$. Note that there exists a $2n+$ one-form Ω_{2n+1}^0 such that

$$(i/2\pi)^{n+1}\mathrm{Tr}(F)^{n+1} = d\Omega_{2n+1}^0(A,F),$$

since $d\,\mathrm{Tr}\,F^{n+1} = 0$. We do not bother to write down the explicit form for Ω_{2n+1}^0, which is well known.[1] However, we have the following relation on $H_+ \cap H_- \supset S^{2n+1}$:

$$\gamma_{2n+1}^G(f) = \Omega_{2n+1}^0(A_+,F_+) - \Omega_{2n+1}^0(A_-,F_-) + d\alpha(A_-,F_-,f).$$

Therefore we have

$$\begin{aligned} \int_{S^{2n+1}} \gamma_{2n+1}^G(f) &= \int_{S^{2n+1}} [\Omega_{2n+1}^0(A_-^f, F_-^f) - \Omega_{2n+1}^0(A_-,F_-)] \\ &= \int_{\partial H_+} \Omega_{2n+1}^0(A_+,F_+) - \int_{\partial H_-} \Omega_{2n+1}^0(A_-,F_-) \\ &= \int_{H_+} d\Omega_{2n+1}^0(A_+,F_+) - \int_{H_-} d\Omega_{2n+1}^0(A_-,F_-) \\ &= \left(\frac{i}{2\pi}\right)^{n+1}\left[\int_{H_+} \mathrm{Tr}(F_+)^{n+1} - \int_{H_-} \mathrm{Tr}(F_-)^{n+1}\right] \\ &= \left(\frac{i}{2\pi}\right)^{n+1}\int_{S^{2n+2}} \mathrm{Tr}\,F^{n+1} = \int_{S^{2n+2}} \mathrm{ch}(F). \end{aligned}$$

Since the integral of the Chern form $\mathrm{ch}(F)$ is an integer, the value of (2.1) is an integer.

[1]See, for example, B. Zumino, in *Current Algebra and Anomalies*, edited by S. B. Treiman, R. Jackiw, B. Zumino, and E. Witten (Princeton U.P., Princeton, 1985); T. Eguchi, P. B. Gilkey, and A. J. Hansen, Phys. Rep. **66**, 213 (1980).
[2]A. Jaffe and C. Taubes, *Vortices and Monopoles* (Birkhäuser, Boston, 1980).
[3]E. Witten, Nucl. Phys. B **223**, 422 and 433 (1983); Phys. Lett. B **117**, 324 (1982); S. Elitzer and V. P. Nair, Nucl. Phys. B **243**, 205 (1984).
[4]S. Okubo, H. Zhang, Y. Tosa, and R. E. Marshak, Phys. Rev. D **37**, 1655 (1988).
[5]H. Zhang, S. Okubo, and Y. Tosa, Phys. Rev. D **37**, 2946 (1988).
[6]H. Zhang and S. Okubo, Phys. Rev. D **38**, 1880 (1988).
[7]S. Okubo and H. Zhang, in *Perspectives on Particle Physics*, edited by S. Matsuda *et al.* (World Scientific, Singapore, 1989).
[8]A. T. Lundell and Y. Tosa, J. Math. Phys. **29**, 1795 (1988).
[9]S. Okubo and Y. Tosa, Phys. Rev. D **40**, 1925 (1989).
[10]Y. Tosa, Phys. Rev. D **39**, 1648 (1989).
[11]Y. Tosa, Phys. Rev. D **40**, 1934 (1989).
[12]R. Bott and R. Seeley, Commun. Math. Phys. **62**, 235 (1978).
[13]S. Okubo and J. Patera, J. Math. Phys. **24**, 2772 (1983); **25**, 219 (1984); Phys. Rev. D **31**, 2669 (1985).
[14]E. Witten, Nucl. Phys. B **223**, 433 (1983).
[15]M. F. Atiyah, R. Bott, and A. Shapiro, Topology **3**, 3 (1964); I. R. Porteous, *Topological Geometry* (Van Nostrand Reinhold, London, 1969); M. Postnikov, *Lie Groups and Lie Algebras* (Mir, Moscow, 1986); D. Husemoller, *Fibre Bundles*, 2nd ed. (Springer, New York, 1975).
[16]M. Karoubi and C. Leruste, *Algebraic Topology via Differential Geometry* (London Mathematical Society Lecture Note Series Vol. 99) (Cambridge U.P., Cambridge, 1987).
[17]A. T. Lundell, Topology **8**, 209 (1969).
[18]R. Bott, Bull. Soc. Math. Fr. **87**, 293 (1959).
[19]E. Dyer and R. Lashof, Ann. Mat. Pura. Appl. **54**, 231 (1961).
[20]M. A. Kervaire, Ill. J. Math. **4**, 161 (1960).
[21]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, 1951).

Addendum to "The equivariant inverse problem in gauge field theories and the uniqueness of the Yang–Mills equations" [J. Math. Phys. 30, 2382 (1989)]

R. J. Noriega
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

C. G. Schifini
Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina and CONICET, Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

(Received 19 July 1989; accepted for publication 17 January 1990)

The main theorem of the previous paper is extended to the case where $L = L(g_{ij};A_i^\alpha;A_{i,j}^\alpha)$. References and notations are the same.

Let $T = T(g_{ij};F_{ij}^\alpha) = L(g_{ij};0;\ -\frac{1}{2}F_{ij}^\alpha)$. Since L^{hk} is gauge invariant, by the replacement theorem[1] $L^{hk} = T^{hk}$. Then T^{hk} is gauge invariant and T has the form required in Theorem 1. Hence, $T = L_1 + L_2 + K$. Since $(L - T)^{hk} = 0$, it follows that $L = L_1 + L_2 + S$, where $S = S(A_i^\alpha;A_{i,j}^\alpha)$. Since $E_\alpha^i(L)$ is a gauge tensorial density, $L_{\alpha\ \beta}^{i,j;h,k} + L_{\alpha\ \beta}^{i,k;h,j}$ is a gauge tensorial density. Now, $\partial L_1/\partial A_{i,j}^\alpha = -2\partial L_2/\partial F_{ij}^\alpha$. Then

$$4L_{1\alpha\beta}^{ij;hk} + 4L_{1\alpha\ \beta}^{ik;hj} + 4L_{2\alpha\beta}^{ij;hk} + 4L_{2\alpha\ \beta}^{ik;hj} + S_{\alpha\ \beta}^{i,j;h,k} + S_{\alpha\ \beta}^{i,k;h,j}$$

is a gauge tensorial density. This is also true for the sum of the first two terms, and the sum of the following two terms is null. Thus $S_{\alpha\ \beta}^{i,j;h,k} + S_{\alpha\ \beta}^{i,k;h,j}$ is a tensorial density. As we proved in the original paper, it has the form $a_{\alpha\beta}\epsilon^{ijhk}$ and it is symmetric in k,j. Thus it is null and $S_{\alpha\ \beta}^{i,j;h,k} = -S_{\alpha\ \beta}^{i,k;h,j}$, from where it follows that

$$E_\alpha^i(S) = S_\alpha^i - S_{\alpha\ \beta}^{i,j;h}A_{h,j}^\beta.$$

We deduce easily that

$$E_\alpha^i(S)(0;-\tfrac{1}{2}F) = -E_\alpha^i(S)(0;-\tfrac{1}{2}F)$$

(making $\overline{x}^i = -x^i$). Then, by the replacement theorem,[1]

$$E_\alpha^i(L) = E_\alpha^i(L_1)(g;0;-\tfrac{1}{2}F;-\tfrac{2}{3}F').$$

Since this equation is tensorial, it is valid for all coordinate systems. Then

$$E_\alpha^i(L_2 + S)(g;0;-\tfrac{1}{2}F;-\tfrac{2}{3}F') = 0,$$

and so

$$5g^{lm}g_{lm,s}l_{\gamma\alpha}\epsilon^{sihk}F_{hk}^\gamma + E_\alpha^i(S)(0;-\tfrac{1}{2}F) = 0.$$

It follows easily that $l_{\alpha\beta} = 0$, i.e., $L_2 = 0$. Thus $E_\alpha^i(S)$ is a tensorial density, and so the same is true for $E_\alpha^i(S)_{\beta\ \alpha}^{h,k;r,s}$. Then it is null, which means that $E_\alpha^i(S)$ is a polynomial of degree $\leqslant 1$ in $A_{i,j}^\alpha$. Then

$$E_\alpha^i(S) = d_{\alpha\beta\gamma}\epsilon^{ijhk}A_{j,h}^\gamma A_k^\beta + C_{\alpha\beta\gamma\theta}\epsilon^{ijhk}A_j^\beta A_h^\gamma A_k^\theta,$$

where $c_{\alpha\beta\gamma\theta}$ is skew symmetric in β, γ, θ. If $B_\alpha^i = E_\alpha^i(S)$, then B_α^i must satisfy

$$B_{\alpha\beta}^{i;j,h} = -B_{\beta\alpha}^{j;i,h},$$

$$B_{\alpha\beta}^{i;j} = B_{\beta\alpha}^{j;i} + \frac{\partial}{\partial x^h}(B_{\alpha\beta}^{i;j,h})$$

(see Ref. 2). We deduce

$$d_{\alpha\beta\gamma} + d_{\beta\alpha\gamma} + d_{\alpha\gamma\beta} = 0, \quad c_{\alpha\beta\gamma\theta} + c_{\beta\alpha\gamma\theta} = 0.$$

If

$$S_1 = \tfrac{1}{3}d_{\alpha\beta\gamma}\epsilon^{ijhk}A_{j,h}^\gamma A_k^\beta A_i^\alpha + \tfrac{1}{4}c_{\alpha\beta\gamma\theta}\epsilon^{ijhk}A_i^\alpha A_j^\beta A_h^\gamma A_k^\theta,$$

we have $E_\alpha^i(S) = E_\alpha^i(S_1)$, and S_1 is a scalar density. Then

$$E_\alpha^i(L) = E_\alpha^i(L_1 + S_1),$$

and besides

$$E^{ij}(L) = E^{ij}(L_1 + S_1).$$

Being that $L_1 + S_1$ is a scalar density, we are in the same situation as the one studied in the original paper. Now, the theorem follows for $L = L(g_{ij};A_i^\alpha;A_{i,j}^\alpha)$.

[1] G. W. Horndeski, Utilitas Math. **19**, 215 (1981).
[2] I. M. Anderson and T. Duchamp, Am. J. Math. **102**, 781 (1980).

0022-2488/90/061503-01$03.00

Adjoint division algebras and SU(3)

Geoffrey Dixon
Department of Physics, Brandeis University, Waltham, Massachusetts and Department of Mathematics, University of Massachusetts, Boston, Massachusetts 02125

(Received 11 October 1989; accepted for publication 31 January 1990)

Adjoint algebras are constructed for the division algebras and their tensor products, and for the full tensor product, adjoint idempotents that project lepton and quark color vectors are also constructed.

I. INTRODUCTION

In previous publications[1-5] I derived all the important features of the standard model starting from the tensor product of the three hypercomplex division algebras. Here I look more closely at some of the distinguishing properties of these algebras and apply them to the construction of SU(3) color projection operators.

II. THEORY

It is generally known that the complex algebra **C** is commutative and associative, the quaternions **Q** are associative but not commutative, and the octonions **O** are neither. Each property lost, however, is balanced by a concommitant rise in multiplicative potency.

The three hypercomplex division algebras are, respectively, 2, 4, and 8 dimensional. The complete algebra of actions on a real n-dimensional vector space is $\mathbf{R}(n)$, the algebra of real $n\times n$ matrices. Hence, the action algebras of **C**, **Q**, and **O** are isomorphic to $\mathbf{R}(2)$, 4 dimensional, $\mathbf{R}(4)$, 16 dimensional, and $\mathbf{R}(8)$, 64 dimensional.

Let $\mathbf{K}_L$, $\mathbf{K}_R$, and $\mathbf{K}_A$ be the algebras of left-, right-, and two-sided actions of the algebra **K** on itself, $\mathbf{K}=\mathbf{C}$, **Q**, or **O**. Isomorphisms for these nine adjoint algebras are listed below:

$$\mathbf{C}_L=\mathbf{C}_R=\mathbf{C}_A=\mathbf{C},$$

$$\mathbf{Q}_L=\mathbf{Q}_R=\mathbf{Q},\quad \mathbf{Q}_A=\mathbf{R}(4),$$

$$\mathbf{O}_L=\mathbf{O}_R=\mathbf{O}_A=\mathbf{R}(8).$$

The complex isomorphisms are easiest to explain. Since **C** is commutative, $\mathbf{C}_L=\mathbf{C}_R=\mathbf{C}_A$, and since it is associative these adjoint algebras can be no more complicated than **C** itself. Hence, all fall short of the action algebra $\mathbf{R}(2)$. We can complete it by augmenting the conventional basis for **C**, 1, and i, with actions 1_* and i_*, defined on x in **C** by $1_*x=x^*$ and $i_*x=ix^*$ (note that i, 1_*, and i_* anticommute, x^* the complex conjugate of x), but we have to go outside of **C** to do so.

Because **Q** is associative, $\mathbf{Q}_L=\mathbf{Q}_R=\mathbf{Q}$, and because **Q** is noncommutative, $\mathbf{Q}_A$ is more complicated than the 2 one-sided adjoint algebras. A complete basis for $\mathbf{Q}_A$ consists of the actions

$$x\to x,\ q_ix,\ xq_i,\ q_ixq_j, \tag{1}$$

$i,j=1,2,3$, and the q_i are a conventional basis for the hypercomplex part of **Q**. Hence, $\mathbf{Q}_A$ is $1+3+3+9=16$ dimensional and must be isomorphic to $\mathbf{R}(4)$.

Let e_a, $a=1,...,7$, be the hypercomplex units of **O** [my multiplication table for the octonions is the conventional one (although not the most natural), with the e_i, $i=1,2,3$, generating a quaternionic subalgebra, and $e_4=e_1e_7$, $e_5=e_2e_7$, and $e_6=e_3e_7$ (see Ref. 6)]. One might suppose that the noncommutativity of **O** would make $\mathbf{O}_A$ bigger than $\mathbf{O}_L$ and $\mathbf{O}_R$, but the nonassociativity of **O** maximizes the potency of one-sided multiplication. As a consequence, any left adjoint action of **O** on **O** can be expressed as a right action, and vice versa. Define the actions

$$e_{ab\cdots c}x=e_a(e_b(\cdots(e_cx)\cdots)), \tag{2}$$

$$xe_{a\cdots bc}=((\cdots(xe_a)\cdots)e_b)e_c. \tag{3}$$

Then, for example, we can express the following left actions as right actions:

$$e_7x=\tfrac{1}{2}x(e_{41}+e_{52}+e_{63}-e_7), \tag{4}$$

$$e_{41}x=\tfrac{1}{2}x(-e_{41}+e_{52}+e_{63}+e_7) \tag{5}$$

(note that $e_4e_1=e_5e_2=e_6e_3=e_7$, which sets the pattern). Hence, $\mathbf{O}_L=\mathbf{O}_R=\mathbf{O}_A$, all isomorphic to $\mathbf{R}(8)$. In particular, a complete basis for $\mathbf{O}_L$ consists of the actions

$$x\to x,\ e_ax,\ e_{ab}x,\ e_{abc}x, \tag{6}$$

which makes this algebra $1+7+21+35=64$ dimensional, as expected. Actions of higher order than those in (6) can be shown to reduce to one of those forms. Note also that the embedding of parentheses in (2) and (3) ensures that these adjoint algebras are associative [or else the isomorphism to $\mathbf{R}(8)$ would be impossible].

Under commutation, certain subsets of $\mathbf{C}_A$, $\mathbf{Q}_A$, and $\mathbf{O}_A$ are isomorphic to the Lie algebras so(2), so(4), and so(8), respectively.

The generator of so(2) = u(1) is the imaginary unit i.

The generators of so(4) are the adjoint operations

$$x\to q_ix,\ xq_i. \tag{7}$$

The operations $x\to q_ix$ and $x\to xq_i$ generate separate and commuting copies of su(2) = so(3). So so(4) = su(2) × su(2). We can define Hermitian conjugates of these operations using the order reversing $x\to x^\dagger$ operation on **Q** (and **O**) which changes the signs of the q_i (and e_a). For any operation on x in $\mathbf{Q}_A$ (or $\mathbf{O}_A$), take $x^\dagger$, reverse

 0022-2488/90/061504-02$03.00

left and right adjoint operations, then take the dagger conjugate of the result. So $q_i x \rightarrow (x^\dagger q_i)^\dagger = -q_i x$ (anti-Hermitian, same for the right action), and $q_i x q_j \rightarrow (q_j x^\dagger q_i)^\dagger = q_i x q_j$ (Hermitian).The same technique applied to the generators (6) of $\mathbf{O}_L$ yields: $e_a x \rightarrow (x^\dagger e_a)^\dagger = -e_a x$ (anti-Hermitian), $e_{ab} x \rightarrow (x^\dagger e_{ab})^\dagger = e_{ba} x = -e_{ab} x$ (anti-Hermitian), and $e_{abc} x \rightarrow (x^\dagger e_{abc})^\dagger = -e_{cba} x = e_{abc} x$ (Hermitian). The anti-Hermitian operations $x \rightarrow e_a x, e_{ab} x$ generate so(8). Different combinations of these close under commutation and generate other Lie algebras. In particular, so(7), spinor representation on **O**,

$$x \rightarrow e_{ab} x;$$

so(7), fundamental representation on **O**, identity invariant,

$$x \rightarrow (e_a - e_{bc})x\colon e_a = e_b e_c;$$

G_2, Lie algebra of the automorphism group of **O**,

$$x \rightarrow (e_{ab} - e_{cd})x\colon e_a e_b = e_c e_d;$$

su(4) = so(6),

$$x \rightarrow (e_a - e_{bc})x\colon e_a = e_b e_c, \quad a,b,c \neq 7;$$

su(3), subalgebra of G_2 leaving e_7 invariant,

$$x \rightarrow (e_{ab} - e_{cd})x\colon e_a e_b = e_c e_d, \quad a,b,c,d \neq 7.$$

This last Lie algebra, su(3), generates the color group in my derivation of the Standard model.

The tensor algebras $\mathbf{P} = \mathbf{C} \otimes \mathbf{Q}$ and $\mathbf{H} = \mathbf{C} \otimes \mathbf{Q} \otimes \mathbf{O}$ derive their commutivity, associativity, and adjoint properties from their constituent algebras; **P** is isomorphic to the Pauli algebra, **C**(2). It is associative but noncomutative; **H** is not only noncommutative and nonassociative, it is also nonalternative [$a(ab) \neq a^2 b$ necessarily]. Adjoint algebra isomorphisms are listed below:

$$\mathbf{P}_L = \mathbf{P}_R = \mathbf{C}(2), \quad \mathbf{P}_A = \mathbf{C}(4),$$

$$\mathbf{H}_L = \mathbf{H}_R = \mathbf{C}(16), \quad \mathbf{H}_A = \mathbf{C}(32).$$

I note in passing that Clifford (Dirac) algebras of the two presently most important space-times in theoretical physics [(1,3) and (1,9)] are constructible in the same way from the algebras $\mathbf{P}_L$ and $\mathbf{H}_L$. I shall investigate this in more detail in a future publication.

With respect to SU(3) **H** transforms as $1 \oplus 3 \oplus \bar{3} \oplus \bar{1}$. I employ the idempotents $\rho_\pm = \frac{1}{2}(1 \pm ie_7)$ to project from **H** these various multiplets. In particular I make the assignments

$$\rho_+ \mathbf{H} \rho_+ \rightarrow \mathbf{1},\ \rho_+ \mathbf{H} \rho_- \rightarrow \mathbf{3},\ \rho_- \mathbf{H} \rho_+ \rightarrow \bar{\mathbf{3}},\ \rho_- \mathbf{H} \rho_- \rightarrow \bar{\mathbf{1}}. \tag{8}$$

For example, let $x = x^a e_a, \quad a = 0,1,\ldots,7, \quad e_0 = 1$. Then, since $e_7 \rho_\pm = \mp i \rho_\pm$,

$$\rho_+ x \rho_+ = (x^0 - ix^7)\rho_+,$$

$$\rho_+ x \rho_- = [(x^1 + ix^4)e_1 + (x^2 + ix^5)e_2 + (x^3 + ix^6)e_3]\rho_-. \tag{9}$$

If the x^a are real the antisinglet and antitriplet are the conjugates of the singlet and triplet, but if complex, which in our case they are, the antimultiplets are new. I interpreted this to imply family replication.

Because the $\rho_\pm$ are independent of **Q**, left multiplication of $\rho_\pm$ can be expressed as a right action. For example, for X in **H**,

$$\rho_+ X \rho_+ = \tfrac{1}{8} X(1 + ie_{41})(1 + ie_{52})(1 + ie_{63}). \tag{10}$$

Define

$$\rho_{\pm 1} = \frac{1 \pm ie_{41}}{2},\ \rho_{\pm 2} = \frac{1 \pm ie_{52}}{2},\ \rho_{\pm 3} = \frac{1 \pm ie_{63}}{2}. \tag{11}$$

We can use these idempotents to completely decompose **H** with respect to SU(3). For example, we make the family assignments:

$$\mathbf{H}\rho_{+1}\rho_{+2}\rho_{+3} \rightarrow \text{leptons},$$

$$\mathbf{H}\rho_{-1}\rho_{+2}\rho_{+3} \rightarrow \text{red quarks},$$

$$\mathbf{H}\rho_{+1}\rho_{-2}\rho_{+3} \rightarrow \text{green quarks},$$

$$\mathbf{H}\rho_{+1}\rho_{+2}\rho_{-3} \rightarrow \text{blue quarks}.$$

The antifamily is obtained by changing the signs of all subscripts. These projections can be expressed differently, and of course the color assignments are arbitrary. With respect to **O** the leptons are linear in 1 and e_7, red quarks in e_1 and e_4, green quarks in e_2 and e_5, and blue quarks in e_3 and e_6.

III. CONCLUSION

I wish to emphasize that these adjoint algebras are secondary to the algebra **H** itself. The elements $\rho_{\pm i}$, $i = 1,2,3$, decompose SU(3) multiplets to the vector level, but they are not elements of **H**. Using idempotents in **H**, namely the $\rho_\pm$, the best one can do is decompose **H** to the SU(3) multiplet level, and in my derivation of the Standard model this was responsible for the exactness of that symmetry.

[1] G. M. Dixon, "Derivation of the Standard Model" (to be published).
[2] G. M. Dixon, J. Phys. G **12**, 561 (1986).
[3] G. M. Dixon, Phys. Lett. B **152**, 343 (1985).
[4] G. M. Dixon, Phys. Rev. D. **29**, 1276 (1984).
[5] G. M. Dixon, Phys. Rev. D. **28**, 833 (1983).
[6] I. R. Porteous, *Topological Geometry* (Van Nostrand–Reinhold, London, 1969).

Gauge field systems on $\mathbb{CP}^n$

Zhong-Qi Ma[a)]
CCAST (World Laboratory), P. O. Box 8730, Beijing, People's Republic of China and School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland

D. H. Tchrakian
School of Theoretical Physics, Dublin Institute of Advanced Studies, 10 Burlington Road, Dublin 4, Ireland and Department of Mathematical Physics, St. Patrick's College, Maynooth, Ireland

(Received 28 June 1989; accepted for publication 7 January 1990)

$SU(n)\times U(1)$ gauge fields on $\mathbb{CP}^n$ are constructed in a gauge where the connection is expressed in $so(n-1)$. The (anti)self-duality relations satisfied by these fields are given, and the $2n$-dimensional gauge field systems whose Bogomolnyi bounds are saturated by these are discussed. Other gauge field systems, not endowed with topological lower bounds, are also discussed. The analogy between all these gauge field systems on $\mathbb{CP}^n$ and S^{2n} is highlighted.

I. INTRODUCTION

The study of gauge field systems in higher dimensions was stimulated by the discovery of topologically nontrivial solutions to the Yang–Mills–Higgs (YMG) system by 't Hooft[1] and by Polyakov,[2] and to the Yang–Mills (YM) system by Belavin *et al.*[3] These are the celebrated monopole and instanton solutions in 3 and 4 dimensions, respectively. In a first study,[4] instantons and monopoles were generalized to all even and odd dimensions, respectively, and subsequently explicit spherical symmetric instanton solutions to the systems in $4p$ dimensions were discovered,[5,6] as well as axially symmetric solutions.[7]

Although the monopole[1,2] system features the Higgs field, the study of the pure YM field system affords a more direct access to the topological properties of gauge field systems. Accordingly, we shall choose to pursue the generalization of the *pure* YM system to higher dimensions, having in mind that ultimately the Higgs field must be reintroduced for physical applications.[8–10] Since this can be done systematically[8–10] by the use of coset-space dimensional reduction,[11] we shall henceforth restrict our considerations only to the generalizations of the *pure* YM system to higher dimensions. These dimensions must of necessity be *even*, since our guiding principle will be the occurrence of topologically nontrivial field configurations and the latter are characterized by their Chern–Pontryagin (C–P) charges, which for systems consisting of gauge fields only are defined on *even* dimensions.

The generalization of the YM system to higher dimensions has developed along two[4,12,13] distinct lines. The first[4] involves the definition of gauge field systems in all even dimensions $2(p+q)$, whose Lagrangian densities are positive definite, given by

$$L_{\rm GYM}(p,q) = e\,{\rm tr}[F(2p)^2 + (i\kappa^2)^{2(q-p)}(2p!/2q!)F(2q)^2], \quad (1.1)$$

$$F(2n) = F\wedge F\wedge\cdots\wedge F, \quad n \text{ times}. \quad (1.1')$$

Here, κ is a dimensional constant and e is the determinant of the Vielbein of the $2(p+q)$-dimensional manifold. The system (1.1) was chosen to that its Euler–Lagrange equations are solved by the duality equation

$$F(2p)_{\mu_1\cdots\mu_{2p}} = (e/2q!)(i\kappa^2)^{(q-p)} \times\epsilon_{\mu_1\cdots\mu_{2p}\nu_1\cdots\nu_{2q}}F(2q)^{\nu_1\cdots\nu_{2q}}, \quad (1.2)$$

and the Bianchi identities.

We call the systems (1.1) generalized YM (GYM) systems because under dimensional reduction they yield residual systems that are dominated[8] by the YMH system at low energies. [The corresponding reduction of (1.2) then yields the Bogomolnyi equations of the residual system in lower dimensions.] Note also that by virtue of (1.1′), only terms quadratic in $\partial_\mu A_\nu$ (any derivative of the connection A_ν) appears in the action density (1.1). So, as in YM theory, there is no difficulty in defining a canonical momentum.

A special feature of GYM systems is that when $p+q$, the dimensional constant κ does not appear in (1.1) and $L_{\rm GYM}(p,p)$ is then conformally invariant in $4p$ dimensions. With $p\neq q$, $L_{\rm GYM}(p,q)$ are *not* conformally invariant.

Another, even more important feature of GYM systems is that the Lagrange density (1.1) is minimized *absolutely* by the self-duality Eq. (1.2), whose solutions are therefore *stable* instantons of the system. Such solutions of course exist only for the appropriate gauge groups with the requisite topological properties with respect to the $2(p+q)$-dimensional manifold. For these field configurations, the action attains the Bogomolnyi bound, which turns out to be the $(p+q)$th Chern–Pontryagin (C–P) charge.

Both of these last two distinguishing features of the GYM system depend critically on the dimensionality of the manifold being $2(p+q)$. This brings us to the second line of generalization, due to Saçlioğlu[12] and Fujii[13] of the YM system, which we refer to as the extended YM(EYM) systems following the nomenclature of Ref. 13.

Here we depart slightly from the definition of the EYM systems given by the authors of Refs. 12 and 13, the reason for which will become clear in Sec. IV. The original definition[12,13] of the EYM system in $2N$ dimensions is

$$L^{(N)} = e\,{\rm tr}(\Gamma_{\mu\nu}\otimes F_{\mu\nu})^n, \quad (1.3)$$

[a)] On leave from IHEP, Academia Sinica, P. O. Box 918(4), Beijing, People's Republic of China.

 0022-2488/90/061506-07$03.00

where $\Gamma_{\mu\nu} = -\frac{1}{4}[\Gamma_\mu,\Gamma_\nu]$ are the SO(2N) matrices given in terms of the Γ matrices in $2N$ dimensions.

Our departure from (1.3) consists of the following: After computing the trace over the SO(2N) matrices, we formally state that the remaining polynomial in $F_{\mu\nu}$ be the EYM Lagrangian density on a manifold of arbitrary dimensionality. For example, for $N=3$ we define

$$L^{(3)} \approx e \operatorname{tr} F_{\mu\nu} F_{\nu\rho} F_{\rho\mu} \tag{1.3'}$$

to be the EYM system in any dimensions, not necessarily equal to $2N$. We note that for N odd, $L_{(N)}$ is not positive definite.

Perhaps the most important feature of the EYM systems $L^{(n)}$ is that they are conformally invariant in $2n$ dimensions, for odd or even n. This differs from the GYM case, which for odd n (i.e., p or q odd) are *not* conformally invariant. For even $n=2p$, and in particular $p=q$, the GYM system in $4p$ dimensions is included in the corresponding EYM system, and the two Lagrangians differ only through terms that feature higher than the quadratic power of the "velocity" field $\partial_\mu A_\nu$. These important features were pointed out by Saçlioğlu,[12] who also observed that for odd n as well, certain duality relations were satisfied by some polynomials of the curvature field strength. The latter however are not Bogomolnyi conditions like (1.2) and do not therefore confer stability on the corresponding solutions. Nevertheless we will find it useful to employ these, and similar duality relations for even n, in Sec. IV below.

While it is not our purpose in this paper to discuss the applications of the above defined systems, we would nevertheless like to motivate our work with physically relevant criteria. Firstly, we note that in any such problem, e.g., relating to quantum fluctuations, stability would be a desirable feature of any solution. This privileges the GYM systems as these are endowed with Bogomolnyi bounds. Secondly, since these systems pertain to higher dimensions mostly, they would usually be defined on an (extra-dimensional) compact manifold, e.g., in some compactification[14] scheme.

It is against this background that the present work is carried out. Our primary task is to find solutions of the Bogomolnyi (self-duality) Eqs. (1.2) on compact manifolds. More specifically, the most interesting class of such manifolds occurring in compactification schemes are the coset spaces, and in particular the symmetric spaces. In fact, it is already known from the work of Ref. 15 that (1.2) are satisfied on the spheres $S^{2(p+q)}$. In Ref. 15 we constructed spin connection gauge fields satisfying (1.2), on particular $2(p+q)$-dimensional (double self-dual) generalized gravitational backgrounds. The double self-duality of the (generalized) gravitational background was checked explicitly only from $S^{2(p+q)}$. In the present paper, we take a more direct approach to the same problem, with no reference to the gravitational dynamics. We construct the (symmetric) field on the corresponding (symmetric) space, using the formalism of Schwarz,[16] Yang,[17] and Gu.[18] Having constructed the symmetric field, the Eqs. (1.2) can then be checked. In practice we restrict ourselves to space $\mathbb{CP}^{p+q}$, as the next example after the symmetric space $S^{2(p+q)}$ already considered.[15] This is a sufficiently nontrivial example, so that it can serve as a prototype for this investigation to be carried out for all the other symmetric spaces. Some of our results, in particular that pertaining to the GYM systems with $p=1$ and $q=(n-1)$ were obtained by Bais and Batenburg,[19] but not in the context of GYM dynamics.

The rest of this paper is planned as follows. In Sec. II, we give our construction of gauge fields on $\mathbb{CP}^n = \mathrm{SU}(n+1)/\mathrm{SU}(n)\times\mathrm{U}(1)$, with gauge group $\mathrm{SU}(n)\times\mathrm{U}(1)$. Our construction, which is given in a gauge where the gauge connection belongs to $\mathrm{SU}(n+1)$, is equivalent to the usual, one[20] used in Ref. 19, as is clear from the fact that the expressions for the (gauge covariant) fieldstrengths agree in both formulations. We hope that our construction is interesting in its own right, since it can be very naturally extended to any symmetric space. In Sec. III we present our main result, which is the duality relations satisfied by the $2p$-form curvature strengths on $\mathbb{CP}^{p+q}$. Having achieved this major task of classifying the self-duality equations solving the appropriate GYM system, we proceed in Sec. IV to the more general consideration of the Euler–Lagrange equation of the EYM systems on $\mathbb{CP}^{p+q}$. These solutions are not guaranteed to be stable, as they are not minimized absolutely by Bogomolnyi equations. Our conclusions of Sec. IV are based on the special examples where $(p+q)=n=2$, 3, and 4 only. In both Secs. III and IV, we illustrate our procedures by employing the (already analyzed) cases of $S^{2(p+q)}$, and highlight the analogy between symmetric gauge fields on S^{2n} and $\mathbb{CP}^n$ for systems with GYM and EYM dynamics, respectively.

II. GAUGE FIELDS ON $\mathbb{CP}^n$

In this section we will construct the symmetric gauge connection and curvature field strength on $\mathbb{CP}^n$. We first write down the symmetry equations to be solved in subsection II A, and then compute the effective forms of $\mathrm{SU}(n+1)$ and $\mathrm{SU}(n)\times\mathrm{U}(1)$ acting on the coordinates of $\mathbb{CP}^n$ in subsection II B. Finally, in subsection II C we give the values of A^a and F^{ab}, and then in Sec. II D we show that F^{ab} satisfies the Yang–Mills field equation.

Before writing down the symmetry equations we record our notation. We denote the coordinates of the $2n$-dimensional symmetric coset space $\mathbb{CP}^n = \mathrm{SU}(n+1)/\mathrm{SU}(n)\times\mathrm{U}(1)$ by x^α, $\alpha = a,\bar{a}$ and with $\overline{x^a} = x^{\bar{a}}$ and $a=1,\ldots,n$. These coordinates are defined in terms of $(n+1)$ complex numbers $z^i, i=1,\ldots,n+1$ as follows:

$$x^a = z^a/z^{n+1}, \quad \text{for } z^{n+1} \neq 0, \tag{2.1}$$

and it is clear that this is one of $(n+1)$ patches.

Using the definitions given in Ref. 21 for $ds^2 = \Sigma_i |dz^i|^2$, we have the metric on $\mathbb{CP}^n$

$$g_{\alpha\beta} = \begin{Bmatrix} g_{a\bar{b}} = -x^{\bar{a}}x^b/\Omega^4 + \delta ab/\Omega^2, \\ g_{\bar{a}\bar{b}} = g_{ab} = 0, \end{Bmatrix} \tag{2.2a}$$

$$\Omega^2 = 1 + \sum_a |x^a|^2 \equiv 1 + r^2. \tag{2.2b}$$

We note that the construction given in this paper is strictly a local one, and that we make no reference to the global definition of the fields.

A. Symmetry equation

The $SU(n+1)$ symmetric gauge connection $A^a(x)$ satisfies the symmetry equation

$$A^a(uz) = \hat{u}^a_b g(u) A^b(z) g^{-1}(u) + i g(u)\partial^a g^{-1}(u), \tag{2.3}$$

where we take $g(u)$ to be an element of $SU(n+1)$ in the fundamental representation, and $\hat{u}^a_b$ is the effective form of $SU(n+1)$ acting on the coordinates x^a of $\mathbb{CP}^n$.

Following the formulations of Schwartz[16] and Yang[17] and Gu[18] for symmetric gauge fields, we specialize (2.3) to the symmetric case without the affine term

$$A^a(z) \equiv A^a(u_z\mathring{z}) = (\hat{u}_z)^a_b(\mathring{z}) g(u_z) A^b(\mathring{z}) g^{-1}(u_z). \tag{2.4}$$

Furthermore, at the fixed point $z=\mathring{z}$, choosing $u = h \in SU(n)\times U(1)$ to be the stability subgroup for $\mathring{z}$, i.e., $h\mathring{z}=\mathring{z}$, (2.4) becomes the purely algebraic equation

$$A^a(\mathring{z}) = \hat{h}^a_b g(h) A^b(\mathring{z}) g^{-1}(h), \tag{2.5}$$

where again $\hat{h}$ is the effective form of $h\in SU(n)\times U(1)$. In our computations, we will choose this $\mathring{z}$ to correspond to the fixed point with $r=0$ in (2.2b).

Equation (2.4) is the symmetry constraint, whose solution is the symmetric gauge connection we are seeking. We first calculate $A^a(\mathring{z})$ by solving the simpler algebraic equation (2.5) and then we have $A^a(z)$ from (2.4). It is this latter form of the connection that we need to calculate the field strength $F^{ab}(z)$.

B. Effective forms $\hat{u}_z(\mathring{z})$ and $\hat{h}$

Expressing the coordinates x^a via $z^a = (x^a, 1)$ the elements u in (2.3) act as

$$z'^i = u^i_j z^j, \tag{2.6}$$

where $(z')^{n+1} \neq 1$ in general. The corresponding coordinate $x'^a = z'^a/z'^{n+1}$ suffers the following transformation:

$$\hat{u}^a_b = \frac{\partial x'^a}{\partial x^b}, \tag{2.7}$$

induced by (2.6). This is the quantity we have to calculate, but only at the fixed point $\mathring{z}$, as required in Eqs. (2.4) and (2.5).

First we calculate $\hat{h}$ for $h\in SU(n)\times U(1)$, which we parametrize as

$$h = wt, \tag{2.8a}$$

$$w = \begin{bmatrix} \hat{w} & \\ & 1 \end{bmatrix} \in SU(n), \tag{2.8b}$$

$$t = \exp[i\varphi\lambda_{(n^2+2n)}] \in U(1). \tag{2.8c}$$

In (2.8c), $\lambda_{(n^2+2n)} = [2n(n+1)]^{-1/2}\,\mathrm{diag}(1,\ldots,1,-n)$ is the last Gell-Mann generator of $SU(n+1)$. From (2.6) and (2.7) it follows that $\hat{w}$ is given by (2.8b) directly, while $\hat{t}$ and $\hat{h}$ can be deduced to be

$$\hat{t} = \exp(i\varphi\sqrt{(n+1)/2n})\cdot 1_n, \tag{2.9a}$$

$$\hat{h} = \hat{w}\hat{t}. \tag{2.9b}$$

To calculate the boost u_z defined by $z = u_z\mathring{z}$ is a little more complicated. Indeed, for $n>2$ it is hard to give a general parametrization, so we shall restrict to $u_z(\mathring{z})$ at the fixed point $\mathring{z}$ only which is what is needed to compute $A^a(z)$ from $A^a(\mathring{z})$ in Eq. (2.4).

This boost u_z depends on the $2n$ parameters x^a. Since $h\mathring{z}=\mathring{z}$, u_z can be expressed with h factored out from the right as follows:

$$u_z = vR, \tag{2.10a}$$

with $v\in SU(n)/SU(n-1)$ and R the $U(1)$ element

$$R = \begin{bmatrix} 1_{n-1} & & \\ & \cos\omega & \sin\omega \\ & -\sin\omega & \cos\omega \end{bmatrix}. \tag{2.10b}$$

Using (2.10) in (2.6) and (2.7) and restricting to the point $z=\mathring{z}$ we find

$$\hat{u}_z(\mathring{z})^a_b = \begin{Bmatrix} v^a_b/\cos\omega, & b\neq n, \\ x^a\Omega^2/r, & b=n. \end{Bmatrix} \tag{2.11}$$

In (2.11), we have not specified the explicit parametrization of $v^a_b\ (b\neq n)$ because when it is substituted in (2.4) the contribution of these terms is evaluated using the unitarity property of the matrix v.

C. Connection and field strength

Using (2.9) in (2.5) we find the symmetric gauge connection at the fixed point $\mathring{z}$ to be

$$A^a(\mathring{z}) = cI_{(n+1)a},\quad A^{\bar{a}}(\mathring{z}) = c^* I_{a(n+1)}, \tag{2.12}$$

where the elements of the matrix I_{ij} are given by

$$(I_{ij})_{kl} = \delta_{ik}\delta_{jl}. \tag{2.13}$$

Then, from (2.4) and (2.11) we compute the connection A^a at an arbitrary point z^a (in this patch with $z^{n+1}\neq 0$)

$$A^a(z) = c(x^b I_{ba} + I_{(n+1)a} - x^a x^b I_{b(n+1)} - x^a I_{(n+1)(n+1)}),$$

$$A^{\bar{a}}(z) = (A^a(z))^+. \tag{2.12$'$}$$

From this we compute the curvature field strength

$$F_{a\bar{b}}(z) = \partial_a A_{\bar{b}} - \partial_{\bar{b}} A_a - i[A_a, A_{\bar{b}}],$$

$$\begin{aligned} F^{a\bar{b}}(z) = (c - c^* - i|c|^2) \times\{ &-\Omega^2 I_{ba} + (x^a x^{\bar{b}} + \delta_{ab}) x^d x^{\bar{c}} I_{dc} \\ &+ x^a\Omega^2 I_{b(n+1)} + (x^a x^{\bar{b}} + \delta_{ab}) x^d I_{d(n+1)} \\ &+ x^{\bar{b}}\Omega^2 I_{(n+1)a} + (x^a x^{\bar{b}} + \delta_{ab}) x^{\bar{d}} I_{(n+1)d} \\ &+ (\delta_{ab} - r^2 x^a x^{\bar{b}}) I_{(n+1)(n+1)} \}, \end{aligned}$$

$$F^{ab}(z) = F^{\bar{a}\bar{b}}(z) = 0, \tag{2.14a}$$

which at the fixed point $\mathring{z}$ takes the simple form

$$F^{a\bar{b}}(\mathring{z}) = (c - c^* - i|c|^2)\{ -i_{ba} + \delta_{ba} I_{(n+1)(n+1)} \}. \tag{2.14b}$$

This form of $F^{a\bar{b}}(\mathring{z})$ will be used repeatedly in Secs. III and IV below, with the particular value of $c=i$ which we shall fix in the next subsection, by requiring that (2.14) obeys the Yang–Mills equation.

Before proceeding however, we make the important remark that even though $A^a(z)$ given by (2.12) belongs to

$SU(n+1)$, the field strength belongs to $SU(n)\times U(1)$. This is manifest from (2.14b) for $F^{a\bar{b}}(\mathring{z})$, and from the symmetry condition (2.4) it follows that $F^{a\bar{b}}(z)$ also belongs to $SU(n)\times n(1)$.

D. Yang–Mills equation

So far we have fixed the symmetric gauge connection and field strength up to factors of c and $(c-c^*-i|c|^2)$, respectively. Now we fix this constant c by requiring that these satisfy the Yang–Mills equation

$$0=D_aF^{a\bar{b}}\equiv(1/\sqrt{g})\partial_a(\sqrt{g}f^{a\bar{b}})-i[A_a,F^{a\bar{b}}], \quad (2.15)$$

where $g=\det g_{\alpha\beta}$. It is straightforward to find that the left-hand side of (2.15) vanishes provided that

$$(n+1)(1+ic)(c-c^*-i|c|^2)=0,$$

which has the pure-gauge (trivial)solution $c=2i$, and the nontrivial one $c=i$.

Henceforth, we shall use $c=i$, and will read Eqs. (2.12) and (2.14) accordingly.

Anticipating what follows in Secs. III and IV, we remark that $c=i$ is the choice that satisfies all the Yang–Mills-like equations.

$$D_{\alpha_1}F(2q)^{\alpha_1,\ldots,\alpha_{2q}}=0,$$

with $F(2q)$ defined as in (1.1′).

We end this section with a general remark. As we stressed in Sec. I, we propose the present example of $\mathbb{CP}^n$ as a typical step in generalizing the formulation of gauge fields on S^{2n} to arbitrary symmetric coset spaces. We now see quite clearly, the analogy between our formulas (2.12) and (2.14) for $A^a(\mathring{z})$ and $F^{a\bar{b}}(\mathring{z})$ on $\mathbb{CP}^n$, and the corresponding formulas for the sape S^{2n}, namely,[22]

$$A^m(\mathring{z})=\tfrac{1}{2}\Gamma^m \quad (2.16a)$$

$$F^{mm'}(\mathring{z})=-\tfrac{1}{4}[\Gamma^m,\Gamma^{m'}]\equiv\Gamma^{mm'},\quad m=1,\ldots,2n, \quad (2.16b)$$

where the connection A^m and $F^{mm'}$ both belong to $SO(2n)=SO_+(2n)\oplus SO_-(2n)$, with $SO_\pm(2n)$ the chiral representation given by the Γ matrices. This is in direct analogy with the present $\mathbb{CP}^n$ case, where the symmetric gauge field strength belongs to $SU(n)\times U(1)$, but the analogy is absent for the gauge connection that belongs to $SU(n+1)$. Indeed, (2.12) and (2.14b) can be cast in analogous form to (2.16a) and (2.16b) by rewriting them in terms of the $SU(n+1)$ Gell-Mann matrices λ^μ with $\mu=n,\ldots,(n^2+2n-1)$.

$$A^\mu(\mathring{z})=\lambda^\mu, \quad (2.17a)$$

$$F^{\mu\nu}(\mathring{z})=[\lambda^\mu,\lambda^\nu]. \quad (2.17b)$$

This particular parametrization was employed by us previously.[23]

III. SELF-DUALITY OF GYM SYSTEMS

All known finite-action topologically stable solutions[3,5,6] of the GYM systems on $\mathbb{R}_{4p}$ are self-dual. (In the presence of interacting Higgs fields, this is not always the case.) The self-duality equations are typified by (1.2), and they serve to saturate the Bogonolnyi bound

$$\int dx\,\mathscr{L}_{\rm GYM}(p,q) \geqslant\frac{2(i\kappa^2)^{q-p}}{(2q)!}\int d^{2(p+q)}x\epsilon^{\mu_1,\ldots,\mu_{2p}\nu_1,\ldots,\nu_{2q}} \times F(2p)_{\mu_1\ldots\mu_{2p}}F(2q)_{\nu_1\ldots\nu_{2q}}. \quad (3.1)$$

In (3.1) $\mathscr{L}_{\rm GYM}(p,q)$ is given by (1.1) and the right-hand side is proportional to the $(p+q)$th Chern–Pontryagin (C–P) integral, which takes on integer values provided that the field strength $F(2)$ satisfies the required boundary conditions.

The self-duality Eqs. (1.2) on $\mathbb{R}_{2(p+q)}$ have nontrivial solutions only when $p=q$ (and the dimensional constant κ does not feature in them). In general, for curved manifolds however, (1.2) can have nontrivial solutions in dimensions $2(p+q)=2n$ with any integer n. For example on S^{2n} these solutions were thoroughly discussed in Ref. 15 by way of illustrating the construction given there. Here we shall consider the case of GYM systems on $\mathbb{CP}^n$, and will present below, the self-duality relations that are relevant.

Before proceeding with the $\mathbb{CP}^n$ case, we briefly recall the case S^{2n}. The field strength at the northpole of S^{2n} is given by (2.16b) in terms of Γ matrices in $2(p+q)$ dimensions. The form Eq. (1.2) takes in this case, follows from the well-known Γ matrix identities

$$\Gamma^{m_1\cdots m_{2p}}=\frac{1}{(2q)!}\epsilon^{m_1\cdots m_{2p}n_1\cdots n_{2q}}\Gamma_{2n+1}\Gamma^{n_1\cdots n_{2q}}, \quad (3.2)$$

where $I^{m_1\cdots m_{2p}}$ is the $p$fold totally antisymmetrized product of the $SO(2n)$ representation matrices $\Gamma^{mm'}$ used in (2.16b), and Γ_{2n+1} is the corresponding chirality matrix. Then (1.2) splits up into the self- and anti-self-duality equations

$$F_\pm(2p)=\pm{}^*(F_\pm(2q)), \quad (3.3a)$$

$$^*(F_\pm(2q))\equiv{}^*F_\pm(2p)^{m_1\cdots m_{2p}} =\pm\frac{1}{(2q)!}\epsilon^{m_1\cdots m_{2p}n_1\cdots n_{2q}}F_\pm(2q)_{n_1\cdots n_{2q}}, \quad (3.3b)$$

where now $F_\pm(2r)$ are constructed from r factors of $F_\pm^{mm'}=\tfrac{1}{2}(1\pm\Gamma_{2n+1})\Gamma^{mm'}$, the field strengths with gauge group $SO_\pm(2n)$. This feature of the $SO(2n)$ field strength splitting into self- and anti-self-dual $SO_\pm(2n)$ field strengths will be reflected by its direct analog in what follows for $\mathbb{CP}^n$.

The gauge field strength at the fixed point of the given patch of $\mathbb{CP}^n$ is given by formula (2.14b), or alternatively by (2.17b).A straightforward but careful computation now leads to the following (anti) self-duality equations

$$^*(F(2q))_{SU_n}=(-)^{p-1}[(p-1)!/(q-1)!]F(2p)_{SU_n}, \quad (3.4a)$$

$$^*(F(2q))_{U_1}=(-1)^p(p!/q!)F(2p)_{U_1}, \quad (3.4b)$$

where $(F(2r))_{SU_n}$ denotes that part of $F(2r)$ belonging to $SU(n)$ in $SU(n+1)$, etc. We note that in the special case $p=1$, this result was already obtained in Ref. 19. There however, this duality relation is *not* employed as an equation saturating a Bogomolnyi bound.

The similarity between the (anti)self-duality equations for the curvature $2p$ forms $F(2p)$, on S^{2n} and on $\mathbb{CP}^n$ given by (3.3a) and (3.4a) and (3.4b), respectively, is obvious. In each case, the full field strength $F(2)$ and the curvature $2p$ form $F(2p)$ belong to $SO(2n)$ and $SU(n)\times U(1)$, respectively. Then, the $SO_+(2n)$ and the $SU(n)$ components of the curvature $2p$ form $F(2p)$ are self-dual, while the $SO_-(2n)$ and the $U(1)$ components are (anti)self-dual. Again, in each case the gauge connections are expressed, c.f., (2.16a) and (2.17a), in a gauge where they take their values in $SO(2n+1)$ and $SU(n+1)$, respectively.

There is however a certain difference between these cases. While the curvature $2p$-forms on S^{2n}, $F_\pm(2p)$ are p fold (antisymmetrized) products of the $SO_\pm(2n)$ field strengths $F_\pm(2)$, the curvature $2p$ forms on $\mathbb{CP}^n$, $F(2p)_{SU_n}$ and $F(2p)_{U_1}$ are *both* p fold products of the *full* $SU(n)\times U(1)$ field strength $F(2)_{SU_n\times U_1}$. This dissimilarity necessitates a brief discussion concerning the dynamics underlying these (topologically stable) field configurations.

Since the (anti)self-duality Eqs. (3.3) and (3.4) saturate the Bogomolnyi bound (3.4), then the nontrivial solutions of (3.3) and (3.4) will correspond to the absolute minimum of the action, and hence will satisfy the Euler–Lagrange equations. The action integrals in question belong to the following GYM systems on S^{2n} and $\mathbb{CP}^n$, respectively:

$$S^{(\pm)}_{GYM}(p,q)=\int_{S^{2n}}dx\,\mathrm{tr}\left[F_\pm(2p)^2+\frac{2p!}{2q!}F_\pm(2q)^2\right], \quad (3.5a)$$

$$S^{(n)}_{GYM}(p,q)=\int_{CP^n}dx\,\mathrm{tr}\left[F(2p)^2_{SU_n}+\frac{2p!}{2q!}\frac{(q-1)!^2}{(p-1)!^2}F(2q)^2_{SU_n}\right], \quad (3.6a)$$

$$S^{(1)}_{GYM}(p,q)=\int_{CP^n}dx\,\mathrm{tr}\left[F(2p)^2_{U_1}+\frac{2p!}{2q!}\frac{q!^2}{p!^2}F(2q)_{U_1}\right]. \quad (3.6b)$$

The nontriviality of the solutions is guaranteed by the nonvanishing of the C–P integrals of (3.1). This can be achieved by requiring suitable boundary conditions, since these C–P integrals are surface integrals.

Now in the case S^{2n}, the C–P integral for the $SO_\pm(2n)$ fields, respectively, are

$$I_\pm=\int_{S^{2(p+q)}}dx\,\mathrm{tr}\,F_\pm(2p)^*(F_\pm(2q))=\int_{S^{2n}}dx\,F_\pm(2)\wedge\cdots\wedge F_\pm(2)\ (n\text{ times}), \quad (3.7)$$

whose integrand is recognized as a total divergence, namely the divergence of the Chern–Simons density.From an alternative viewpoint, we can verify easily that the (anti) self-duality equations (3.3a) and (3.3b) together with the Bianchi identities solve the Euler–Lagrange equations, which in general are quite complicated, so we do not elaborate on this aspect here.

The situation in the case $\mathbb{CP}^n$ is somewhat more involved. There the right-hand side of the inequality (3.1) takes the form

$$I_n=\int_{CP^n}dx\,\mathrm{tr}\,F(2p)_{SU_n}{}^*(F(2q))_{SU_n}, \quad (3.8a)$$

$$I_1=\int_{CP^n}dx\,\mathrm{tr}\,F(2p)_{U_1}{}^*(F(2q))_{U_1}, \quad (3.8b)$$

and unlike (3.7), it is not immediately obvious here that the integrands in I_n and I_1 are both total divergences. It is however easy to verify that this is so. It turns out that I_1 is always of the form

$$I_1\approx\mathrm{tr}\,\lambda^2_{n(n+2)}\int dx\,F(2)_{U_1}\wedge F(2)_{U_1}\wedge\cdots\wedge F(2)_{U_1},\quad n\text{ times}, \quad (3.9)$$

which is manifestly a surface integral, and I_n consists of a sum of terms $I^{(i)}_n$, $i=1,\cdots,n$

$$I^{(i)}_n=\int dx\,\underbrace{F(2)_{U_1}\wedge\cdots\wedge F(2)_{U_1}}_{(n-i)\text{ times}}\wedge\mathrm{tr}\,\underbrace{F(2)_{SU_n}\wedge\cdots\wedge F(2)_{SU_n}}_{i\text{ times}}, \quad (3.10)$$

which can easily be verified to be surface integrals. Thus both I_n and I_1 in (3.8) are the surface integrals giving the Bogomolnyi bounds of the GYM systems of (3.6) on $\mathbb{CP}^n$, saturated by the (anti)self-duality equations (3.4).

The above argument depends on working in a suitable gauge where the SU_n (and U_1) field strengths are expressed directly in terms of $SU(n)$ (and U_1) valued connections. This is perfectly in order, since we have employed only gauge covariant equations, c.f. the Bogomolnyi bounds (3.1) and the (anti)self-duality Eqs (3.4), as well as the Euler–Lagrange equations corresponding to (3.6). The latter equations are not exhibited here explicitly since they are in general lengthy and not instructive for our purposes.

In summary, we state that the $SU(n)\times U(1)$ gauge fields (2.14) and (2.17) are *stable* solutions of the systems given by the action integrals (3.6), since they are solutions of the (anti)self-duality Eqs. (3.4).

Before closing this section, we make two important qualitative remarks concerning common features of the GYM field configurations on S^{2n} and $\mathbb{CP}^n$ discussed above.

Remark 1: The total (topological) C–P charge of both systems is zero. By the total charge, we mean the sum of both the instanton and anti-instanton charges characterizing the self- and anti-self-dual solutions. Thus we have the following C–P integrals in each case

$$\int_{S^{2n}}dx\,\mathrm{tr}(F(2p)_++F(2p)_-)^*(F(2q)_++F(2q)_-)=I_++I_-=0, \quad (3.10')$$

$$\int_{CP^n}dx\,\mathrm{tr}(F(2p)_{SU_n}+F(2p)_{U_1})^*(F(2q)_{SU_n}+F(2q)_{U_1})=I_n+I_1=0, \quad (3.10'')$$

where in (3.10′) we have used $I_\pm$ as defined by (3.7), and in (3.10″) I_n and I_1 as defined by (3.8a) and (3.8b).

Remark 2: The gauge invariant stress tensors $\Theta^{\mu\nu}_{SU_n}$ and $\Theta^{\mu\nu}_{U_1}$ of the systems (3.6a) and (3.6b), defined by

$$\Theta^{\mu\nu}_{SU_n}(p+q) = \int dx\, \mathrm{tr}\Big\{\Big[(F(2p)^2_{SU_n})^{\mu\nu} - \frac{1}{2p}\delta^{\mu\nu}F(2p)^2_{SU_n}\Big] + \frac{2p!(q-1)!^2}{2q!(p-1)!^2}\Big[(F(2q)^2_{SU_n})^{\mu\nu} - \frac{1}{2q}\delta^{\mu\nu}F(2q)^2_{SU_n}\Big]\Big\}, \tag{3.11a}$$

$$\Theta^{\mu\nu}_{U_1}(p+q) = \int dx\, \mathrm{tr}\Big\{\Big[(F(2p)^2_{U_1})^{\mu\nu} - \frac{1}{2p}\delta^{\mu\nu}F(2p)^2_{U_1}\Big] + \frac{2p!}{2q!}\frac{q!^2}{q!^2}\Big[(F(2q)^2_{U_1})^{\mu\nu} - \frac{1}{2q}\delta^{\mu\nu}F(2q)^2_{U_1}\Big]\Big\}, \tag{3.11b}$$

vanish identically for the (anti)self-dual field configurations satisfying (3.4a) and (3.4b). In (3.11) we have used the notation $(F(2r)^2)^{\mu\nu} \equiv F(2r)^{\mu\rho_1\cdots\rho_{2r-1}}F(2r)^{\nu\rho_1\cdots\rho_{2r-1}}$.

This property of the vanishing of the stress tensor for anti-self-dual field configuration is shared[15] with the gauge fields on S^{2n}.

IV. EYM SYSTEMS ON $\mathbb{CP}^n$

The (anti) self-duality of the solutions discussed above automatically guarantees their stability. Field configurations whose action is not bounded from below by a Bogomolnyi inequality like (3.1) are not guaranteed to be stable, and in this sense are less interesting than the solutions presented in Sec. III. On the other hand many well-known field configurations satisfying Euler–Lagrange equations are not (anti)self-dual, and it is therefore of some interest as a matter of completeness, to study these solutions in the context of the dynamics giving rise to them. The question of the stability of these solutions goes beyond the scope of the present work, and we plan to report on it elsewhere.

The simplest example of non-self-dual solutions, are the gauge fields on S^{2n} and $\mathbb{CP}^n$, respectively, which satisfy the Yang–Mills equation. In all dimensions except four (with $n=2$), there occur no selfduality equations.

It also turns out that the Euler–Lagrange equations of *all* the GYM systems (1.1) are satisfied on S^{2n} and $\mathbb{CP}^n$, and again, except in dimension $d=2n$, these solutions are not endowed with (anti)self-duality equations saturating a Bogomolnyi bound.

Giving up the requirement that a solution be (anti)self-dual naturally leads one to consider the dynamics of the EYM systems (1.3). As explained in Sec. I, these systems are not endowed with Bogomolnyi equations, but nevertheless products of the curvature field strength satisfy certain duality relations which together with the Bianchi identities imply the Euler–Lagrange equations.

It can be deduced from (1.3) that the Euler–Lagrange equation corresponding to $L^{(N)}$ has the form

$$D_\alpha \mathscr{F}^{\alpha\beta}_{(N)} = 0, \tag{4.1}$$

where $\mathscr{F}^{\alpha\beta}_{(N)}$ are given in terms of powers of $F(2)$, and can be readily computed in each case. Here (4.1) is referred to a d-dimensional space, not necessarily $d=2N$. (This means that in the case of the *spheres* S^d, we cannot infer that the Euler–Lagrange equations for the system on $\mathbb{R}_d$ are also satisfied,[12,13] except when $d=2N$, by exploiting the conformal invariance of $L^{(N)}$ in that case.)

The tensor fields $\mathscr{F}^{\alpha\beta}_{(N)}$ for $N=2$ are simply the curvature two-forms, and then (4.1) is the Yang–Mills equation. We list $\mathscr{F}^{\alpha\beta}_{(N)}$ here, for $N=2,3$ and 4. (Fujii[13] has given also the case of $N=5$, which we shall not consider here.) These are

$$\mathscr{F}^{\alpha\beta}_{(2)} = F^{\alpha\beta}, \tag{4.2a}$$

$$\mathscr{F}^{\alpha\beta}_{(3)} = [F^{\alpha\lambda}, F^{\beta}_{\lambda}], \tag{4.2b}$$

$$\mathscr{F}^{\alpha\beta}_{(4)} = \{F^{\alpha\beta}, F^{\gamma\delta}F_{\gamma\delta}\} + F^{\gamma\delta}F^{\alpha\beta}F_{\gamma\delta} - 2\{F_{\gamma\delta}, F^{\gamma[\alpha}F^{\beta]\delta}\} - 2F^{\alpha]\gamma}F_{\gamma\delta}F^{\delta[\beta}. \tag{4.2c}$$

Our result is the following: The Euler–Lagrange Eqs. (4.1) are satisfied for $\mathscr{F}^{\alpha\beta}_{(N)}$ $(N=2,3,4)$ given by (4.2a)–(4.2c) by the symmetric gauge fields on S^{2n} and $\mathbb{CP}^n$. Note that n here is completely unrelated to N characterizing $L^{(N)}$ of (1.3). (In particular for $n=N$, the solutions on S^{2n} can be related via a stereographic projection, to the solutions on $\mathbb{R}_{2n}$ given in Ref. 13.) We expect that this result is true for EYM systems characterized with arbitrary N.

It is straightforward to establish our results. In particular, to show that the $SU(n)\times U(1)$ gauge field on $\mathbb{CP}^n$ satisfies (4.1) for each $\mathscr{F}^{\alpha\beta}_{(N)}$ of (4.2) can be verified straightforwardly by using our formulas derived in Sec. II. The corresponding demonstration for the $SO_\pm(2n)$ gauge fields on S^{2n} can be performed analogously.

Here our main aim is to highlight the parallel between the dynamics of EYM systems on the two symmetric spaces S^{2n} and $\mathbb{CP}^n$ which we have found. To this end, it is interesting to find out just how far this parallel goes.

The above result, namely that the Euler–Lagrange Eqs. (4.1) are satisfied, can follow also from the Bianchi identities if some duality relation like

$$\mathscr{F}^{\alpha\beta}_{(N)} \simeq \epsilon^{\alpha\beta\gamma_1\cdots\gamma_{2n-2}}F_{\gamma_1\gamma_2}F_{\gamma_3\gamma_4}\cdots F_{\gamma_{2n-3}\gamma_{2n-2}}, \tag{4.3}$$

were satisfied. This approach was particularly emphasized in Ref. 12.

Now the $SO_\pm(2n)$ gauge field on S^{2n} given by (2.16b) clearly satisfies the duality relations (4.3), by virtue of well-known (γ matrix) spinor identities. In this case, i.e., the $SO_\pm(2n)$ gauge field on S^{2n}, the Euler–Lagrange Eqs. (4.1) of the EYM systems characterized with arbitrary N are satisfied.

The corresponding statement for $SU_n\times U_1$ gauge fields on $\mathbb{CP}^n$ turns out not to be quite as straightforward. This is the last question we study in this paper, and we consider it relevant in that it highlights how far the analogy between the two symmetric spaces S^{2n} and $\mathbb{CP}^n$ goes, and where it stops.

We consider each case (4.2a)–(4.2c) in turn. The first one, $n=2$, is already known to satisfy (4.1) by construction, or alternatively because it satisfies the duality relations (3.4). The second one, $n=3$, also satisfies (4.1). In this case the duality relation satisfied is qualified as follows: Since $\mathscr{F}^{\alpha\beta}_{(3)}$ belongs only to $SU(n)$, we define a dual field obtained from (4.3) by subtracting the $U(1)$ part

$$*\mathcal{F}^{\alpha\beta} = (1/2^{2n-1})\epsilon^{\alpha\beta\gamma_1,\dots,\gamma_{2n-2}} \times (F_{\gamma_1\gamma_2}\cdots F_{\gamma_{2n-3}\gamma_{2n-2}})\mathrm{SU}_n. \quad (4.4)$$

Then the duality relation that holds is

$$\mathcal{F}^{\alpha\beta}_{(3)} = [-(i)^{n-1}n/(n-2)!]*\mathcal{F}^{\alpha\beta}. \quad (4.5)$$

It then follows from (4.5) and (4.4) that Eq. (4.1) with $n = 3$ is satisfied by the $\mathrm{SU}(n)\times \mathrm{U}(1)$ field on $\mathbb{CP}^n$ for all n.

Finally, we consider the Eq. (4.3) for $\mathcal{F}^{\alpha\beta}_{(4)}$ of (4.2c). Again we would need a duality relation like (4.3), but we find that this does not occur. Thus the analogy between the symmetric gauge fields on S^{2n} and $\mathbb{CP}^n$ breaks down at this point. The cause is, the privileged status of S^{2n} in the definition of the EYM system. For example, if this definition was so modified that $\mathcal{F}^{\alpha\beta}_{(4)}$ were replaced by $\tilde{\mathcal{F}}^{\alpha\beta}_{(4)}$

$$\tilde{\mathcal{F}}^{\alpha\beta}_{(4)} = \tfrac{1}{2}\{[[F^{\alpha\beta},F^{\gamma\delta}],F_{\gamma\delta}] + [[F_{\gamma\delta},F^{\gamma[\alpha}],F^{\beta]\delta}], \quad (4.6)$$

then indeed the duality relation

$$\tilde{\mathcal{F}}^{\alpha\beta}_{(4)} = -(i)^n[2n(n+1)/(n-2)!]*\mathcal{F}^{\alpha\beta}$$

would follow, with $*\mathcal{F}^{\alpha\beta}$ defined by (4.4).

Replacing (4.2c) by (4.6) however changes the definition of an EYM system, and enlarges the context in which dynamical systems can be defined rather arbitrarily, so we stop our study at this point.

This completes our discussion of the parallel between the EYM dynamics of symmetric gauge fields on S^{2n} and $\mathbb{CP}^n$.

ACKNOWLEDGMENTS

D. H. Tch. thanks Dr. B. P. Dolan for numerous helpful discussions and Zh-Q. M. would like to thank Professor J. T. Lewis for his hospitality at the Dublin Institute for Advanced Studies.

[1]G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[2]A. M. Polyakov, JETP Lett. **20**, 194 (1974).
[3]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).
[4]D. H. Tchrakian, J. Math. Phys. **21**, 166 (1980).
[5]D. H. Tchrakian, Phys. Lett. B **150**, 360 (1985).
[6]B. Brossman, T. W. Kephart, and J. D. Stasheff, Commun. Math. Phys. **96**, 431 (1984).
[7]A. Chakrabarti, T. N. Sherry, and D. H. Tchrakian, Phys. Lett. B **162**, 340 (1985).
[8]D. H. Tchrakian, Phys. Lett. B **155**, 255 (1985).
[9]J. Burzlaff and D. H. Tchrakian, Nuovo Cimento A **92**, 107 (1986).
[10]G. M. O'Brien and D. H. Tchrakian, Phys. Rev. D **35**, 1468 (1987).
[11]P. Forgács and N. S. Manton, Commun. Math. Phys. **72**, 15 (1980).
[12]C. Saçlioğlu, Nucl. Phys. B **277**, 487 (1986).
[13]K. Fujii, Lett. Math. Phys. **12**, 363 (1986).
[14]E. Cremmer and J. Scherk, Nucl. Phys. B **108**, 409 (1976), **118**, 61 (1977).
[15]G. M. O'Brien and D. H. Tchrakian, J. Math. Phys. **29**, 1212 (1988).
[16]A. S. Schwarz, Commun. Math. Phys. **56**, 79 (1977).
[17]C-N. Yang, Phys. Rev. Lett. **33**, 445 (1974).
[18]C-H. Gu, Fudan J. Nat. Sci. **2**, 51 (1976); Phys. Rep. **80**, 251 (1980); Zh-Q. Ma, Nucl. Phys. B **231**, 172 (1984).
[19]F. A. Bais and P. Batenburg, Nucl. Phys. B **269**, 363 (1986).
[20]A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977); J. Nowakowski and A. Trautman, J. Math. Phys. **19**, 1100 (1978); J. Tafel and S. Shnider, *ibid.* **21**, 2236 (1980).
[21]G. W. Gibbons and C. N. Pope, Commun. Math. Phys. **61**, 239 (1978).
[22]Zh-Q. Ma, G. M. O'Brien, and D. H. Tchrakian, Phys. Rev. D **33**, 1177 (1986); D. O'Sé, T. N. Sherry, and D. H. Tchrakian, J. Math. Phys. **27**, 325 (1986).
[23]Zhong-Qi Ma and D. H. Tchrakian, Phys. Rev. D **38**, 3872 (1988).

Parastatistics, supersymmetry and parasupercoherent states

J. Beckers and N. Debergh
Theoretical and Mathematical Physics, Institute of Physics, B.5, University of Liège, B-4000 Liège 1, Belgium

(Received 20 June 1989; accepted for publication 17 January 1990)

Coherent states are constructed in $p=2$-parasupersymmetric quantum mechanics in connection with the two relative para-Bose and para-Fermi sets of trilinear structure relations. They are called parasupercoherent states. Parasupersymmetric operators [such as the Hamiltonian and (two) annihilation operators] are introduced and discussed. In particular, the parasuperspectrum is determined and compared with the recent results obtained by Rubakov and Spiridonov [Mod. Phys. Lett. A **3**, 1337 (1988)]. The superalgebra contents subtended by osp(2/2) and osp(3/2) are analyzed and exploited in order to get constants of motion in both contexts through parallel properties on the supersymmetric harmonic oscillator.

I. INTRODUCTION

During the last three decades, *three* very important fields in modern theoretical physics have appeared implying relatively elaborate mathematical developments where Lie group theory and associated graded Lie algebras[1] (or superalgebras) play a prominent part. In the 1950's, Green[2] has introduced the so-called "parastatistics"[3] while Glauber[4] and Klauder[4] have shown in the 1960's the main properties of the so-called "coherent states" as well as their interest.[5] In the 1970's, the third field was introduced in particle physics for combining bosons and fermions in common (super)multiplets; it has been called "supersymmetry."[6] Finally, in the 1980's, we notice that, among these three fields—parastatistics, coherent states, and supersymmetry—*two* of them have already been arbitrarily combined and superposed:

(a) parastatistics *and* coherent states have led to the construction of para-Bose (or parabosonic) coherent states, as presented by Sharma *et al.*,[7,8] for example:

(b) supersymmetry *and* coherent states have been combined in order to study supercoherent states, as was shown in recent papers;[9–11]

(c) parastatistics *and* supersymmetry have been superposed, leading to a first introduction of parasupersymmetric quantum mechanics as initiated very recently by Rubakov and Spiridonov[12] (see also Biswas and Soni[13]).

In this study, as a prolongation of our recent works,[14,15] we want to superpose the *three* above fields with a specific aim: the study of coherent states in the *simplest* ($p=2$)-parasupersymmetric theory where we are dealing with parabosonic and parafermionic degrees of freedom while, for example, Rubakov and Spiridonov have only considered symmetries between bosons and parafermions. In order to illustrate our developments, we will, for simplicity, consider only *a pair of parafields* but this can be easily generalized on the basis of Green's[2] and Greenberg–Messiah's[16] contributions, the only restriction being to consider parabosonic and parafermionic degrees of freedom of the *same* order.[16]

This article is constructed as follows. In Sec. II, we will fix our notations for combining parastatistics and supersymmetry. We will also give a basis[14] well adapted for the study of the parasupersymmetric Hamiltonian corresponding to bosons[12] and parafermions (Sec. II A) *or* to parabosons and parafermions (Sec. II B). The latter case will be subtended by the *two* relative para-Bose and para-Fermi sets of *trilinear* structure relations issued from Greenberg–Messiah's developments.[16] These two contexts will be studied within the supersymmetric point of view through the standard procedure of supersymmetrization "à la Witten."[17] In Sec. III, we will study the implications of such a $p=2$-parasupersymmetric quantum mechanics dealing with Bose-like and Fermi-like harmonic oscillators. The energy spectrum will be obtained and discussed. Section IV will be devoted to the coherent states associated with this parasupersymmetric theory—we call them the parasupersymmetric coherent states or parasupercoherent states. In fact, we will construct two sets of parasupercoherent states (Secs. IV A and B) according to the two relative para-Bose and para-Fermi sets and we will discuss more particularly (Sec. IV C) those which correspond to the relative para-Fermi context in $p=2$-parasupersymmetric quantum mechanics. An analysis of some expectation values and their interest in connection with the uncertainty principle will also be presented. Finally, in Sec. V, we will emphasize the superalgebra contents of the preceding developments through the orthosymplectic Lie structures osp(2/2) and osp(3/2) (respectively, in Sec. V A and B). We will give in this way the constants of motion of the corresponding contexts and discuss their correspondence (Sec. V C).

As far as bosonic quantum harmonic oscillators (in one spatial dimension) are concerned in this article, we take as units their masses, their angular frequencies, and the Dirac constant $\hbar$.

II. PARASTATISTICS AND SUPERSYMMETRY

Let us first (Sec. II A) fix our notations according essentially to the Rubakov–Spiridonov work[12] and give[14] an *ad hoc* basis for the study of the parasupersymmetric Hamiltonian when bosons and parafermions are included. Then, let us (Sec. II B) extend these considerations to parabosons

 0022-2488/90/061513-11$03.00

and parafermions by constructing the supersymmetric version of the parabosonic Hamiltonian. This second aim will be realized (Sec. II B 1 and II B 2) through the *two* relative para-Bose and para-Fermi sets of trilinear structure relations issued from Greenberg–Messiah's developments[16] when a pair of parafields are concerned. Initially, we want to point out that the resulting parasupersymmetric Hamiltonian has the same form in each of these contexts.

A. Bosons and parafermions

Proposed by Rubakov and Spiridonov,[12] the second-order parasupersymmetric quantum mechanics of one bosonic and one parafermionic degrees of freedom can be summarized in an interesting way when the specific superpotentials refer to a one-dimensional harmonic oscillator in a homogeneous magnetic field. By taking $W_1 = W_2 = -x$, the corresponding parasupersymmetric Hamiltonian takes the form

$$H = H_b + H_{pf} = (a^\dagger a + \tfrac{1}{2}) - \tfrac{1}{2}[b^\dagger, b], \qquad (2.1)$$

where the bosonic (b) and parafermionic (pf) parts are defined as usual. Here the bosonic annihilation (a) and creation ($a^\dagger$) operators satisfy the *bilinear* currently expected commutation relations[18,19]

$$[a,a^\dagger] = 1, \quad [a,a] = [a^\dagger,a^\dagger] = 0, \qquad (2.2)$$

while the parafermionic operator b and its Hermitian conjugate $b^\dagger$ are realized in terms of 3×3 matrices[3,12] and verify *trilinear* relations summarized by

$$b^3 = 0, \quad bb^\dagger b = 2b, \quad b^2 b^\dagger + b^\dagger b^2 = 2b \qquad (2.3)$$

and their Hermitian conjugates. An *ad hoc* realization is given by

$$b = \sqrt{2}\begin{pmatrix} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \end{pmatrix}, \quad b^\dagger = \sqrt{2}\begin{pmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot \end{pmatrix},$$
$$J_3 = \tfrac{1}{2}[b^\dagger, b] = \begin{pmatrix} 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 \end{pmatrix}, \qquad (2.4)$$

leading to the diagonal Hamiltonian (2.1) obtained from the conserved parasupercharges

$$Q = -2iab, \quad Q^\dagger = 2ib^\dagger a^\dagger, \qquad (2.5)$$

which are such that[12]

$$Q^3 = 0, \quad Q^2 Q^\dagger + QQ^\dagger Q + Q^\dagger Q^2 = QH,$$
$$[H,Q] = 0$$

and

$$(Q^\dagger)^3 = 0, \quad (Q^\dagger)^2 Q + Q^\dagger Q Q^\dagger + Q(Q^\dagger)^2 = Q^\dagger H,$$
$$[H,Q^\dagger] = 0. \qquad (2.6)$$

After Rubakov and Spiridonov,[12] we point out that this parasupersymmetric theory and the superalgebra (2.6) imply no restriction on the parasupersymmetric vacuum energy: the spectrum is threefold degenerate except for the few lowest energy levels. In fact, there are in this harmonic context a unique nondegenerate vacuum state (hereafter called $|\psi_0\rangle$) with negative energy ($E_0 = -1/2$) corresponding to the eigenvalue $+1$ of J_3, a twofold degenerate state called $|\psi_1\rangle$ with the energy $E_1 = 1/2$ corresponding to the eigenvalues 1 and 0 of J_3, as well as an infinite set of threefold degenerate states called $|\psi_n\rangle$ corresponding to the three eigenvalues $\pm 1, 0$ of J_3.

Due to the fact that the parafermionic number operator is given by

$$N_{pf} = J_3 + \mathbb{1}_3, \qquad (2.7)$$

we notice that the purely bosonic states have the structure $(0,0,*)^T$ while the purely parafermionic ones are of the following forms: $(0,*,0)^T$ or $(*,0,0)^T$. Then, referring to the eigenvalues $0,1,2,\ldots,n,\ldots$ of the bosonic particle number operator $N = a^\dagger a$, we have constructed the following basis:[14]

$$|\psi_0\rangle = (|0\rangle,0,0)^T,$$
$$|\psi_1\rangle = (0,|0\rangle,0)^T + (|1\rangle,0,0)^T, \qquad (2.8)$$
$$|\psi_n\rangle = \alpha_n(0,|n-1\rangle,0)^T + \beta_n(|n\rangle,0,0)^T$$
$$+ \gamma_n(0,0,|n-2\rangle)^T, \quad n \geqslant 2,$$

where α_n, β_n, γ_n are arbitrary complex numbers. Through the well-known information[19]

$$N|n\rangle = a^\dagger a|n\rangle = n|n\rangle,$$
$$a|n\rangle = \sqrt{n}|n-1\rangle, \qquad (2.9)$$
$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle,$$

we immediately get as previously interpreted

$$H|\psi_n\rangle = (n - 1/2)|\psi_n\rangle, \quad n \geqslant 0. \qquad (2.10)$$

B. Parabosons and parafermions

Let us now consider the parabosonic case instead of the bosonic one. Then the Hamiltonian H_b in Eq. (2.1) has to be replaced by

$$H_{pb} = \tfrac{1}{2}(aa^\dagger + a^\dagger a) = \tfrac{1}{2}\{a,a^\dagger\} \qquad (2.11)$$

according to Sharma *et al.*,[7] for example, with the more general commutation relation(s)

$$[a,H_{pb}] = a, \quad [a^\dagger,H_{pb}] = -a^\dagger. \qquad (2.12)$$

Let us just recall[20] that Eq. (2.12) follows from Eq. (2.2) but the reverse is in general not true. Such a parabosonic Hamiltonian can now be supersymmetrized, so that we will be really led to the simplest parasupersymmetric $p = 2$ theory dealing with para-Bose and para-Fermi operators having the same order as required[16] when we want to ensure the validity of the mixed trilinear relations (see hereafter). We are thus concerned with a pair of parafields characterized by well-defined structure relations which are distributed in *two* nonequivalent ways leading to the so-called *relative para-Bose* and *relative para-Fermi* sets after Greenberg and Messiah.[16]

1. Relative para-Bose supersymmetry

After the Green[2] and Greenberg–Messiah[16] studies, the nontrivial *trilinear* relations for the *purely* parabosonic case reduce in our context to

$$[a,\{a^\dagger,a\}] = 2a, \quad [a^\dagger,\{a,a\}] = -4a, \qquad (2.13a)$$

while, for the *purely* parafermionic case, they reduce to

$$[b,[b^\dagger,b]] = 2b, \tag{2.13b}$$

a compatible relation with Eqs. (2.3). Moreover, there are *mixed* structure relations between a, b, and their Hermitian conjugates which take the forms

$$[\{a^\dagger,a\},b] = 0, \quad [[b^\dagger,b],a] = 0, \tag{2.13c}$$

$$[\{a,a\},b] = [\{a^\dagger,a^\dagger\},b] = 0, \tag{2.13d}$$

and

$$\begin{aligned} &[\{a,b\},a^\dagger] = -[\{a^\dagger,b\}a] = 2b, \\ &\{\{a,b^\dagger\},b\} = \{\{a,b\},b^\dagger\} = 2a, \\ &[\{a,b\},a] = [\{a^\dagger,b\},a^\dagger] \\ &\qquad = \{\{a,b\},b\} = \{\{a,b^\dagger\},b^\dagger\} = 0. \end{aligned} \tag{2.13e}$$

Let us point out the physical meaning of four of the above relations: the first equation (2.13a) simply translates our Eq. (2.12); Eq. (2.13b) is its analogous but in the parafermionic context; both Eqs. (2.13c) mean that the respective Hamiltonians $H_{\rm pb}$ and $H_{\rm pf}$ commute with the other annihilation operator as expected.

Within such a *relative para-Bose set* of trilinear structure relations, our supersymmetric theory is subtended by the parasupersymmetric Hamiltonian $H_{\rm PSS}$ given by

$$H_{\rm PSS} = H_{\rm pb} + H_{\rm pf} = \tfrac{1}{2}\{a^\dagger,a\} - (1/2)[b^\dagger,b], \tag{2.14}$$

which has to be compared with the supersymmetric Hamiltonian (2.1) considered by Rubakov and Spiridonov.

It is remarkable that it can be obtained through the *standard* procedure of supersymmetrization à la Witten[17] from the supercharges defined by

$$Q_1 = \tfrac{1}{2}\{a,b\}, \quad Q_1^\dagger = \tfrac{1}{2}\{b^\dagger,a^\dagger\}, \tag{2.15}$$

i.e.,

$$H_{\rm PSS} = \{Q_1,Q_1^\dagger\}, \tag{2.16}$$

when typically the *mixed* trilinear structure relations (2.13e) are explicitly used. Moreover, we have

$$\begin{aligned} &\{Q_1,Q_1\} = \{Q_1^\dagger,Q_1^\dagger\} = 0, \\ &[H_{\rm PSS},Q_1] = [H_{\rm PSS},Q_1^\dagger] = 0, \end{aligned} \tag{2.17}$$

so that the supersymmetry is subtended by the same superalgebra as in ordinary supersymmetric quantum mechanics and in contradistinction with the Rubakov–Spiridonov superalgebra[12] [see Eqs. (2.6)]. The specific content of the theory governed by the parasupersymmetric Hamiltonian (2.14) will be described in Sec. III.

Let us finally notice that the above considerations are analogous to those that have led Biswas and Soni[13] to construct what they called "the algebraic structure of generalized parastatistics" which is nothing else in our context than the relative para-Bose set obtained by Greenberg and Messiah.[16]

2. Relative para-Fermi supersymmetry

After Greenberg and Messiah,[16] there is another set of trilinear structure relations that is called the *relative para-Fermi set* and which is not equivalent to the above one. It is characterized by the *same* structure relations (2.13a)–(2.13d) given above but by, instead of Eqs. (2.13e), the new *mixed* following ones:

$$\begin{aligned} &\{a,[a^\dagger,b]\} = \{a^\dagger,[b,a]\} = 2b, \\ &[b^\dagger,[b,a]] = [b,[b^\dagger,a]] = 2a, \\ &[b,[a,b]] = [b^\dagger,[a,b^\dagger]] \\ &\qquad = \{a,[b,a]\} = \{a^\dagger,[b,a^\dagger]\} = 0. \end{aligned} \tag{2.18}$$

The set (2.13a)–(2.13d) *and* (2.18) of trilinear structure relations will refer to the relative para-Fermi considerations. It has already been exploited by Palev[21] in connection with orthosymplectic Lie superalgebras (see Sec. V).

Inside such a relative para-Fermi theory, the supersymmetric Hamiltonian $H_{\rm PSS} \equiv (2.14)$ can once again be analyzed. Here, also, we propose a standard procedure of supersymmetrization à la Witten but from the supercharges defined by

$$Q_2 = \tfrac{1}{2}[a,b], \quad Q_2^\dagger = \tfrac{1}{2}[b^\dagger,a^\dagger], \tag{2.19}$$

such that

$$H_{\rm PSS} = \{Q_2, Q_2^\dagger\} \tag{2.20}$$

and

$$\begin{aligned} &\{Q_2,Q_2\} = \{Q_2^\dagger, Q_2^\dagger\} = 0, \\ &[H_{\rm PSS},Q_2] = [H_{\rm PSS},Q_2^\dagger] = 0. \end{aligned} \tag{2.21}$$

Here again, the specific *mixed* trilinear structure relations (2.18) play the prominent role in order to get Eqs. (2.20)–(2.21).

The parasupersymmetric Hamiltonian (2.14) is thus a *common* operator obtained in both relative paracontexts and its implications and properties will be interesting to obtain explicitly (see Sec. III).

As a final comment in this section, let us draw attention to the fact that both superalgebras (2.16)–(2.17) *and* (2.20)–(2.21) imply the Rubakov–Spiridonov superalgebra (2.6) but that the reverse is, in general, not true. This is completely consistent with the fact that bosons are particular parabosons.

III. $p=2$-PARASUPERSYMMETRIC QUANTUM MECHANICS

In this section, let us develop the quantum mechanical study associated with the parasupersymmetric Hamiltonians

$$H^{(\varepsilon)}_{\rm PSS} = H_{\rm pb} + \varepsilon H_{\rm pf} = \tfrac{1}{2}\{a,a^\dagger\} + (\varepsilon/2)[b^\dagger,b], \tag{3.1}$$

where $\varepsilon = \pm 1$ refers to specific choices that can be developed in a complete parallel way. Let us notice that these two possibilities leave the physical content unchanged due to the typical property[3] of Fermi-like oscillators under the transformation $b \to b^\dagger, b^\dagger \to b$ implying $J_3 \to -J_3$ [see Eq. (2.4)]. According to recent contributions[12,14] let us first choose $\varepsilon = -1$ so that we are dealing with the parasupersymmetric Hamiltonian (2.14), which, in the realization (2.4), takes the diagonal form

$$H^{(-)}_{\rm PSS} = \begin{pmatrix} \tfrac{1}{2}\{a,a^\dagger\} - 1 & & \\ & \tfrac{1}{2}\{a,a^\dagger\} & \\ & & \tfrac{1}{2}\{a,a^\dagger\} + 1 \end{pmatrix}. \tag{3.2}$$

By considering the relative para-Bose context (2.13), the supercharges (2.15) and the superalgebra (2.16)–

(2.17), it is not difficult to convince ourselves that the following properties are coming out. All the energy eigenvalues are semipositive definite,

$$E = \langle\psi| H_{\rm PSS}^{(-)} |\psi\rangle \geqslant 0, \quad \forall|\psi\rangle, \tag{3.3a}$$

and E vanishes iff we have

$$Q_1|\psi\rangle = 0 \; and \; Q_1^\dagger \; |\psi\rangle = 0. \tag{3.3b}$$

Thus, in our basis (2.8), $|\psi_0\rangle$ is a unique nondegenerate state with zero energy—the ground state in our $p = 2$ theory—and all the other states $|\psi_n\rangle$ $(n \geqslant 1)$ are degenerate and correspond to positive energies.

The energy spectrum can be completely determined by remembering the parabosonic properties[7] associated with the Hamiltonian (2.11). Recalling that the para-Bose number states $|n\rangle$ are defined by

$$N_{\rm pb}|n\rangle = (H_{\rm pb} - h_0)|n\rangle = n|n\rangle, \tag{3.4}$$

with

$$a|2l\rangle = \sqrt{2l}\,|2l-1\rangle, \quad a|2l+1\rangle = \sqrt{2(l+h_0)}\,|2l\rangle,$$
$$a^\dagger|2l\rangle = \sqrt{2(l+h_0)}\,|2l+1\rangle,$$
$$a^\dagger|2l+1\rangle = \sqrt{2l+2}\,|2l+2\rangle, \tag{3.5}$$

where h_0 is the lowest (non–negative) eigenvalue of $H_{\rm pb}$. We evidently deduce that

$$H_{\rm pb}|n\rangle = \tfrac{1}{2}\{a,a^\dagger\}|n\rangle = (n + h_0)|n\rangle. \tag{3.6}$$

Consequently, we get on the states (2.8)

$$H_{\rm PSS}^{(-)}\,|\psi_n\rangle = (n + h_0 - 1)|\psi_n\rangle, \quad \forall n \geqslant 0, \tag{3.7}$$

and the energy spectrum of the $(p = 2h_0 = 2)$ theory is given by

$$E_n = n, \quad \forall n = 0,1,2,\ldots . \tag{3.8}$$

It shows a perfect supersymmetric character[17] as expected through Eqs. (3.3) and the classification of all the admissible parasupersymmetric states can be pointed out through the action of Q_1, $Q_1^\dagger$, and $Q_1^\dagger\, Q_1$. In a parallel way with the context of usual supersymmetric quantum mechanics,[17,22] we notice that the parafermionic number operator $N_{\rm pf} \equiv (2.7)$ as well as the parabosonic one $N_{\rm pb} \equiv (3.4)$ are such that with the supercharges Q_1 and $Q_1^\dagger$

$$[Q_1,N_{\rm pf}] = Q_1, \quad [Q_1^\dagger,N_{\rm pf}] = -Q_1^\dagger,$$
$$[Q_1,N_{\rm pb}] = Q_1, \quad [Q_1^\dagger,N_{\rm pb}] = -Q_1^\dagger, \tag{3.9}$$

and have the required conserved character:

$$[H_{\rm PSS}^{(-)},N_{\rm pf}] = [H_{\rm PSS}^{(-)},N_{\rm pb}] = 0. \tag{3.10}$$

If we define the action of these number operators on parabosonic and parafermionic states $|{\rm pb}\rangle$ and $|{\rm pf}\rangle$, respectively, by

$$N_{\rm pf}|{\rm pb}\rangle = 0 \Rightarrow n_{\rm pf} = 0, \quad N_{\rm pb}|{\rm pf}\rangle = 0 \Rightarrow n_{\rm pb} = 0, \tag{3.11}$$

we learn that, among the eight possible states, the states $|{\rm pb}\rangle$, $Q_1^\dagger\, Q_1|{\rm pb}\rangle$, $Q_1|{\rm pf}\rangle$, and $Q_1^\dagger|{\rm pf}\rangle$ are purely parabosonic states $(n_{\rm pb} \neq 0)$ while the states $|{\rm pf}\rangle$, $Q_1^\dagger\, Q_1|{\rm pf}\rangle$, $Q_1|{\rm pb}\rangle$, and $Q_1^\dagger|{\rm pb}\rangle$ are purely parafermionic states $(n_{\rm pf} \neq 0)$.

Within our basis (2.8), it is finally easy to show that the state $|\psi_1\rangle$ is twofold degenerate while the states $|\psi_n\rangle$, $n \geqslant 2$ are threefold degenerate. We thus get the spectrum quoted in Fig. 1(a) for parabosons and parafermions in comparison with Fig. 1(b) characterizing the current supersymmetric context[22] between bosons and fermions. One parafermion is on the nondegenerate groundstate and corresponds to the $+1$-eigenvalue of J_3, etc.

Let us notice that the spectrum in Fig. 1(b) is also readily obtained from our considerations as a particular case. Indeed, inside the realization (2.4) for b and $b^\dagger$, we, respectively, distinguish the 2×2 submatrices σ_- and σ_+ which play the role of the fermionic operators in the usual context of supersymmetric quantum mechanics. We thus get here, when dealing with bosons and fermions:

$$\sigma_- = \begin{pmatrix} \cdot & \cdot \\ 1 & \cdot \end{pmatrix}, \quad \sigma_+ = \begin{pmatrix} \cdot & 1 \\ \cdot & \cdot \end{pmatrix}, \quad \tfrac{1}{2}[\sigma_+,\sigma_-] = \tfrac{1}{2}\sigma_3 = H_f. \tag{3.12}$$

The supersymmetric Hamiltonian becomes, as expected,

$$H_{\rm SS}^{(-)} = H_{\rm bf} = H_{\rm b} - H_{\rm f} = \begin{pmatrix} a^\dagger a & 0 \\ 0 & a^\dagger a + 1 \end{pmatrix} \tag{3.13a}$$

and the fermionic number operator is given by

$$N_{\rm f}^{(-)} = \tfrac{1}{2}(\mathbb{1} - \sigma_3). \tag{3.13b}$$

These operators act on the 2-dimensional subspace directly obtained from the space subtended by the basis (2.8) and now characterized by the states

$$|\psi_0'\rangle = \begin{pmatrix} |0\rangle \\ 0 \end{pmatrix},$$
$$|\psi_n'\rangle = \alpha_n'\begin{pmatrix} 0 \\ |n-1\rangle \end{pmatrix} + \beta_n'\begin{pmatrix} |n\rangle \\ 0 \end{pmatrix}, \quad \forall n \geqslant 1. \tag{3.14}$$

This immediately leads (in particular) to the spectrum quoted in Fig. 1(b) showing that one boson is on the unique (nondegenerate) ground state while all the other states are twofold degenerate and contain one boson and one fermion.

For completeness, let us also mention that our discussion on $H_{\rm PSS}^{(-)}$ contains the immediate properties of the Ru-

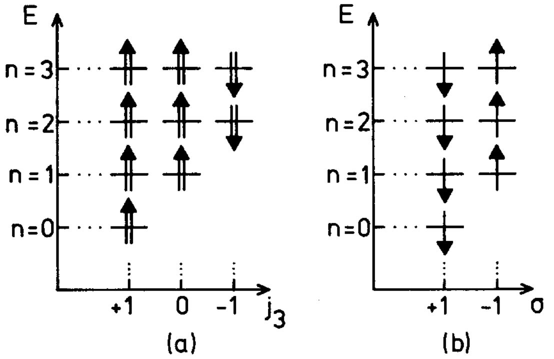


FIG. 1.(a) The parasuperspectrum of the $(p = 2)$ and $(\varepsilon = -1)$ theory where ⇑ (⇓) refers to parafermions (parabosons), $(+1, 0, -1)$ being the j_3 eigenvalues of the operator J_3. (b) The well-known spectrum of $(N = 2)$- and $(\varepsilon = -1)$-supersymmetric quantum mechanics where ↑ (↓) refers to fermions (bosons), $(+1, -1)$ being the σ eigenvalues of the Pauli matrix σ_3.

bakov–Spiridonov context.[12] When dealing with bosons and parafermions, the consistency between Eqs. (2.9) and (3.5) requires $h_0 = 1/2$. We thus get, from our basis (2.8) and Eq. (3.7), that the Rubakov–Spiridonov spectrum is

$$E_n^{(\mathrm{RS})} = n - \tfrac{1}{2}, \quad \forall n, \tag{3.15}$$

leading to the already quoted properties [see Sec. II A and Eq. (2.10) in particular].

As a final comment on this quantum mechanical approach, let us notice that the other choice in Eq. (3.1), i.e., the study of the parasupersymmetric Hamiltonian

$$H_{\mathrm{PSS}}^{(+)} = \tfrac{1}{2}\{a,a^\dagger\} + \tfrac{1}{2}[b^\dagger,b] \tag{3.16}$$

is evidently completely similar to the above developments. Correspondingly the parafermionic number operator has to be chosen as

$$N_{\mathrm{pf}}^{(+)} = \mathbb{1}_3 - J_3 = \begin{pmatrix} 0 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 2 \end{pmatrix} \tag{3.17}$$

and the *ad hoc* basis replacing the vectors (2.8) is now specified by $\{|\Psi_0\rangle, |\Psi_1\rangle, |\Psi_n\rangle, \ n\geqslant 2\}$ given by

$$\begin{aligned} |\Psi_0\rangle &= (0,0,|0\rangle)^T, \\ |\Psi_1\rangle &= (0,|0\rangle,0)^T + (0,0,|1\rangle)^T, \\ |\Psi_n\rangle &= \alpha_n'(0,|n-1\rangle,0)^T + \beta_n'(0,0,|n\rangle)^T \\ &\quad + \gamma_n'(0,0,|n-2\rangle)^T, \quad n\geqslant 2. \end{aligned} \tag{3.18}$$

The degeneracies are the same as in the $(\varepsilon = -1)$ context and the spectrum [Fig. 2(a)] has similar characteristics up to correspondences with the J_3 eigenvalues. In Fig. 2(b) we have also quoted the corresponding spectrum in the current supersymmetric context between bosons and fermions.

IV. PARASUPERSYMMETRIC COHERENT STATES

In this section we propose to construct two sets of parasupersymmetric *coherent* states (Secs. IV A and B), in accordance with the previous two contexts of relative para-Bose and para-Fermi sets, respectively.

As supersymmetry has to be included in our constructions, let us recall that the supersymmetric Hamiltonian has

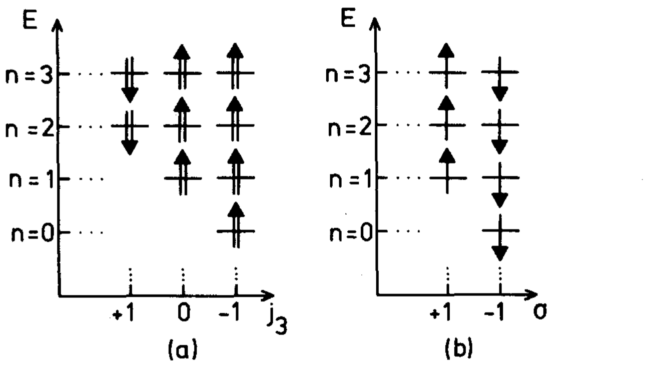


FIG. 2. (a) The parasuperspectrum of the $(p=2)$ and $(\varepsilon = +1)$ theory where ⇑ (⇓) refers to parafermions (parabosons). (b) The well-known spectrum of the $(N=2)$- and $(\varepsilon = +1)$-supersymmetric quantum mechanics where ↑ (↓) refers to fermions (bosons).

an always obvious *even* ($\mathscr{E}$) character while the supercharges are odd ($\mathscr{O}$). The Lie superalgebra[1] will then be characterized by the structure relations ensuring that

$$[\mathscr{E},\mathscr{E}]\to\mathscr{E}, \quad [\mathscr{O},\mathscr{E}]\to\mathscr{O}, \quad \{\mathscr{O},\mathscr{O}\}\to\mathscr{E}. \tag{4.1}$$

Moreover, it is interesting to point out, at the start of our discussion, that these even or odd characters are always guaranteed inside the above operators whatever the even or odd properties of the a- and b-annihilation operators are *except that* the last ones have to be of *opposite* parities. If a is even (as usual), b has to be odd (as usual) and this case will lead to the relative para-Bose set of trilinear relations given by Eqs. (2.13), according to the superstructure bilinear relations (4.1). If a is odd, b has to be even and this case will lead to the relative para-Fermi set given by Eqs. (2.13a)–(2.13d) *and* (2.18), once again in perfect agreement with Eqs. (4.1).

From a general point of view, the above considerations illustrate, in particular, that parastatistics,[2] an *older* proposal with respect to supersymmetry,[6] includes as well as supersymmetry the typical ideas leading to Lie superalgebras through the simultaneous superposition of commutators and anticommutators as a generalized Lie bracket.[1] These characteristics will be more effectively exploited in Sec. V.

As a last part of this section, we then study (Sec. IV C) the parasupercoherent states obtained in the relative para-Fermi context in order to determine their normalization factor and measure when $h_0 = 1$, i.e., when the $p=2$ parasupersymmetric theory is undertaken. We can thus analyze their impact on expectation values and on the uncertainty principle.

A. Para-Bose-supersymmetric coherent states

In connection with the Rubakov–Spiridonov contribution,[12] we have recently constructed[14] parasupercoherent states inside the boson–parafermion theory associated with the Hamiltonian (2.1).

Let us now consider the corresponding construction but in the parabosonic–parafermionic context subtended by the Hamiltonian $H_{\mathrm{PSS}} \equiv (2.14)$.

Due to the nice properties of the basis (2.8) in the parabosonic–parafermionic case, i.e., when the eigenvalues n refer to the para-Bose number operator $N_{\mathrm{pb}} \equiv (3.4)$, we can easily apply the following parasupersymmetric operator:

$$A = a\mathbb{1}_3 + (1/4)a^\dagger\{b^\dagger,b^\dagger\}. \tag{4.2}$$

It evidently has an *even* parity if as already mentioned inside the relative para-Bose set (2.13) we are considering $a,a^\dagger$ as even operators and $b,b^\dagger$ as odd ones. Then, according to the structure relations (4.1) and (2.13), we can show that

$$[A,H_{\mathrm{PSS}}] = A, \quad [A^\dagger,H_{\mathrm{PSS}}] = -A^\dagger, \tag{4.3}$$

which are the fundamental information for seeing this new operator A as an effective even parasupersymmetric annihilation operator. This can be corroborated by the following results:

$$A|\psi_0\rangle = 0, \quad A|\psi_n\rangle = |\psi_{n-1}\rangle, \quad \forall n\geqslant 1, \tag{4.4}$$

with explicit coefficients α_n, β_n, γ_n deduced from Eq. (2.8). These numbers depend on n:

$$\alpha_n = \frac{1}{\sqrt{(n-1)!}}, \quad \beta_n = \frac{1}{\sqrt{n!}}\left[1 - \frac{n(n-1)}{4}\right],$$
$$\gamma_n = \frac{1}{2\sqrt{(n-2)!}}, \quad \forall n \geqslant 2.$$

As usual in quantum mechanics,[19,4] we can then search for parasupercoherent states as eigenfunctions $|Z\rangle_{h_0}$ of this operator (4.2) by asking for

$$A|Z\rangle_{h_0} = z|Z\rangle_{h_0}, \tag{4.5}$$

where z is any complex number.

Let us mention that we want to maintain the h_0 dependence in our developments because, in a parallel way with Sharma *et al.*'s results[7] they are valid for arbitrary h_0's and because our contexts will fix $h_0 = 1$ in the parabosonic–parafermionic context or $h_0 = \frac{1}{2}$ in the bosonic–parafermionic one as already mentioned in Sec. III.

Indexing by h_0, the para-Bose number eigenstates defined by Eq. (3.4), we can develop our new parasupercoherent states $|Z\rangle_{h_0}$ in such a basis. Explicitly, we have to take care of the structure of the $|\psi_0\rangle$, $|\psi_1\rangle$, and $|\psi_n\rangle$ given in Eq. (2.8) *and* to distinguish between even and odd para-Bose number states in the summations due to the relations (3.5). We thus write

$$|Z\rangle_{h_0} = \sum_{l=0}^{\infty} a_{2l}\begin{pmatrix}|2l\rangle_{h_0}\\0\\0\end{pmatrix} + \sum_{l=0}^{\infty} a_{2l+1}\begin{pmatrix}|2l+1\rangle_{h_0}\\0\\0\end{pmatrix}$$
$$+ \sum_{l=0}^{\infty} b_{2l}\begin{pmatrix}0\\|2l\rangle_{h_0}\\0\end{pmatrix} + \sum_{l=0}^{\infty} b_{2l+1}\begin{pmatrix}0\\|2l+1\rangle_{h_0}\\0\end{pmatrix}$$
$$+ \sum_{l=0}^{\infty} c_{2l}\begin{pmatrix}0\\0\\|2l\rangle_{h_0}\end{pmatrix} + \sum_{l=0}^{\infty} c_{2l+1}\begin{pmatrix}0\\0\\|2l+1\rangle_{h_0}\end{pmatrix}. \tag{4.6}$$

Introducing the development (4.6) in Eq. (4.5), we get the following recurrence relations:

$$a_{2l}z = a_{2l+1}\sqrt{2(l+h_0)} + c_{2l-1}\sqrt{2l},$$
$$a_{2l+1}z = a_{2l+2}\sqrt{2(l+2)} + c_{2l}\sqrt{2(l+h_0)},$$
$$b_{2l}z = b_{2l+1}\sqrt{2(l+h_0)}, \quad c_{2l}z = c_{2l+1}\sqrt{2(l+h_0)},$$
$$b_{2l+1}z = b_{2l+2}\sqrt{2l+2}, \quad c_{2l+1}z = c_{2l+2}\sqrt{2l+2},$$

so that a relatively elaborate but systematic treatment shows that all the coefficients appearing in Eq. (4.6) can be determined in terms of the only three unknowns a_0, b_0, and c_0.

We effectively get for arbitrary l

$$a_{2l} = a_0 P_l z^{2l} - c_0 P_l 2l(l+h_0-1)z^{2l-2},$$
$$b_{2l} = b_0 P_l z^{2l}, \quad c_{2l} = c_0 P_l z^{2l}, \tag{4.7a}$$

and

$$a_{2l+1} = a_0 Q_l z^{2l+1} - c_0 Q_l 2l(l+h_0)z^{2l-1},$$
$$b_{2l+1} = b_0 Q_l z^{2l+1}, \quad c_{2l+1} = c_0 Q_l z^{2l+1}, \tag{4.7b}$$

where

$$P_l \equiv \frac{1}{2^l}\left[\frac{\Gamma(h_0)}{\Gamma(l+1)\Gamma(l+h_0)}\right]^{1/2}$$

and

$$Q_l \equiv \frac{1}{2^l}\left[\frac{\Gamma(h_0)}{2\Gamma(l+1)\Gamma(l+h_0+1)}\right]^{1/2}. \tag{4.7c}$$

Finally, by introducing the parabosonic coherent states[7] that are defined in our notations by

$$|z\rangle_{h_0} = \sum_{l=0}^{\infty} P_l z^{2l}|2l\rangle_{h_0} + \sum_{l=0}^{\infty} Q_l z^{2l+1}|2l+1\rangle_{h_0}, \tag{4.8}$$

we can write the parasupercoherent states on the compact form

$$|Z\rangle_{h_0} = a_0\begin{pmatrix}|z\rangle_{h_0}\\0\\0\end{pmatrix} + b_0\begin{pmatrix}0\\|z\rangle_{h_0}\\0\end{pmatrix} + c_0\begin{pmatrix}|*\rangle\\0\\|z\rangle_{h_0}\end{pmatrix}, \tag{4.9}$$

the ket $|*\rangle$ being given by

$$|*\rangle = -\tfrac{1}{2}|z''\rangle_{h_0} - (h_0 - \tfrac{1}{2})\left\{\sum_{l=0}^{\infty} 2lP_l z^{2l-2}|2l\rangle_{h_0} + \sum_{l=0}^{\infty} 2lQ_l z^{2l-1}|2l+1\rangle_{h_0}\right\}, \tag{4.10}$$

where

$$|z''\rangle_{h_0} = \frac{\partial^2}{\partial z^2}|z\rangle_{h_0}.$$

Let us only notice that the states (4.9) and (4.10) exactly reduce to the parasupercoherent states obtained[14] in the bosonic–parafermionic context *when* $h_0 = \frac{1}{2}$. It is easy to convince ourselves that, in this case, the states (4.8) reduce to the ordinary bosonic coherent states associated with the usual Heisenberg–Weyl context,[4] i.e.,

$$|z\rangle_{1/2} = \sum_{n=0}^{\infty} \frac{z^n}{(n!)^{1/2}}|n\rangle. \tag{4.11}$$

The parasupercoherent states (4.9) can then be further exploited (normalization factor and scalar product, completeness relation, etc.) but, due to the interest of the following case, we do not pursue such a study. The conclusion of this section is that the states (4.9) do exist and can be handled in the $h_0 = 1$ context for our $p = 2$-parasupersymmetric theory.

B. Para-Fermi-supersymmetric coherent states

We have already proposed[15] another parasupersymmetric annihilation operator—called B—inside the boson–parafermion theory associated with the Rubakov–Spiridonov Hamiltonian (2.1). We have also constructed the corresponding parasupercoherent states.[15] In such a context, these states appear as the closest parasupersymmetric states to the classical ones.

Here, let us extend these considerations but in the parabosonic–parafermionic case subtended by the Hamiltonian $H_{PSS} \equiv (2.14)$ and dealing with the basis (2.8) where the eigenvalues n refer to the number operator $N_{pb} \equiv (3.4)$. We consider the parasupersymmetric operator

$$B = (1/\sqrt{2})\left[b^{\dagger} + \tfrac{1}{4}b\{a,a\}\right], \tag{4.12}$$

which has an *even* character iff we are considering $a,a^\dagger$ as odd operators and $b,b^\dagger$ as even ones, a set of properties dealing with the relative para-Fermi structure relations (2.13a)–(2.13d) and (2.18). Notice that B is the Hermitian conjugate of $A\equiv(4.2)$ *but* with the simultaneous substitution $a\leftrightarrow b$.

Once again the operator B is such that

$$[B,H_{\rm PSS}]=B,\quad [B^\dagger,H_{\rm PSS}]=-B^\dagger, \tag{4.13}$$

and

$$B|\psi_0\rangle=0,\quad B|\psi_n\rangle=|\psi_{n-1}\rangle,\quad \forall n>0. \tag{4.14}$$

It is thus a new effective, even, parasupersymmetric annihilation operator that can lead to new parasupercoherent states hereafter denoted $|Y\rangle_{h_0}$ and defined by

$$B|Y\rangle_{h_0}=y|Y\rangle_{h_0}, \tag{4.15}$$

where y is any complex number.

Let us develop the states $|Y\rangle_{h_0}$ in terms of the para-Bose number eigenstates defined by Eq. (3.4). We evidently get a formula analogous with (4.6) but in terms of coefficients r_{2l},r_{2l+1}, s_{2l},s_{2l+1}, and t_{2l},t_{2l+1} in correspondence with the respective a_{2l},a_{2l+1}, b_{2l},b_{2l+1}, and c_{2l},c_{2l+1}.

Here we get the information for arbitrary l

$$r_{2l}=r_0P_ly^{2l},\quad s_{2l}=r_0P_ly^{2l+1},\quad t_{2l}=(r_0/2)P_ly^{2l+2} \tag{4.16a}$$

$$r_{2l+1}=r_1R_ly^{2l},\quad s_{2l+1}=r_1R_ly^{2l+1},\quad t_{2l+1}=(r_1/2)R_ly^{2l+2} \tag{4.16b}$$

where P_l is still given by Eq. (4.7c) and

$$R_l\equiv(1/2^l)\left[\frac{\Gamma(h_0+1)}{\Gamma(l+1)\Gamma(l+h_0+1)}\right]^{1/2}. \tag{4.16c}$$

Now, it can be shown that, in terms of the parabosonic coherent states (4.8), our parasupercoherent states have the final form

$$|Y\rangle_{h_0}=r_0\begin{pmatrix}|y\rangle_{h_0}\\ y|y\rangle_{h_0}\\ (y^2/2)|y\rangle_{h_0}\end{pmatrix} \tag{4.17}$$

after we have required that $r_1=r_0\,y(2h_0)^{-1/2}$, a restriction imposed for the inclusion of the nonparastatistical context.

These states $|Y\rangle_{h_0}$ have to be normalized and have to satisfy the (over-)completeness relation.[5] By noticing that[7]

$$_{h_0}\langle y|y\rangle_{h_0}=\Gamma(h_0)(|y|^2/2)^{1-h_0}\times[I_{h_0-1}(|y|^2)+I_{h_0}(|y|^2)], \tag{4.18}$$

the requirement[5]

$$_{h_0}\langle Y|Y\rangle_{h_0}=1 \tag{4.19}$$

does fix the normalization factor r_0. We obtain

$$|r_0|^2=\frac{(|y|^2/2)^{h_0-1}}{\Gamma(h_0)(1+|y|^2+\frac14|y|^4)}\times[I_{h_0-1}(|y|^2)+I_{h_0}(|y|^2)]^{-1}. \tag{4.20}$$

Here, in Eqs. (4.18) and (4.20), we are dealing with the modified Bessel functions of the kth order[23] given by

$$I_k(z)=\sum_{m=0}^{\infty}\frac{(\frac12 z)^{k+2m}}{m!\Gamma(k+m+1)}. \tag{4.21}$$

Moreover, the resolution of the identity operator,[5] i.e.,

$$\int|Y\rangle_{h_0\,h_0}\langle Y|\mu_{h_0}(|y|^2)d^2y=I, \tag{4.22}$$

requires the existence of a measure when the integration is realized over the whole complex y plane. A lengthy and elaborate calculation that is parallel to the one performed for para-Bose coherent states[7] leads to the result

EL4

$$\mu_{h_0}(|y|^2)=(1/2\pi)(|y|^2/2)^{1-h_0}\times\Gamma(h_0)[I_{h_0-1}(|y|^2)+I_{h_0}(|y|^2)]\times\int_{-\infty}^{+\infty}M_{h_0}(x)e^{-i|y|^2x}dx, \tag{4.23}$$

where

$$M_{h_0}(x)=\frac{1}{\pi}\sum_{n=0}^{\infty}\frac{(2ix)^n}{n!}\times\frac{\Gamma([n/2]+1)\Gamma([(n+1)/2]+h_0)}{\Gamma(h_0)}, \tag{4.24}$$

the notation $[k]$ standing for the largest integer smaller than or equal to k. It has to be noticed that the series contained in Eq. (4.24) converges only for $|x|<1$. We have thus shown that our parasupercoherent states (4.17) are normalized and are (over)-complete.

An effective test of the above developments is the particular context $h_0=\frac12$ corresponding to the case of the ordinary harmonic oscillator in the bosonic sector and, correspondingly, to the bosonic–parafermionic case in the supersymmetric theory subtended by the Hamiltonian (2.1). If we put $h_0=\frac12$ in Eq. (4.20), we immediately get

$$|r_0|^2=(1/\sqrt{\pi})(|y|^2/2)^{-1/2}(1+|y|^2+\tfrac14|y|^4)^{-1}\times[I_{-1/2}(|y|^2)+I_{1/2}(|y|^2)]^{-1}. \tag{4.25}$$

By noticing that[23]

$$\sqrt{\frac{\pi}{2x}}I_{1/2}(x)=\frac{1}{x}\sinh x$$

and

$$\sqrt{\frac{\pi}{2x}}I_{-1/2}(x)=\frac{1}{x}\cosh x,$$

we get

$$I_{-1/2}(|y|^2)+I_{1/2}(|y|^2)=\sqrt{2/\pi|y|^2}\exp(|y|^2)$$

and finally

$$|r_0|^2=(1+|y|^2+\tfrac14|y|^4)^{-1}\exp(-|y|^2). \tag{4.26}$$

This result is identical with the normalization factor already obtained[15] in the Rubakov–Spiridonov developments for coherent states in parasupersymmetric quantum mechanics. Moreover, if we put $h_0=\frac12$ in Eq. (4.24), we obtain, by considering separate even and odd powers in the sum, that

$$M_{1/2}(x)=(1/\pi)(1-ix)^{-1}$$

and, consequently, the relation (4.23) becomes

$$\mu_{1/2}(|y|^2) = \frac{e^{|y|^2}}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{-i|y|^2 x}}{\pi(1-ix)}\,dx. \quad (4.27)$$

Now, using the information,[24]

$$\int_{-\infty}^{+\infty} (a-ix)^{-\nu} e^{-iyx}\,dx = 2\pi \frac{y^{\nu-1}e^{-ay}}{\Gamma(\nu)},$$

valid when $y>0$, $\mathrm{Re}\,a>0$, and $\mathrm{Re}\,\nu>0$, we can evidently choose $a=\nu=1$, so that

$$\int_{-\infty}^{+\infty} (1-ix)^{-1} e^{-i|y|^2 x}\,dx = 2\pi e^{-|y|^2}.$$

Equation (4.27) finally gives

$$\mu_{1/2}(|y|^2) = 1/\pi, \quad (4.28)$$

the well-known[4,7] measure for the harmonic oscillator leading to the usual resolution of the identity operator, i.e.,

$$\frac{1}{\pi}\int |y|\rangle_{1/2\ 1/2}\langle Y|d^2y = I.$$

C. Parasupercoherent states and the uncertainty principle

Having constructed two sets $\{|Z\rangle_{h_0}\}$ and $\{|Y\rangle_{h_0}\}$ of states, respectively, given by Eqs. (4.9) and (4.17), we have to carry on their properties in our $p=2$-parasupersymmetric theory corresponding to the specific value $h_0=1$.

Let us here only consider the second set due to its nice properties within the $h_0=1/2$ context as well as due to the simpler form of the states (4.17).

If, for $h_0=1$, the normalization factor is readily obtained from Eq. (4.20) as

$$|r_0|^2 = (1+|y|^2+(1/4)|y|^4)^{-1}[I_0(|y|^2) + I_1(|y|^2)]^{-1}, \quad (4.29)$$

it is not an easy task to give the final form of the measure issued from Eqs. (4.23) and (4.24), due to the appearance of hypergeometric, Whittaker and Bessel functions[23] as well as of relatively complicated integrals.[24,25] The first problem consists in finding the expression of $M_1(x)$ issued from Eq. (4.24) and the second one is the integration contained in Eq. (4.23). Let us just summarize the main steps of these calculations.

By decomposing the summation in even and odd powers, we can show that

$$M_1(x) = \frac{1}{\pi}\left[{}_2F_1(1,1;\tfrac{1}{2};-x^2) + 2ix\,{}_2F_1\left(1,2;\frac{3}{2};-x^2\right)\right], \quad (4.30)$$

where the hypergeometric functions[23] ${}_2F_1(a,b;c;z)$ converge for $|z|<1$ or for $|z|=1$ (if $\mathrm{Re}\,(a+b-c)<0$). Now, by taking the following properties[25] into account:

$$2\frac{\Gamma(\alpha+\beta+1/2)\Gamma(1/2)}{\Gamma(\alpha+1/2)\Gamma(\beta+1/2)}\,{}_2F_1\left(\alpha,\beta;\frac{1}{2};z\right) = {}_2F_1\left(2\alpha,2\beta;\alpha+\beta+\frac{1}{2};\frac{1-\sqrt{z}}{2}\right) + {}_2F_1\left(2\alpha,2\beta;\alpha+\beta+\frac{1}{2};\frac{1+\sqrt{z}}{2}\right) \quad (4.31a)$$

and

$$\frac{(\alpha-1/2)(\beta-1/2)\Gamma(\alpha+\beta+1/2)\Gamma(1/2)}{(\alpha+\beta-1/2)\Gamma(\alpha+1/2)\Gamma(\beta+1/2)} \times\sqrt{z}\,{}_2F_1\left(\alpha,\beta;\frac{3}{2};z\right) = {}_2F_1\left(2\alpha-1,2\beta-1;\alpha+\beta-\frac{1}{2};\frac{1+\sqrt{z}}{2}\right) - {}_2F_1\left(2\alpha-1,2\beta-1;\alpha+\beta-\frac{1}{2};\frac{1-\sqrt{z}}{2}\right), \quad (4.31b)$$

we, respectively, get with the change of variable $x=ip$:

$$6\,{}_2F_1\left(1,1;\frac{1}{2};p^2\right) = {}_2F_1\left(2,2;\frac{5}{2};\frac{1-p}{2}\right) + {}_2F_1\left(2,2;\frac{5}{2};\frac{1+p}{2}\right) \quad (4.32a)$$

and

$$\frac{3}{2}p\,{}_2F_1\left(1,2;\frac{3}{2};p^2\right) = {}_2F_1\left(1,3;\frac{5}{2};\frac{1+p}{2}\right) - {}_2F_1\left(1,3;\frac{5}{2};\frac{1-p}{2}\right). \quad (4.32b)$$

With these relations, the integral contained in Eq. (4.23) now reads

$$-\frac{i}{3\pi}\int_{-i\infty}^{+i\infty} dp\, e^{p|y|^2}\left[\frac{1}{2}\,{}_2F_1\left(2,2;\frac{5}{2};\frac{1-p}{2}\right) + \frac{1}{2}\,{}_2F_1\left(2,2;\frac{5}{2};\frac{1+p}{2}\right) - 4\,{}_2F_1\left(1,3;\frac{5}{2};\frac{1+p}{2}\right) + 4\,{}_2F_1\left(1,3;\frac{5}{2};\frac{1-p}{2}\right)\right].$$

By exploiting inverse Laplace transforms given by[24]

$$\frac{1}{2i\pi}\int_{-i\infty}^{+i\infty} dp\, e^{p|y|^2}\,{}_2F_1\left(\alpha,\beta;\gamma;\frac{1}{2}-\frac{p}{\lambda}\right) = \frac{\lambda\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)}(\lambda|y|^2)^{(1/2)(\alpha+\beta-3)} \times W_{(1/2)(\alpha+\beta+1)-\gamma,(1/2)(\alpha-\beta)}(\lambda|y|^2)$$

and valid only for $\mathrm{Re}\,\alpha>0$ and $\mathrm{Re}\,\beta>0$, we get specific Whittaker functions[23] in the results. This property applied to the four above integrals lead to the final evaluation of the integral contained in Eq. (4.23), i.e.,

$$\sqrt{\pi}(2|y|^2)^{1/2}[\tfrac{1}{2}W_{0,0}(2|y|^2) + 2W_{0,-1}(2|y|^2)] + \sqrt{\pi}(-2|y|^2)^{1/2}[2W_{0,-1}(-2|y|^2) - \tfrac{1}{2}W_{0,0}(-2|y|^2)].$$

Finally, this last expression can be put on the following form when relating[23] Whittaker functions and modified Bessel functions $K_\nu(z)$:

$$|y|^2[K_0(|y|^2) + K_0(-|y|^2) + 4(K_1(|y|^2) - K_1(-|y|^2))].$$

The final expression of the measure (4.23) is thus, for $h_0=1$,

$$\mu_1(|y|^2) = (|y|^2/2\pi)[I_0(|y|^2) + I_1(|y|^2)] \times [K_0(|y|^2) + K_0(-|y|^2) + 4(K_1(|y|^2) - K_1(-|y|^2))] \quad (4.33)$$

ensuring the (over)-completeness of the parasupercoherent states $|Y\rangle_1$.

As a last comment on the parasupercoherent states $|Y\rangle_{h_0}$, let us end this subsection by studying their impact on the uncertainty principle through the archetypal example of the Heisenberg relation on position and momentum. Remembering the relation between the x and p operators with the a and $a^\dagger$ ones, we evidently have for arbitrary h_0

$$x|Y\rangle_{h_0} = \frac{r_0}{\sqrt{2}}(a + a^\dagger)\begin{pmatrix} |y\rangle_{h_0} \\ y|y\rangle_{h_0} \\ (y^2/2)|y\rangle_{h_0} \end{pmatrix} \quad (4.34)$$

and

$$p|Y\rangle_{h_0} = \frac{ir_0}{\sqrt{2}}(a^\dagger - a)\begin{pmatrix} |y\rangle_{h_0} \\ y|y\rangle_{h_0} \\ (y^2/2)|y\rangle_{h_0} \end{pmatrix}. \quad (4.35)$$

By using the parabosonic coherent states (4.8) ($z \leftrightarrow y$) and the properties (3.5), it is relatively easy, on the one hand, to show that the expectation values $\langle x\rangle$ and $\langle p\rangle$ are independent of h_0. We effectively obtain

$$\langle x\rangle = \sqrt{2}\,\mathrm{Re}\,y \quad \text{and} \quad \langle p\rangle = \sqrt{2}\,\mathrm{Im}\,y. \quad (4.36)$$

On the other hand, the expectation values of x^2 and p^2 are, respectively, obtained on the forms

$$\langle x^2\rangle_{h_0} = 2\,\mathrm{Re}^2\,y + \frac{h_0 I_{h_0-1}(|y|^2) + (1-h_0)I_{h_0}(|y|^2)}{I_{h_0-1}(|y|^2) + I_{h_0}(|y|^2)} \quad (4.37a)$$

and

$$\langle p^2\rangle_{h_0} = 2\,\mathrm{Im}^2\,y + \frac{h_0 I_{h_0-1}(|y|^2) + (1-h_0)I_{h_0}(|y|^2)}{I_{h_0-1}(|y|^2) + I_{h_0}(|y|^2)}. \quad (4.37b)$$

We thus get as in the para-Bose context[7]

$$\begin{aligned}(\Delta x)^2_{h_0} &= \langle x^2\rangle_{h_0} - \langle x\rangle^2 \\ &= \frac{h_0 I_{h_0-1}(|y|^2) + (1-h_0)I_{h_0}(|y|^2)}{I_{h_0-1}(|y|^2) + I_{h_0}(|y|^2)} \\ &= \langle p^2\rangle_{h_0} - \langle p\rangle^2 \\ &= (\Delta p)^2_{h_0}. \end{aligned} \quad (4.38)$$

We immediately deduce that, in the $h_0 = 1/2$ context, we get

$$(\Delta x)^2_{1/2} = (\Delta p)^2_{1/2} = \tfrac{1}{2}, \quad (4.39)$$

so that the corresponding states are the closest states to the classical ones as expected.[15] In the $h_0 = 1$ context, we get

$$(\Delta x)^2_1 = (\Delta p)^2_1 = \frac{I_0(|y|^2)}{[I_0(|y|^2) + I_1(|y|^2)]}. \quad (4.40)$$

Here, due to the fact that $I_1 \geqslant 0$, we always have

$$I_0/(I_0 + I_1) \leqslant 1$$

and

$$0 \leqslant (\Delta x)_1(\Delta p)_1 \leqslant 1. \quad (4.41)$$

Let us also notice that *if* $|y| \ll 1$, $I_0 \sim 1$, and $I_1 \sim \frac{1}{2}|y|^2$, so that

$$(\Delta x)^2_1(\Delta p)^2_1 \sim 1/(1 + \tfrac{1}{2}|y|^2)^2 \sim 1, \quad (4.42)$$

which is a result already obtained (and discussed) on some supercoherent states.[9]

Such results can directly be related to the discussion on para-Bose coherent states[7] and lead to the same conclusions in the *supersymmetric* context. Let us only recall here that the commutator $[x,p]$ is a c number iff $h_0 = 1/2$, while it is not the case when $h_0 = 1$:

$$[x,p] = i[a,a^\dagger]. \quad (4.43)$$

V. PARASTATISTICS, SUPERSYMMETRY, AND LIE SUPERALGEBRAS

It has already been pointed out[3,21,13] that the structure relations among parafields have to deal with (simple) Lie superalgebras.[1]

If we remember that, for each pair of parafields, there exist (besides the straight commutation relations for bosons and the straight anticommutation relations for fermions) only two relative para-Bose and relative para-Fermi sets (see Sec. II), we have to mention that Palev[21] has noticed that the creation and annihilation operators of parafields generate the simple Lie superalgebra $B(n/m)$. If we limit ourselves to *only one pair* of such parafields, this superalgebra is simply $B(1,1)$, this Kac notation[1] corresponding to the orthosymplectic Lie superalgebra osp(3/2). These considerations[21] were thus associated with the relative para-Fermi set given here in Eqs. (2.13a)–(2.13d) *and* (2.18).

More recently, Biswas and Soni[13] have claimed to propose new mixed relations leading to what they called "generalized parastatistics." In fact, as it is easy to convince oneself, their proposal coincides with the above relative para-Bose set which reduces to our Eqs. (2.13) for a pair of parafields. Nevertheless, Biswas and Soni[13] have then shown the interesting result that to such a relative para-Bose set is associated the simple Lie superalgebra C(2) if we limit ourselves to only one pair of parafields. This superalgebra C(2) is the Kac notation[1] corresponding to the orthosymplectic Lie superalgebra osp(2/2).

Both of the above superalgebras osp(3/2) and osp(2/2) have already been exploited[26–31] in supersymmetric quantum mechanics[17] and more particularly in the study of the supersymmetric harmonic oscillator (in one spatial dimension leading to pairs of bosonic and fermionic degrees of freedom). Moreover, they have recently been related one to the other through the one-to-one correspondence[32]

$$\mathrm{osp}(3/2) \leftrightarrow \mathrm{osp}(2/2) \,\square\, \mathrm{sh}(2/2), \quad (5.1)$$

where the second superalgebra appearing in the semidirect sum is the Heisenberg–Weyl superalgebra.[29] This correspondence (5.1) is associated with a "character reversal" phenomenon[32] recently explained[33,34] from a detailed analysis of the even and odd root systems of the superalgebras belonging to the following specific chain:

$$\text{osp}(3/4) \supset \text{osp}(3/2) \supset \text{osp}(2/2). \tag{5.2}$$

It is possible[33] to show that osp(3/4) also contains the semidirect sum (5.1), so that the correspondence can be understood through the interchange of the even and odd properties of the nontrivial generators of the superalgebra sh(2/2).

If all these results refer to invariance superalgebras of the (1-dimensional) *supersymmetric* harmonic oscillator as well as to its constants of motion, a natural question then arises in the *parasupersymmetric* context we are dealing with in this article, a question strengthened by the appearance of the same superalgebras osp(3/2) and osp(2/2) but in connection with the relative para-Fermi and para-Bose sets of trilinear structure relations. What are the invariance superalgebras and the associated constants of motion in the $p = 2$-parasupersymmetric theory developed in the preceding sections?

The answer is given in Secs. V A and V B by new explicit realizations of the respective superalgebras and, in the Sec. V C, by a discussion of the results as well as of their common and distinct features.

A. Relative para-Bose supersymmetry and the superalgebra osp(2/2)

For the generalized version of Jacobi's identity quoted by Greenberg and Messiah,[16] i.e.,

$$[[A,B]_\varepsilon,C] + [[C,A]_\eta,B]_{-\eta\varepsilon} + \eta\varepsilon[[B,C],A]_{-\eta\epsilon} = 0, \tag{5.3}$$

where $\varepsilon = \pm 1$ and $\eta = +1$ for the para-Bose context, we propose to rewrite

$$\varepsilon = (-1)^{\pi(I)(\pi(J)+\pi(K))}, \quad \eta = (-1)^{\pi(K)(\pi(I)+\pi(J))} \tag{5.4}$$

when they are expressed in terms of even $[\pi(I) = 0]$ or odd $[\pi(I) = 1]$ parities according to the notations

$$c_I \equiv (a,b):\ c_{I_{|\pi(I)=0}} = a \text{ (even)} \tag{5.5a}$$

and

$$c_{I_{|\pi(I)=1}} = b \text{ (odd)}. \tag{5.5b}$$

We can thus summarize all the structure relations (2.13) by:[13]

$$\begin{aligned}
[c_I,\langle c_J,c_K\rangle]_\pm &= 0,\\
[c_I,\langle c_J^\dagger,c_K\rangle]_\pm &= 2\delta_{IJ}c_K,\\
[c_I^\dagger,\langle c_J,c_K\rangle]_\pm &= -2(-1)^{\pi(I)\pi(J)}\delta_{IJ}c_K\\
&\quad \times -2(-1)^{\pi(K)(\pi(I)+\pi(J))}\delta_{IK}c_J,
\end{aligned} \tag{5.6}$$

where, for self-consistency, we recall that

$$[c_I,c_J]_\pm = c_Ic_J - (-1)^{\pi(I)\pi(J)}c_Jc_I$$

and

$$\langle c_I,c_J\rangle = c_Ic_J + (-1)^{\pi(I)\pi(J)}c_Jc_I. \tag{5.7}$$

Let us now realize osp(2/2) in direct correspondence with recent developments[29] on the supersymmetric harmonic oscillator but here in the parasupersymmetric theory. We get

$$\begin{aligned}
H_{\text{pb}} &= (1/2)\{a^\dagger,a\},\ C_+ = (i/2)e^{-2it}\{a^\dagger,a^\dagger\},\\
C_- &= -(i/2)e^{2it}\{a,a\},
\end{aligned} \tag{5.8}$$

and

$$H_{\text{pf}} = (1/2)[b^\dagger,b], \tag{5.9}$$

as the four even generators while the four odd supercharges read

$$Q_1 = (1/2)\{a,b\}, \quad Q_1^\dagger = (1/2)\{b^\dagger,a^\dagger\}, \tag{5.10}$$

$$S_1 = (1/2)e^{-2it}\{a^\dagger,b\}, \quad S_1^\dagger = (1/2)e^{2it}\{b^\dagger,a\}, \tag{5.11}$$

according to the choices (2.15). The superstructure relations issued from Eq. (4.1) with the generators (5.8)–(5.11) give immediately the superalgebra osp(2/2).

It is easy to convince oneself that, within the $p = 2$ theory developed in Sec. II B 1 and characterized by the parasupersymmetric Hamiltonian $H_{\text{PSS}} \equiv (2.14)$, the above eight generators are conserved and are associated with constants of motion as usual in the Hamiltonian formalism.

B. Relative para-Fermi supersymmetry and the superalgebra osp(3/2)

In correspondence with the generalized Jacobi identity (5.3), we need here $\varepsilon = \pm 1$, but $\eta = -1$ for the para-Fermi context.[16] The validity of the formulas (5.4) is once again ensured but when

$$c_I \equiv (a,b):\ c_{I_{|\pi(I)=0}} = b \text{ (even)} \tag{5.12a}$$

and

$$c_{I_{|\pi(I)=1}} = a \text{ (odd)}. \tag{5.12b}$$

In this case, all the structure relations (2.13a)–(2.13d) and (2.18) are summarized by

$$\begin{aligned}
[c_I,[c_J,c_K]_\pm]_\pm &= 0,\\
[c_I,[c_J^\dagger,c_K]_\pm]_\pm &= 2\delta_{IJ}c_K,\\
[c_I^\dagger,[c_J,c_K]_\pm]_\pm &= 2(-1)^{\pi(I)\pi(J)}\delta_{IJ}c_K\\
&\quad - 2(-1)^{\pi(K)(\pi(I)+\pi(J))}\delta_{IK}c_J.
\end{aligned} \tag{5.13}$$

We thus construct the 12 following generators where the 6 even ones are

$$\begin{aligned}
H_{\text{pb}} &= \tfrac{1}{2}\{a^\dagger,a\}, \quad C_+ = (i/2)e^{-2it}\{a^\dagger,a^\dagger\},\\
C_- &= -(i/2)e^{2it}\{a,a\},
\end{aligned} \tag{5.14}$$

$$H_{\text{pf}} = \tfrac{1}{2}[b^\dagger,b], \quad T_+ = e^{it}b^\dagger, \quad T_- = e^{-it}b, \tag{5.15}$$

while the 6 odd ones read

$$Q_2 = \tfrac{1}{2}[a,b], \quad Q_2^\dagger = \tfrac{1}{2}[b^\dagger,a^\dagger], \tag{5.16}$$

$$S_2 = \tfrac{1}{2}e^{-2it}[a^\dagger,b], \quad S_2^\dagger = \tfrac{1}{2}e^{2it}[b^\dagger,a], \tag{5.17}$$

$$P_+ = i\sqrt{2}e^{-it}a^\dagger, \quad P_- = -i\sqrt{2}e^{it}a, \tag{5.18}$$

according to the choices (2.19). Let us mention that Eqs. (5.14) lead to the so(2.1) content [as well as Eqs. (5.8)], while Eqs. (5.15) give rise to the so(3) content.[26] The whole set of superstructure relations leads here to the superalgebra osp(3/2) containing osp(2/2) $\equiv \{H_{\text{pb}},C_\pm,H_{\text{pf}},Q_2,Q_2^\dagger,S_2,S_2^\dagger\}$ as a subsuperalgebra, containing itself the subsuperalgebra osp(1/2) generated by

$$K_{\pm} \equiv \tfrac{1}{2}C_{\pm}, \quad K_0 = \tfrac{1}{2}H_{pb}$$

and

$$F_+ \equiv \tfrac{1}{2}(S_2 + iQ_2^\dagger), \; F_- = \tfrac{1}{2}(S_2^\dagger - iQ_2).$$

The superalgebra osp(3/2) appears as an invariance superalgebra of the $p = 2$ parasupertheory for a pair of parafields in the relative para-Fermi context, the above 12 generators being conserved and associated with constants of motion.

C. Discussion

Having obtained *different* information in the para-Bose and para-Fermi contexts, we thus lose the correspondence (5.1) in this parasupersymmetric theory. The reason is clear when analyzed in comparison with the nonparastatistical point of view and the proof already mentioned.[33–34] The main point is that, in the parasupertheory, we know that the commutator $[a, a^\dagger]$ and the anticommutator $\{b, b^\dagger\}$ are not the identity (as it is the case in the supercontext). Then, the Heisenberg superalgebra sh(2/2) which contains the identity operator (due to its central extension) has no meaning in the parasupercase. If we define two even (say $P_\pm$) and two odd (say $T_\pm$) generators in terms of single powers of a, $a^\dagger$, b, or $b^\dagger$, the superalgebra corresponding to osp(2/2)□$\{P_\pm, T_\pm\}$ becomes infinite in the para-Bose case, while, in the supercontext, it closes and leads to the semidirect sum given in (5.1). Thus each relative paracontext has its own properties essentially characterized by typical parities as given in Eqs. (5.5) and (5.12), and different sets of constants of motion in particular.

Note added in proof: We have just proposed another nonequivalent approach to parasupersymmetric quantum mechanics [see J. Beckers and N. Debergh, Nucl. Phys. B (to be published)].

[1] V. G. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); Lecture Notes Math. **676**, 587 (1978).
[2] H. S. Green, Phys. Rev. **90**, 270 (1953).
[3] Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (University of Tokyo, Tokyo, 1982).
[4] R. J. Glauber, Phys. Rev. **131**, 2766 (1963); J. R. Klauder, J. Math. Phys. **4**, 1055 (1963).
[5] J. R. Klauder and B. S. Skagerstam, *Coherent States—Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985).
[6] P. Fayet and S. Ferrara, Phys. Rep. C **32**, 250 (1977); *Supersymmetry in Physics* (North-Holland, Amsterdam, 1985).
[7] J. K. Sharma, C. L. Mehta, and E. C. G. Sudarshan, J. Math. Phys. **19**, 2089 (1978).
[8] J. K. Sharma, C. L. Mehta, and E. C. G. Sudarshan, J. Math. Phys. **22**, 78 (1981).
[9] C. Aragone and F. Zypman, J. Phys. A: Math. Gen. **19**, 2267 (1986).
[10] M. Gunaydin, J. Math. Phys. **29**, 275 (1988).
[11] A. B. Balantekin, H. A. Schmitt, and B. R. Barrett, J. Math. Phys. **29**, 1634 (1988).
[12] V. A. Rubakov and V. P. Spiridonov, Mod. Phys. Lett. A **3**, 1337 (1988).
[13] S. N. Biswas and S. K. Soni, J. Math. Phys. **29**, 16 (1988).
[14] J. Beckers and N. Debergh, Mod. Phys. Lett. A **4**, 1209 (1989).
[15] J. Beckers and N. Debergh, Mod. Phys. Lett. A **4**, 2289 (1989).
[16] O. W. Greenberg and A. M. L. Messiah, Phys. Rev. B **138**, 1155 (1965).
[17] E. Witten, Nucl. Phys. B **188**, 513 (1981); B **202**, 253 (1982).
[18] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Mécanique Quantique* (Hermann, Paris, 1977), Vol. 1.
[19] R. Shankar, *Principles of Quantum Mechanics* (Plenum, New York and London, 1980).
[20] E. P. Wigner, Phys. Rev. **77**, 711 (1950).
[21] T. D. Palev, J. Math. Phys. **23**, 1100 (1982).
[22] D. Lancaster, Nuovo Cimento. A **79**, 28 (1984).
[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables* (National Bureau of Standards Applied Mathematics, Washington, DC, 1965), 4th ed.
[24] H. Bateman, *Tables of Integral Transforms*, edited by A. Erdelyi (MacGraw-Hill, New York, 1954), Vol. 1.
[25] I. S. Gradshteyn and J. M. Ryzhik *Table of Integrals, Series and Products* (Academic, New York, 1965), 4th ed.
[26] J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).
[27] M. J. Englefield, J. Phys. A: Math. Gen. **21**, 1309 (1988).
[28] M. de Crombrugghe and V. Rittenberg, Ann. Phys. (N.Y.) **151**, 199 (1983).
[29] J. Beckers and V. Hussin, Phys. Lett. A **118**, 319 (1986).
[30] J. Beckers, D. Dehin, and V. Hussin, J. Phys. A: Math. Gen. **20**, 1137 (1987).
[31] J. Beckers, D. Dehin, and V. Hussin, J. Phys. A: Math. Gen. **21**, 651 (1987).
[32] J. Beckers, D. Dehin, and V. Hussin, J. Math. Phys. **29**, 1705 (1988).
[33] J. Beckers and J. F. Cornwell, J. Phys. A: Math. Gen. **22**, 925 (1989).
[34] J. Beckers and J. F. Cornwell, J. Math. Phys. **30**, 1655 (1989).

Classification of all star and grade star irreps of gl(n|1)

M. D. Gould and R. B. Zhang
Department of Mathematics, University of Queensland, St. Lucia 4067, Australia

(Received 29 August 1989; accepted for publication 24 January 1990)

A method for inducing nondegenerate forms on irreducible gl$(m|n)$ modules that implies some general results on star and grade star representations is investigated. These results are applied to obtain a complete classification, in terms of highest weights, of the irreducible star and grade star representations of gl$(n|1)$. It is demonstrated that while gl$(n|1)$ admits a large class of star representations the irreducible grade star representations are comparatively rare. Moreover, for $n \neq 2$ all grade star irreducible representations are also star representations and, for $n > 2$, are atypical. The superalgebra gl$(2|1)$ proves to be a special case and admits a two-parameter family of four-dimensional typical grade star irreducible representations that are not star representations. In particular, typical grade star irreducible representations of gl$(n|1)$ exist only for $n = 1,2$.

I. INTRODUCTION

The theory of Lie superalgebras and their representations plays a fundamental role in the understanding and exploitation of supersymmetry in physical systems. The concept of supersymmetry first arose in elementary particle physics[1] and has since been applied in a variety of other areas including nuclear physics[2] and condensed matter physics.[3–5] A comprehensive review of Lie superalgebra representation theory and its various physical applications is provided in Kostelecky and Campbell.[6]

The representation theory of the simple basic classical Lie superalgebras was first investigated in the definitive work of Kac,[7,8] who introduced the now familiar categorization of finite-dimensional *irreducible representations* (irreps) into typical and atypical types. Typical irreps, first classified by Kac,[8] have many properties in common with the finite-dimensional irreps of simple Lie algebras and in particular are given explicitly by an induced module construction[8] allowing for a straightforward determination of their dimensions and characters. By contrast, the structure of atypical irreps is far from well understood and has required the introduction of new techniques such as supertableaux methods,[9–13] those based on shift operators and weight space techniques,[14–16] and those arising from a modification of the Kac induced module construction[17] for atypicals. We also mentioned recent work[18] on the calculation of the characters of the atypical irreps.

A physically important problem that has thus far received comparatively little attention is the classification of all * and grade * irreps for a basic classical simple Lie superalgebra. Star and grade star representations were first introduced by Scheunert *et al.*[19] who demonstrated that all simple basic classical Lie superalgebras admit at most two types of * irreps and two types of grade * irreps: These representations are natural generalizations of Hermitian representations for simple Lie algebras and as such are most likely to be of direct physical interest. However, it should be noted that the classification of an irrep according to whether it is of * or grade * type (or possibly both or neither of these) is apparently unrelated to its typicality: Indeed, both typical and atypical * (and grade *) irreps exist, as well as both typical and atypical irreps which are neither * nor grade *. Therefore, it is evident that the determination of * and grade * irreps for a basic classical Lie superalgebra is going to yield an entirely new classification scheme for the irreps.

The * and grade * irreps were explicitly constructed for the Lie superalgebras osp(2|1) and sl(2|1) by Scheunert *et al.*[20] We also mentioned previous work on certain infinite-dimensional * and grade * representations arising from noncompact real forms of the Lie superalgebras osp(2|1),[14,21] osp(1|4),[22] osp(3|2),[21,23,24] and osp(4|2),[21,25] of relevance to supergravity theories. However, despite these case studies, the classification of * and grade * irreps for a basic classical Lie superalgebra has not received a systematic treatment. Therefore, it is our aim in this series of papers to classify the two types of * and two types of grade * irreps [herein referred to as type (1) and type (2) * and grade * irreps, respectively] for the Lie superalgebras gl$(n|1)$ and C(n).

In this paper it is shown that the dual of a type (1) * irrep of gl$(m|n)$ is a type (2) * irrep (and conversely), as distinct from the grade * case, where the dual of a type (1) [resp. (2)] grade * irrep is again grade * of type (1) [resp. (2)]: However, it turns out that the type (1) and (2) grade * cases are effectively interchanged by a reversal of $\mathbb{Z}_2$ grading. We use these results to give a complete classification of all irreducible type (1) and (2) * and grade * irreps of gl$(n|1)$. It is shown that while there exists a relatively large class of * irreps the grade * irreps are comparatively rare. Indeed, all grade * irreps of gl$(n|1)$ have at most two $\mathbb{Z}$-graded levels and are also * irreps except for the special case of gl(2|1), which possesses a two-parameter family of four-dimensional typical grade * irreps which are not * irreps. In particular, the only cases when typical grade * irreps for gl$(n|1)$ exist is when $n = 1,2$. The classification of * and grade * irreps of the Lie superalgebra C(n) will be given in the second paper of this series.

From the point of view of future research it would clearly be of interest to extend the results of this paper to the Lie superalgebras gl$(m|n)$ and osp$(m|n)$ for general m, n. Our work demonstrates that for the Lie superalgebras gl$(n|1)$ and C$(n+1) = $ osp$(2|2n)$, grade * irreps are comparatively rare, being equivalent, also, to * irreps except for the special cases of gl(2|1) and C(1), which admit an additional class of four-dimensional typical grade * irreps. On the basis of these results we anticipate that grade * irreps of $gl(m|n)$ and $osp(m|n)$ in general are likely to be rare and thus of limited utility in physical applications. By contrast, we expect a relatively large class of * irreps for $gl(m|n)$, which remain to be classified. For such irreps of $gl(m|n)$ it would be of interest to determine the $gl(m|n)\downarrow gl(m|n-1)$ branching rules and obtain the matrix elements of the $gl(m|n)$ generators in the resulting Gel'fand–Tsetlin basis. Finally, it would also be of interest to consider extensions to noncompact real forms of the Lie superalgebras $gl(m|n)$ and $osp(m|n)$.

II. PRELIMINARIES AND INDUCED FORMS

In ungraded index notation, the $gl(m|n)$ basic elements $E^a{}_b$ $(1\leqslant a, b\leqslant m+n)$ satisfy the graded commutation relations

$$[E^a{}_b, E^c{}_d] = \delta^c_b E^a{}_d - (-1)^{[(a)+(b)][(c)+(d)]}\delta^a_d E^c{}_b, \quad (1)$$

where for $a\leqslant m$ (referred to herein as even indices) we define $(a)=0$ and for $a>m$ (referred to herein as odd indices) we define $(a)=1$. It is also sometimes useful to write these generators in graded index notation, where we introduce even indices $i, j,...$ $(=1,...,m)$ and odd indices μ, $\nu,...(=1,...,n)$, in terms of which our even $gl(m)\oplus gl(n)$ generators are given by the operators $E^i{}_j(1\leqslant i,j\leqslant m)$, $E^\mu{}_\nu(1\leqslant\mu,\nu\leqslant n)$, respectively, and our odd generators are given by the operators $E^i{}_\mu$, $E^\mu{}_i(1\leqslant i\leqslant m, 1\leqslant\mu\leqslant n)$. We remark that the bracket on the lhs of Eq. (1) refers to the usual commutator except in the case where both generators $E^a{}_b$, $E^c{}_d$ are odd, in which case it refers to the anticommutator.

As a basis for a Cartan subalgebra of $gl(m|n)$ we choose the commuting operators $E^i{}_i$ $(1\leqslant i\leqslant m)$, $E^\mu{}_\mu$ $(1\leqslant\mu\leqslant n)$ whose eigenvalues serve to label the weights of the representations. We denote the weights of $gl(m|n)$, in the notation of Kac,[8] by

$$\Lambda = (\lambda\,|\chi) = \sum_{i=1}^m \lambda_i\epsilon_i + \sum_{\mu=1}^n \chi_\mu\delta_\mu,$$

so that with this convention, the root system of $gl(m|n)$ is given by the set of even roots

$$\pm(\epsilon_i-\epsilon_j),\quad 1\leqslant i<j\leqslant m,\quad \pm(\delta_\mu-\delta_\nu),\quad 1\leqslant\mu<\nu\leqslant n,$$

together with the set of odd roots

$$\pm(\epsilon_i-\delta_\mu),\quad 1\leqslant i\leqslant m,\quad 1\leqslant\mu\leqslant n.$$

Following Kac,[8] we choose as a system of simple roots the distinguished set

$$\epsilon_i-\epsilon_{i+1}\,(1\leqslant i<m),\quad \alpha_s=\epsilon_m-\delta_1,$$
$$\delta_\mu-\delta_{\mu+1}\,(1\leqslant\mu<n),$$

so that the sets of even and odd positive roots are given, respectively, by

$$\Phi_0{}^+ = \{\epsilon_i-\epsilon_j|1\leqslant i<j\leqslant m\}\cup\{\delta_\mu-\delta_\nu|1\leqslant\mu<\nu\leqslant n\},$$
$$\Phi_1{}^+ = \{\epsilon_i-\delta_\mu|1\leqslant i\leqslant m, 1\leqslant\mu\leqslant n\}.$$

We denote the half-sum of the even and odd positive roots, respectively, by

$$\rho_0 = \frac{1}{2}\sum_{\alpha\in\Phi_0{}^+}\alpha = \frac{1}{2}\sum_{i=1}^m (m+1-2i)\epsilon_i + \frac{1}{2}\sum_{\mu=1}^n (n+1-2\mu)\delta_\mu, \quad (2a)$$

$$\rho_1 = \frac{1}{2}\sum_{\alpha\in\Phi_1{}^+}\alpha = \frac{n}{2}\sum_{i=1}^m\epsilon_i - \frac{m}{2}\sum_{\mu=1}^n\delta_\mu = \frac{1}{2}(\dot{n}|-\dot{m}), \quad (2b)$$

and set

$$\rho = \rho_0-\rho_1. \quad (2c)$$

Throughout this work we let (,) denote the nondegenerate bilinear form define on the weights by[8]

$$(\Lambda,\Lambda') = \sum_{i=1}^m \lambda_i\lambda'_i - \sum_{\mu=1}^n \chi_\mu\chi'_\mu,$$

where $\Lambda=(\lambda|\chi)$, $\Lambda'=(\lambda'|\chi')$.

Every finite-dimensional $gl(m|n)$ module V admits a $\mathbb{Z}_2$ gradation (compatible with the superalgebra grading)

$$V = V_{\bar 0}\oplus V_{\bar 1},$$

where $V_{\bar 0}$ (resp., $V_{\bar 1}$) is referred to as the even (resp., odd) component of V. We then define, for homogeneous $v\in V$, the parity factor (v) by $(v)=0$ (resp. 1) according to whether $v\in V_{\bar 0}$ (resp., $V_{\bar 1}$). Following Kac,[8] the finite-dimensional irreducible $gl(m|n)$ modules are uniquely characterized by their highest weights Λ, where Λ is a dominant weight for $gl(m)\oplus gl(n)$: We denote the set of dominant weights for $gl(m)\oplus gl(n)$ [and hence $gl(m|n)$] by D_+. For $\Lambda\in D_+$ we let $V(\Lambda)$ denote the finite-dimensional irreducible $gl(m|n)$ module with highest weight Λ and we denote the finite-dimensional irreducible $gl(m)\oplus gl(n)$ module with highest weight Λ by $V_0(\Lambda)$. Throughout, we denote the set of *distinct* weights in $V(\Lambda)$ [resp. $V_0(\Lambda)$] by $\Pi(\Lambda)$ [resp. $\Pi_0(\Lambda)$]. Following Kac,[8] we say that $\Lambda\in D_+$ and the corresponding irreducible module $V(\Lambda)$ are *typical* if

$$(\Lambda+\rho,\alpha)\neq 0,\quad \forall\alpha\in\Phi_1{}^+;$$

otherwise, $V(\Lambda)$ and Λ are called *atypical*.

We note that every finite-dimensional irreducible $gl(m|n)$ module admits a natural $\mathbb{Z}$ gradation[7,8]

$$V(\Lambda) = \bigoplus_{k=0}^d V_k(\Lambda) \quad (3)$$

[$V_d(\Lambda)\neq(0)$ assumed], in which case we say that $V(\Lambda)$ *admits d+1 levels*. The $\mathbb{Z}$ gradation in (3) induces the following partitioning of the weights in $V(\Lambda)$:

$$\Pi(\Lambda) = \bigcup_{k=0}^d \Pi_k(\Lambda)$$

where $\Pi_k(\Lambda)$ is the set of distinct weights in the subspace $V_k(\Lambda)$. We note that the spaces $V_k(\Lambda)$ of Eq. (3) are to constitute modules for the even subalgebra $gl(m)\oplus gl(n)$,[7,8] from which it follows[26] that $\Pi_k(\Lambda)$ is stable under the Weyl group of $gl(m)\oplus gl(n)$ [also referred to as the Weyl group of $gl(m|n)$].

In order to study the * and grade * irreps of $gl(m|n)$ it is convenient to consider a natural method for inducing (nondegenerate) sesquilinear forms on $V(\Lambda)$ from a given Hermitian irreducible $gl(m) \oplus gl(n)$ module $V_0(\Lambda)$: Note that this automatically imposes the restriction that $\Lambda \in D_+$ must be real. Let us therefore assume that $V_0(\Lambda)$ is a Hermitian irreducible $gl(m) \oplus gl(n)$ module, so that $V_0(\Lambda)$ is equipped with a (positive definite) inner product $\langle | \rangle$ satisfying

$$\langle E^i{}_j v | w \rangle = \langle v | E^j{}_i w \rangle, \quad \langle E^\mu{}_\nu v | w \rangle = \langle v | E^\nu{}_\mu w \rangle,$$
$$\langle v | w \rangle^* = \langle w | v \rangle, \quad \forall v, \; w \in V_0(\Lambda).$$

It is then possible to extend the form $\langle | \rangle$ to all of $V(\Lambda)$ in four different ways. In order to account for these possibilities we introduce two grading parameters θ, ϵ which can take the values 0 or 1. For a given pair of values θ, ϵ we extend $\langle | \rangle$ to all of $V(\Lambda)$ by defining

$$\langle V_0(\Lambda) | V_k(\Lambda) \rangle = 0, \quad 0 < k \leqslant d,$$
$$\langle E^\mu{}_i v | w \rangle = (-1)^{\theta(v) + \epsilon} \langle v | E^i{}_\mu w \rangle, \tag{4}$$
$$\langle \alpha v_1 + \beta v_2 | w \rangle = \alpha^* \langle v_1 | w \rangle + \beta^* \langle v_2 | w \rangle, \quad \forall \alpha, \; \beta \in \mathbb{C}.$$

With this definition the above form is sesquilinear and the decomposition (3) is orthogonal under the form $\langle | \rangle$. We now have to demonstrate that this form is well defined, so that given $v, w \in V(\Lambda)$, the value $\langle v | w \rangle$ is unique (i.e., independent of the ways in which it is evaluated). This result and some other properties of the form are proved in the following lemma.

Lemma 1: (i) $\langle | \rangle$ is well defined; (ii) $\langle E^i{}_j v | w \rangle = \langle v | E^j{}_i w \rangle$, $\langle E^\mu{}_\nu v | w \rangle = \langle v | E^\nu{}_\mu w \rangle$; (iii) $\langle v | w \rangle = \langle w | v \rangle^*$; and (iv) $\langle E^i{}_\mu v | w \rangle = (-1)^{(w) \cdot \theta + \epsilon} \times \langle v | E^\mu{}_i w \rangle$, $\forall v, w \in V(\Lambda)$.

Proof: For simplicity we prove the result only for the case $\theta = \epsilon = 0$, the remaining cases follow by a similar argument. We proceed to prove (i)–(iii) together by induction on the $\mathbb{Z}$-grading index k, where the results are guaranteed for $k = 0$ by construction. We therefore assume that (i)–(iii) hold for $V_{k-1}(\Lambda)$ and note that every vector v in $V_k(\Lambda)$ may be written in the form

$$v = \sum_{i,\mu} E^\mu{}_i v_{i,\mu}, \quad v_{i,\mu} \in V_{k-1}(\Lambda).$$

To prove (i) it suffices to demonstrate that for $v = 0$,

$$\langle v | w \rangle = 0, \quad \forall w \in V_k(\Lambda)$$

is always true. Now $v = 0$ implies

$$0 = E^j{}_\nu v = \sum_{i,\mu} E^j{}_\nu E^\mu{}_i v_{i,\mu}$$
$$= \sum_{i,\mu} (\delta^\mu{}_\nu E^j{}_i + \delta^j{}_i E^\mu{}_\nu - E^\mu{}_i E^j{}_\nu) v_{i,\mu}. \tag{5}$$

Thus for any $w \in V_{k-1}(\Lambda)$,

$$\langle v | E^\gamma_j w \rangle = \sum_{i,\mu} \langle E^\mu_i v_{i,\mu} | E^\gamma_j w \rangle$$
$$= \sum_{i,\mu} \langle v_{i,\mu} | (\delta^\gamma_\mu E^i_j + \delta^i_j E^\gamma_\mu - E^\gamma_j E^i_\mu) w \rangle.$$

However, $v_{i,\mu}, w \in V_{k-1}(\Lambda)$, and the induction hypothesis allows us to rewrite the above equation as

$$\langle v | E^\gamma_j w \rangle = \sum_{i,\mu} \langle (\delta^\mu_\gamma E^j_i + \delta^j_i E^\mu_\gamma - E^\mu_i E^j_\gamma) v_{i,\mu} | w \rangle,$$

which vanishes by virtue of Eq. (5). Since $V_k(\Lambda)$ is spanned by vectors of the form $E^\gamma_j w$, $w \in V_{k-1}(\Lambda)$, this shows that

$$\langle v | w \rangle = 0, \quad \forall w \in V_k(\Lambda),$$

from which we conclude that (i) is true.

In reference to (ii) we have, for all $v, w \in V_{k-1}(\Lambda)$,

$$\langle E^i_j E^\mu_r v | E^\gamma_l w \rangle = -\delta^i_r \langle E^\mu_j v | E^\gamma_l w \rangle + \langle E^\mu_r E^i_j v | E^\gamma_l w \rangle$$
$$= \langle v | (E^i_j E^r_\mu E^\gamma_l - \delta^i_r E^j_\mu E^\gamma_l) w \rangle$$
$$= \langle v | E^r_\mu E^j_i E^\gamma_l w \rangle = \langle E^\mu_r v | E^j_i E^\gamma_l w \rangle.$$

Again, because the vectors $E^\mu_r v$, $v \in V_{k-1}(\Lambda)$ span $V_k(\Lambda)$, we have

$$\langle E^i_j v | w \rangle = \langle v | E^j_i w \rangle, \quad \forall v, \; w \in V_k(\Lambda).$$

A similar argument applies to the generators E^μ_ν and establishes (ii), as required.

For (iii) we have, for all $v, w \in V_{k-1}(\Lambda)$,

$$\langle E^\mu_i v | E^\gamma_j w \rangle = \langle v | E^i_\mu E^\gamma_j w \rangle$$
$$= \langle v | (\delta^i_j E^\gamma_\mu + \delta^\gamma_\mu E^i_j) w \rangle - \langle E^j_\nu v | E^i_\mu w \rangle. \tag{6}$$

On the other hand,

$$\langle E^\gamma_j w | E^\mu_i v \rangle^* = \langle w | E^j_\nu E^\mu_i v \rangle^*,$$

from which, in view of the inductive hypothesis, we obtain

$$\langle E^\gamma_j w | E^\mu_i v \rangle^* = \langle E^j_\nu E^\mu_i v | w \rangle$$
$$= \langle (\delta^\mu_\nu E^j_i + \delta^j_i E^\mu_\nu - E^\mu_i E^j_\nu) v | w \rangle$$
$$= \langle v | (\delta^i_j E^\nu_\mu + \delta^\nu_\mu E^i_j) w \rangle - \langle E^j_\nu v | E^i_\mu w \rangle$$
$$= \langle E^\mu_i v | E^\gamma_j w \rangle,$$

where in the second to last step we employed the inductive hypothesis and the last equality follows from Eq. (6). Since the vectors $E^\mu_i v$, $v \in V_{k-1}(\Lambda)$ span $V_k(\Lambda)$ we thus deduce

$$\langle v | w \rangle^* = \langle w | v \rangle, \quad \forall v, w \in V_k(\Lambda),$$

which proves parts (iii). Hence, by induction, we conclude that parts (i)–(iii) of Lemma 1 are true.

Finally, in reference to (iv) we have, for all $v, w \in V(\Lambda)$,

$$\langle E^i_\mu v | w \rangle = \langle w | E^i_\mu v \rangle^* = \langle E^\mu_i w | v \rangle^* = \langle v | E^\mu_i w \rangle,$$

where (iii) has been repeatedly used. This proves Lemma 1, as required.

Let us agree to call any sesquilinear form on $V(\Lambda)$ satisfying the conditions of Lemma 1 *invariant* of type (θ, ϵ). Another important property of the induced form $\langle | \rangle$ is given by the following lemma.

Lemma 2: The form $\langle | \rangle$ induced on $V(\Lambda)$ is the unique (up to scalar multiples) invariant nondegenerate sesquilinear form of type (θ, ϵ) on $V(\Lambda)$.

Proof: To prove nondegeneracy we note that the kernel of the form $\langle | \rangle$ is given by

$$K = \{ v \in V(\Lambda) | \langle v | w \rangle = 0, \quad \forall w \in V(\Lambda) \}.$$

Following Lemma 1, K constitutes a $\mathbb{Z}$-graded $gl(m|n)$ submodule of $V(\Lambda)$ which cannot equal $V(\Lambda)$ since by construction the restriction of $\langle | \rangle$ to $V_0(\Lambda)$ is nondegenerate, i.e., $K \cap V_0(\Lambda) = (0)$. Thus since $V(\Lambda)$ is irreducible, we conclude $K = (0)$, as required.

To prove uniqueness let $\langle | \rangle_1$ be another nondegenerate invariant sesquilinear form of type (θ,ϵ), on $V(\Lambda)$. We then define a vector space mapping ξ: $V(\Lambda)\to V(\Lambda)$, $\xi(v)=v'$, where v' is given by

$$\langle v'|w\rangle_1=\langle v|w\rangle,\quad \forall w\in V(\Lambda).$$

We note that ξ is well defined and 1–1 since the forms are nondegenerate; it also gives rise to a homomorphism (and hence isomorphism) of $gl(m|n)$ modules. In view of Schur's lemma[27] it follows that ξ reduces to a scalar multiple α of the identity on $V(\Lambda)$, so that

$$\langle v|w\rangle_1=\alpha\langle v|w\rangle,\quad \forall v,\ w\in V(\Lambda),$$

which proves the result.

The above shows that on each finite-dimensional irreducible $gl(m|n)$ module $V(\Lambda)$, Λ real, we may induce a nondegenerate invariant sesquilinear form of type (θ,ϵ) which is uniquely determined by its restriction to the maximal $\mathbb{Z}$-graded component $V_0(\Lambda)$. Such a form has all the properties of an inner product except that it is not generally positive definite. As shall be discussed in Sec. III, if the induced form $\langle|\rangle$ is positive definite we say that $V(\Lambda)$ is a star module of type (1) [resp. (2)] if $\theta=0$ and $\epsilon=0$ (resp. 1), while we say that $V(\Lambda)$ is a grade star module of type (1) [resp. (2)] if $\theta=1$ and $\epsilon=0$ (resp. 1).

It is worth noting that even when the form $\langle|\rangle$ is not positive definite it may still be applied to obtain matrix elements, etc. even for irreps which are neither * or grade *. To this end we note that the Gel'fand invariants of $gl(m|n)$ and its canonical subalgebras are all self-adjoint under the above form, from which we deduce that Gel'fand states corresponding to different eigenvalues are necessarily orthogonal under the induced form.

We conclude this section by noting that the above-mentioned induced inner product construction applies to any finite-dimensional indecomposable $gl(m|n)$ module $\overline{V}(\Lambda)$ admitting a $\mathbb{Z}$ gradation

$$\overline{V}(\Lambda)=\bigoplus_{k=0}^{d}\overline{V}_k(\Lambda),$$

where $\overline{V}_0(\Lambda)=V_0(\Lambda)$ is an irreducible $gl(m)\oplus gl(n)$ submodule of (real) highest weight $\Lambda\in D_+$. Such a form then satisfies the properties of Lemma 1 (i.e., is invariant), but will be degenerate unless $\overline{V}(\Lambda)=V(\Lambda)$ is irreducible. Therefore, the quotient module defined by

$$V(\Lambda)=\overline{V}(\Lambda)/K,\quad K=\{v\in\overline{V}(\Lambda)|\langle v|w\rangle=0,\quad \forall w\in\overline{V}(\Lambda)\}$$

is necessarily irreducible. In particular, such a form can be defined on the corresponding Kac-induced module[8] and suggests a convenient way of extracting an irreducible module from an indecomposable one.

III. STAR AND GRADE STAR $gl(m|n)$ MODULES

Following Scheunert *et al.*[19] $gl(m|n)$ admits two types of irreducible * representations. We say that $V(\Lambda)$ is an irreducible * module of type (1) [resp. (2)] if $V(\Lambda)$ can be equipped with an inner product (necessarily positive definite) $\langle|\rangle$ on which the generators $E^a{}_b$ satisfy the hermiticity requirements

$$\pi_\Lambda{}^\dagger(E^a{}_b)=\pi_\Lambda(E^b{}_a)\,[\text{resp.}\,(-1)^{[(a)+(b)]}\pi_\Lambda(E^b{}_a)],$$

where π_Λ is the representation afforded by $V(\Lambda)$. Equivalently, $V(\Lambda)$ is * of type (1) if

$$\langle E^a{}_b v|w\rangle=\langle v|E^b{}_a w\rangle$$

and * of type 2 if

$$\langle E^a{}_b v|w\rangle=(-1)^{[(a)+(b)]}\langle v|E^b{}_a w\rangle.$$

To define grade * modules we first recall that the grade adjoint $A^\ddagger$ of a homogeneous operator A on a $\mathbb{Z}_2$-graded Hilbert space V with the inner product $\langle|\rangle$ is defined by[19,27]

$$\langle Av|w\rangle=(-1)^{(A)(v)}\langle v|A^\ddagger w\rangle,$$

where, as usual, (A) denotes the parity of A. Equivalently,

$$A^\ddagger=(A^T)^*,$$

where T denotes the graded transpose.[27] We note that the grade adjoint coincides with the normal Hermitian conjugate for even operators and

$$(A^\ddagger)^\ddagger=(A^T)^T=(-1)^{(A)}A.$$

We then say that $V(\Lambda)$ is an irreducible grade * module of type (1) [resp. (2)] if $V(\Lambda)$ can be equipped with a (positive definite) inner product $\langle|\rangle$ on which the generators $E^a{}_b$ satisfy the graded-hermiticity conditions

$$\pi_\Lambda^\ddagger(E^a{}_b)=(-1)^{[(a)+(b)]\epsilon}\pi_\Lambda(E^b{}_a),$$

$$\epsilon=(a)\,[\text{resp.}\,(b)].$$

Equivalently, $V(\Lambda)$ is grade * of type (1) if

$$\langle E^a_b v\langle w\rangle=(-1)^{[(a)+(b)][(a)+(v)]}\langle v|E^b_a w\rangle$$

and grade * of type (2) if

$$\langle E^a_b v|w\rangle=(-1)^{[(a)+(b)][(b)+(v)]}\langle v|E^b_a w\rangle.$$

In graded index notation, on * or grade * irreps $V(\Lambda)$, the even $gl(m)\oplus gl(n)$ generators always satisfy

$$\langle E^i_j v|w\rangle=\langle v|E^j_i w\rangle,\quad \langle E^\mu_\nu v|w\rangle=\langle v|E^\nu_\mu w\rangle,$$

which is just the condition that $V(\Lambda)$ give rise to a Hermitian representation of $gl(m)\oplus gl(n)$: This implies immediately that the components of the highest weight Λ must be real. On the other hand, our odd generators must satisfy

$$\langle E^\mu_i v|w\rangle=\langle v|E^i_\mu w\rangle,\tag{7a}$$

$$\langle E^\mu_i v|w\rangle=-\langle v|E^i_\mu w\rangle\tag{7b}$$

in the type (1) and (2) * cases, respectively, while for the type (1) and (2) grade * cases we have, respectively,

$$\langle E^\mu_i v|w\rangle=(-1)^{(v)}\langle v|E^i_\mu w\rangle,\tag{8a}$$

$$\langle E^\mu_i v|w\rangle=-(-1)^{(v)}\langle v|E^i_\mu w\rangle.\tag{8b}$$

We note, in the notation of Sec. II, that the inner product $\langle|\rangle$ is a nondegenerate invariant sesquilinear form of type $\theta=0$, $\epsilon=0$, 1 in cases (7a) and (7b), respectively, and of type $\theta=1$, $\epsilon=0$, 1 in the respective cases (8a) and (8b). In view of the uniqueness of the induced form, it follows that the above inner product necessarily coincides the form induced on $V(\Lambda)$ by the restriction of $\langle|\rangle$ to $V_0(\Lambda)$. Therefore, we see that $V(\Lambda)$ is a * or grade * module if and only if $V_0(\Lambda)$ is a Hermitian irrep of $gl(m)\oplus gl(n)$ and the corresponding form induced on $V(\Lambda)$ is positive definite (i.e., gives rise to an inner product).

The study of $gl(m|n)$ * modules $V(\Lambda)$ is facilitated by noting that the type (1) and (2) cases are related by duality. We recall that the dual representations $\tilde{\pi}_\Lambda$ and π_Λ is defined by[27]

$$\tilde{\pi}_\Lambda(E^a{}_b) = -\pi_\Lambda^T(E^a{}_b),$$

where T denotes the super transpose. Introducing a homogeneous basis $\{v_\alpha\}$ for $V(\Lambda)$, let us write $(\alpha)=0,1$ according to whether v_α is odd or even, respectively. Then in this basis we have[27]

$$\pi_\Lambda^T(E^a{}_b)_{\alpha\beta} = -(-1)^{[(a)+(b)](\beta)}\pi_\Lambda(E^a{}_b)_{\beta\alpha}.$$

If $V(\Lambda)$ is a type (1) * module, so that

$$\pi_\Lambda^\dagger(E^a{}_b) = \pi_\Lambda(E^b{}_a),$$

then the dual representation $\tilde{\pi}_\Lambda$ must satisfy

$$\begin{aligned}\tilde{\pi}_\Lambda{}^\dagger(E^a{}_b)_{\alpha\beta} &= -\pi_\Lambda^T(E^a{}_b)^*{}_{\beta\alpha}\\ &= -(-1)^{[(a)+(b)](\alpha)}\pi_\Lambda(E^a{}_b)^*_{\alpha\beta}\\ &= -(-1)^{[(a)+(b)](\alpha)}\pi_\Lambda(E^b{}_a)_{\beta\alpha}\\ &= (-1)^{[(a)+(b)][(\alpha)+(\beta)]}\tilde{\pi}_\Lambda(E^b{}_a)_{\alpha\beta}.\end{aligned}$$

Using $(\alpha)+(\beta)=(a)+(b)$ we thereby obtain

$$\tilde{\pi}_\Lambda{}^\dagger(E^a{}_b) = (-1)^{[(a)+(b)]}\tilde{\pi}_\Lambda(E^b{}_a),$$

so that $\tilde{\pi}_\Lambda$ is a type (2) * representation; that is, the dual of a type (1) * representation is a type (2) * representation and conversely. Denoting the irreducible module dual to $V(\Lambda)$ by $V^*(\Lambda)$, we thus obtain the following proposition.

Proposition 1: For $\Lambda\in D_+$, $V(\Lambda)$ is an irreducible type (1) * module of $gl(m|n)$ if and only if $V^*(\Lambda)$ is an irreducible * module of type (2). □

As distinct from the above situation for * modules, we have the following result concerning grade * modules.

Proposition 2: For $\Lambda\in D_+$, $V(\Lambda)$ is an irreducible type (1) [resp (2)] grade * module if and only if $V^*(\Lambda)$ is an irreducible grade * module of type (1) [resp. (2)].

Proof: Suppose π_Λ affords a type (1) grade * irrep of $gl(m|n)$, so that

$$\pi_\Lambda^\ddagger(E^a{}_b) = (-1)^{[(a)+(b)](a)}\pi_\Lambda(E^b{}_a)$$

or

$$\pi_\Lambda^T(E^a{}_b)^* = (-1)^{[(a)+(b)](a)}\pi_\Lambda(E^b{}_a).$$

Using the definition of the dual representation $\tilde{\pi}_\Lambda$, the above equations yield

$$\begin{aligned}\tilde{\pi}_\Lambda(E^a{}_b) &= -(-1)^{[(a)+(b)](a)}\pi_\Lambda(E^b{}_a)^*\\ &= -(-1)^{[(a)+(b)](b)}(\pi_\Lambda^T)^\ddagger(E^b{}_a),\end{aligned}$$

where we have used the result

$$\begin{aligned}(\pi_\Lambda^T)^\ddagger(E^a{}_b) &= (\pi_\Lambda^T)^T(E^a{}_b)^*\\ &= (-1)^{[(a)+(b)]}\pi_\Lambda(E^a{}_b)^*.\end{aligned}$$

Hence,

$$\begin{aligned}\tilde{\pi}_\Lambda^\ddagger(E^a{}_b) &= -(-1)^{[(a)+(b)](a)}\pi_\Lambda^T(E^b{}_a)\\ &= (-1)^{[(a)+(b)](a)}\tilde{\pi}_\Lambda(E^b{}_a),\end{aligned}$$

so that $\tilde{\pi}_\Lambda$ is also grade * of type (1). A similar argument holds for the type (2) case. □

It is important to note that unlike the * case, the type of grade * module depends on the $\mathbb{Z}_2$ grading (even or odd) chosen for the maximal $\mathbb{Z}$ graded $gl(m)\oplus gl(n)$ component $V_0(\Lambda)$: This follows because the actual definition of grade adjoint depends on the choice of $\mathbb{Z}_2$ grading. We note, however, in view of Eqs. (8a) and (8b), that the type (1) and (2) grade * cases are interchanged by a reversal of $\mathbb{Z}_2$ grading. In other words, a type (1) grade * module $V(\Lambda)$ in which $V_0(\Lambda)$ is chosen to have *odd* $\mathbb{Z}_2$ grading is equivalent to regarding $V(\Lambda)$ as a type (2) grade * module in which $V_0(\Lambda)$ is regarded as *even*. Therefore, it suffices to classify all type (1) and (2) grade * irreps with the standard choice of $\mathbb{Z}_2$ grading for $V(\Lambda)$, where the maximal $\mathbb{Z}$-graded component $V_0(\Lambda)$ is chosen to be even: Throughout the paper we adopt this standard choice of $\mathbb{Z}_2$ grading.

However, it should be noted that in Proposition 2 the $\mathbb{Z}_2$ grading of the maximal component of $V^*(\Lambda)$ is assumed to be given by the grading of the minimal component of $V(\Lambda)$. Therefore, with the standard $\mathbb{Z}_2$ grading for $V^*(\Lambda)$, Proposition 2 states that if $V(\Lambda)$ is a type (1) [resp. (2)] grade * module, then $V^*(\Lambda)$ is also a type (1) [resp. (2)] grade * module if $V(\Lambda)$ has an *odd* number of levels, while if $V(\Lambda)$ has an *even* number of levels, then $V^*(\Lambda)$ is necessarily grade * of type (2) [resp. (1)].

As noted in Sec. I, the * and grade * irreps of $gl(m|n)$ are those most likely to be of interest in physical applications. The remainder of this paper is devoted to a classification of the finite-dimensional irreducible * and grade * $gl(n|1)$ modules. It is hoped that this work will provide insight into the difficult problem of determining all * and grade * irreps for $gl(m|n)$ in general.

IV. STAR MODULES FOR gl(n|1)

In this section we present a detailed study of the * irreps for $gl(n|1)$. In particular, we will prove Theorems 1 and 2 which specify the necessary and sufficient conditions on the highest weight of an irreducible module in order that it be star of type (1) or (2). This gives a complete classification of the star representations in terms of their highest weights.

For simplicity, we alter our notation and denote the even $gl(n)\oplus gl(1)$ generators by E^i_j $(1\leqslant i,j\leqslant n)$, Ω, respectively, and the odd generators of $gl(n|1)$ by E^i, E_i $(1\leqslant i\leqslant n)$. In this case the set of even positive roots Φ_0^+ consists simply of the positive roots $\epsilon_i-\epsilon_j$ $(1\leqslant i\leqslant j\leqslant n)$ of $gl(n)$ and our odd positive roots are given by

$$\Phi_0^+ = \{\epsilon_i - \delta_1 \,|\, 1\leqslant i\leqslant n\}.$$

We recall[8] that $(\rho,\alpha_s)=0$, where $\alpha_s=\epsilon_n-\delta_1$ is the odd simple root and $(\rho,\alpha)=(\rho_0,\alpha)\in\mathbb{Z}^+$ for $\alpha\in\Phi_0^+$. Throughout, we denote our $gl(n|1)$ weights by $\Lambda=(\lambda\,|\,\omega)$, where $\omega\in\mathbb{C}$ and λ is a $gl(n)$ weight, and we let W denote the Weyl group of $gl(n)$ [and hence, $gl(n|1)$].

We now note that if $V(\Lambda)$ is an irreducible type 1 * module of $gl(n|1)$ with the inner product $\langle\,|\,\rangle$, then for v^Λ, the highest weight vector of $V(\Lambda)$, we must have

$$0\leqslant\langle E_i v^\Lambda|E_i v^\Lambda\rangle=\langle v^\Lambda|E^iE_i v^\Lambda\rangle$$
$$=\langle v^\Lambda|(E^i{}_i+\Omega)v^\Lambda\rangle$$
$$=(\Lambda,\epsilon_i-\delta_1)\langle v^\Lambda|v^\Lambda\rangle,$$

where in the above we exploited the fact that $E^iv^\Lambda=0$ $(1\leqslant i\leqslant n)$. Thus in order for $V(\Lambda)$ to be a type 1 * module Λ must be real and satisfy

$$(\Lambda,\alpha)\geqslant 0,\quad \forall\in\Phi_1^+.$$

We have the following result concerning such weights.

Proposition 3: For $\Lambda\in D_+$, the following conditions are equivalent: (i) $(\Lambda,\alpha_s)\geqslant 0$, (ii) $(\nu,\alpha)\geqslant 0$, $\forall\alpha\in\Phi_1^+$, $\nu\in\Pi(\Lambda)$.

Proof: Clearly (i) is a special case (ii), so that it suffices to show that (i) implies (ii). For $\alpha_i=\epsilon_i-\epsilon_{i+1}$ $(1\leqslant i<n)$ we have

$$(\alpha_i,\alpha_s)=(\epsilon_i-\epsilon_{i+1},\epsilon_n-\delta_1)=-\delta_{n,i+1}\leqslant 0,$$
$$(\alpha_s,\alpha_s)=0.$$

Now for $\nu\in\Pi(\Lambda)$ we may write

$$\nu=\Lambda-\sum_i n_i\alpha_i-n_s\alpha_s,\quad n_i,n_s\in\mathbb{Z}^+,$$

so that

$$(\nu,\alpha_s)=(\Lambda,\alpha_s)-\sum_i n_i(\alpha_i,\alpha_s)\geqslant(\Lambda,\alpha_s).$$

Thus $(\Lambda,\alpha_s)\geqslant 0$ implies

$$(\nu,\alpha_s)\geqslant 0,\quad \forall\nu\in\Pi(\Lambda).$$

Now for $\alpha\in\Phi_1^+$ arbitrary, there exists $\sigma\in W$ such that $\alpha=\sigma(\alpha_s)$, so that for $(\Lambda,\alpha_s)\geqslant 0$,

$$(\nu,\alpha)=(\nu,\sigma(\alpha_s))=(\sigma^{-1}(\nu),\alpha_s)\geqslant 0,\quad \forall\nu\in\Pi(\Lambda),$$

where we have used the invariance of (,) under W and the fact that W acts transitively on Φ_1^+. This proves the result.

Let us agree to call a weight $\Lambda\in D_+$ * *permissible* if Λ is real and satisfies $(\Lambda,\alpha_s)\geqslant 0$: We denote the set of *-permissible dominant weights by D_+^*. Clearly, as noted above, in order for $V(\Lambda)$ to be a type (1) * module it is necessary that $\Lambda\in D_+^*$. We now investigate the converse.

We first note that $V(\Lambda)$ is a direct sum of irreducible $gl(n)\oplus gl(1)$ submodules $V_0(\nu)$, where ν is dominant, and each such module occurs with unit multiplicity.[17] It follows that the $gl(n)\oplus gl(1)$ invariant

$$\eta=E_iE^i$$

(summation on i from 1 to n) must reduce to scalar multiple of the identity on each such irreducible $gl(n)\oplus gl(1)$ submodule $V_0(\nu)$. For $\Lambda\in D_+^*$, we have the following result concerning the eigenvalues of η.

Proposition 4: Suppose $\Lambda\in D_+^*$. Then (i) For all $\nu\in\Pi(\Lambda)$, $(\Lambda-\nu,\Lambda+\nu)\geqslant 0$, $(\Lambda-\nu,\Lambda+\nu+2\rho)\geqslant 0$. (ii) The eigenvalues of η on $V(\Lambda)$ are all non-negative.

Proof: Using the partitioning

$$\Pi(\Lambda)=\bigcup_{k=0}^{d}\Pi_k(\Lambda)$$

we proceed to prove (i) by induction on the Z-grading index k. For $k=0$ we have, from a well-known classical Lie algebra result,[26]

$$(\Lambda-\nu,\Lambda+\nu)\geqslant 0,\quad \forall\nu\in\Pi_0(\Lambda).$$

Also, since $\Lambda-\nu$, $\nu\in\Pi_0(\Lambda)$ is a sum of even positive roots we have

$$(\Lambda-\nu,\Lambda+\nu+2\rho)\geqslant(\Lambda-\nu,2\rho)\geqslant 0,\quad \forall\nu\in\Pi_0(\Lambda),$$

so that (i) holds for all $\nu\in\Pi_0(\Lambda)$. We now assume that (i) holds for all $\nu\in\Pi_{k-1}(\Lambda)$ and note that for every $\mu\in\Pi_k(\Lambda)$, there exists $\alpha\in\Phi_1^+$ such that

$$\nu=\mu-\alpha\ \in\Pi_{k-1}(\Lambda).$$

Then

$$(\Lambda-\mu,\Lambda+\mu)=(\Lambda-\nu+\alpha,\Lambda+\nu-\alpha)$$
$$=(\Lambda-\nu,\Lambda+\nu)+2(\mu,\alpha)$$
$$\geqslant(\Lambda-\nu,\Lambda+\nu)\geqslant 0,$$

where the second to last inequality follows from proposition (3) and the last inequality follows from the inductive hypothesis. Hence,

$$(\Lambda-\mu,\Lambda+\mu+2\rho)$$
$$=(\Lambda-\mu,\Lambda+\mu)+(\Lambda-\mu,2\rho)\geqslant(\Lambda-\mu,2\rho).$$

However, in terms of the even simple roots $\alpha_i=\epsilon_i-\epsilon_{i+1}$ $(1\leqslant i<n)$, we have

$$\Lambda-\mu=\sum_i n_i\alpha_i+n_s\alpha_s,\quad n_i,n_s\in\mathbb{Z}_1^+,$$

so that

$$(\Lambda-\mu,\rho)=\sum_i n_i(\alpha_i,\rho)+n_s(\alpha_s,\rho)=\sum_i n_i(\alpha_i,\rho)\geqslant 0.$$

This is enough to prove (i) for all $\mu\in\Pi_k(\Lambda)$ and hence, part (i) follows by induction.

As to (ii) we note that the second-order invariant I_2 of $gl(n|1)$ may be written as (summation on i from 1 to n is assumed)

$$I_2=\sigma_2+E_iE^i-E^iE_i-\Omega^2,$$

where σ_2 is the second-order invariant of $gl(n)$. To determine the eigenvalue γ of the $gl(n)\oplus gl(1)$ invariant $\eta=E_iE^i$ on an irreducible $gl(n)\oplus gl(1)$ submodule $V_0(\nu)$ of $V(\Lambda)$, let v^ν be the maximal state of $V_0(\nu)$. We then have

$$E_iE^iv^\nu=\gamma v^\nu,\quad E^iE_iv^\nu=\bar\gamma v^\nu,$$

so that

$$(\gamma+\bar\gamma)v^\nu=(E_iE^i+E^iE_i)v^\nu$$
$$=\sum_{i=1}^{n}(E^i_i+\Omega)v^\nu=(\nu,2\rho_1)v^\nu. \tag{9}$$

On the other hand,

$$(\gamma-\bar\gamma)v^\nu=(E_iE^i-E^iE_i)v^\nu=(I_2-\sigma_2+\Omega^2)v^\nu.$$

By recalling that I_2 and $\sigma_2-\Omega^2$ are the quadratic Casimir elements of $gl(n|1)$ and its even subalgebra $gl(n)\oplus gl(1)$, respectively, we obtain

$$(\gamma-\bar\gamma)v^\nu=[(\Lambda,\Lambda+2\rho)-(\nu,\nu+2\rho_0)]v^\nu.$$

The above equation, together with Eq. (9), then yields, for the eigenvalues γ of the $gl(n)\oplus gl(1)$ invariant $\eta=E_iE^i$,

$$2\gamma = (\Lambda,\Lambda + 2\rho) - (\nu,\nu + 2\rho)$$
$$= (\Lambda - \nu,\Lambda + \nu + 2\rho) \geqslant 0,$$

where the last inequality follows from part (i). This proves the result. □

Our aim now is to demonstrate that for $\Lambda \in D^*_+$, $V(\Lambda)$ is a type (1) * module. To this end we assume that $V_0(\Lambda)$ gives rise to a Hermitian representation of $gl(n) \oplus gl(1)$ and we let $\langle | \rangle$ be the corresponding unique nondegenerate invariant sesquilinear form of type $(\theta,\epsilon) = (0,0)$ induced on $V(\Lambda)$. It then suffices to demonstrate that $\langle | \rangle$ is positive definite (i.e., gives rise to an inner product). We first note the following property of the form $\langle | \rangle$.

Lemma 3: With the form $\langle | \rangle$ defined on $V(\Lambda)$ above, the $gl(n)$ Gel'fand invariants are Hermitian, so that $gl(n) \oplus gl(1)$ submodules with different highest weights are orthogonal.

Proof: As noted previously, the definition of the form $\langle | \rangle$ guarantees that the $gl(n)$ Gel'fand invariants are Hermitian and hence, eigenstates corresponding to different eigenvalues are orthogonal under the form. The result then follows from the fact that the $gl(n)$ Gel'fand invariants uniquely label the finite-dimensional irreducible $gl(n)$ modules, together with the result[17] that all irreducible $gl(n) \oplus gl(1)$ submodules of $V(\Lambda)$ occur with unit multiplicity. □

By utilizing Lemma 3 and Proposition 4 we obtain the following result for the above induced form.

Proposition 5: For $\Lambda \in D^*_+$, the form $\langle | \rangle$ on $V(\Lambda)$ is positive definite.

Proof: We employ the **Z** gradation of Eq. (3) and proceed by induction on the **Z**-grading index k, where the result holds for $k=0$ by construction. We now assume that the form is positive definite on $V_{k-1}(\Lambda)$ and note that for $v \in V_k(\Lambda)$, $E^i v \in V_{k-1}(\Lambda)$, from which we obtain, in view of the inductive hypothesis,

$$\langle v | E_i E^i v \rangle = \langle E^i v | E^i v \rangle \geqslant 0, \quad \forall v \in V_k(\Lambda). \tag{10}$$

We now proceed in two steps. (i) $V_k(\Lambda)$ is a direct sum of irreducible $gl(n) \oplus gl(1)$ submodules $V_0(\nu)$. Let us assume initially that $v \neq 0$ belongs to such an irreducible submodule $V_0(\nu)$. Then from Proposition (4), we have, for the invariant $\eta = E_i E^i$,

$$\eta v = \gamma v, \quad \gamma \geqslant 0,$$

so that

$$\gamma \langle v | v \rangle = \langle v | \eta v \rangle = \sum_i \langle v | E_i E^i v \rangle \geqslant 0, \tag{11}$$

where the last inequality follows from Eq. (10). Now if $\gamma \langle v | v \rangle \neq 0$ we must have $\gamma > 0$ and hence, $\langle v | v \rangle > 0$. If, on the other hand, $\gamma \langle v | v \rangle = 0$, Eq. (11) implies

$$0 = \sum_i \langle v | E_i E^i v \rangle = \sum_i \langle E^i v | E^i v \rangle.$$

Then by the inductive hypothesis we deduce

$$E^i v = 0, \quad 1 \leqslant i \leqslant n$$

and hence, $v \in V_0(\Lambda)$, in which case $\langle v | v \rangle > 0$ by construction. Thus we necessarily have, for $v \neq 0$ in an irreducible $gl(n) \oplus gl(1)$ submodule of $V_k(\Lambda)$, $\langle v | v \rangle > 0$.

(ii) We now consider the general case and suppose that $0 \neq v \in V_k(\Lambda)$ is arbitrary. Then v may be expanded as

$$v = \sum_\alpha v_\alpha,$$

where each $v_\alpha \neq 0$ belongs to an irreducible $gl(n) \oplus gl(1)$ submodule of $V_k(\Lambda)$. From Lemma 3 the v_α are orthogonal, so that

$$\langle v | v \rangle = \sum_\alpha \langle v_\alpha | v_\alpha \rangle > 0,$$

where the last inequality follows from step (i). This shows that $\langle v | v \rangle > 0$ for all nonzero $v \in V_k(\Lambda)$, so that $\langle | \rangle$ is positive definite on $V_k(\Lambda)$. Hence, the result is proved by induction. □

We thus arrive at the following classification scheme for the type 1 * irreps of $gl(n|1)$.

Theorem 1: For $\Lambda \in D_+$, $V(\Lambda)$ is equivalent to a type (1) irreducible * module if and only if Λ is real and $(\Lambda,\alpha_s) \geqslant 0$. □

Recalling Proposition 1, we see that Theorem 1 essentially characterizes both types of * representations. To characterize type (2) * modules more explicitly, we note that the weights in $V^*(\Lambda)$ are the negative of the weights in $V(\Lambda)$ and if Λ^- is the lowest weight of $V(\Lambda)$, then $-\Lambda^-$ is the highest weight of $V^*(\Lambda)$. We thus obtain, in view of Theorem 1 and Proposition 1, the following proposition.

Proposition 6: For $\Lambda \in D_+$, $V(\Lambda)$ is equivalent to a type (2) irreducible * module of $gl(n|1)$ if and only if Λ is real and $(\Lambda^-,\alpha_s) \leqslant 0$. □

Now let I_Λ be the index set introduced in Ref. 17, viz. $I_\Lambda = \{1,...,n\}$ if Λ is typical, while if $(\lambda + \rho, \epsilon_i - \delta_1) = 0$ we have

$$I_\Lambda = \{j > i\} \cup \{j < i | (\Lambda,\epsilon_j - \epsilon_i) > 0\}.$$

We note that the integer $d_\Lambda = |I_\Lambda|$ is the maximal **Z**-graded level index of $V(\Lambda)$; i.e., $V(\Lambda)$ admits $d_\Lambda + 1$ levels. We define θ by $\theta = 1$ if Λ is typical; for Λ atypical,

$$\theta = \begin{cases} 1, & 1 \in I_\Lambda, \\ 0, & \text{otherwise.} \end{cases}$$

Then for Λ^-, the lowest weight of $V(\Lambda)$, we have[17]

$$(\Lambda^-,\alpha_s) = \lambda_1 + \omega + d_\Lambda - \theta, \quad \Lambda = (\lambda | \omega),$$
$$= (\Lambda,\epsilon_1 - \delta_1) + d_\Lambda - \theta.$$

In view of Proposition 6 we obtain the following theorem.

Theorem 2: For $\Lambda \in D_+$, $V(\Lambda)$ is equivalent to an irreducible type (2) * module of $gl(n|1)$ if and only if Λ is real and $(\Lambda,\epsilon_1 - \delta_1) + d_\Lambda - \theta \leqslant 0$.

Corollary: If $\Lambda \in D_+$ is typical, then $V(\Lambda)$ is equivalent to a type (2) * module if and only if $(\Lambda+\rho,\epsilon_1-\delta_1)<0$ and Λ is real.

Proof: For Λ typical we have $d_\Lambda=n$, $\theta=1$ and the condition of Theorem 2 reduces to $(\Lambda+\rho,\epsilon_1-\delta_1)\leqslant 0$. Obviously, for Λ typical we cannot have $(\Lambda+\rho,\epsilon_1-\delta_1)=0$.□

V. GRADE STAR MODULES FOR gl(n|1)

This section is devoted to the study of grade * irreps of gl(n|1). The main results obtained are Theorems 3 and 4 which, respectively, classify all type (1) and (2) grade * modules. As will be demonstrated, when $n>2$, only two very special classes of two-level irreps are grade *. However, gl(2|1) proves to be a special case for which four-dimensional typical grade * irreps exist.

Suppose now that $V(\Lambda)$ is a type 1 grade * gl(n|1) module, so that Λ is real and $V(\Lambda)$ is equipped with a (positive definite) inner product $\langle|\rangle$ which is invariant of type $\theta=1$, $\epsilon=0$, viz. for all v, $w\in V(\Lambda)$,

$$\langle E_i v|w\rangle=(-1)^{(v)}\langle v|E^i w\rangle,$$
$$\langle E^i v|w\rangle=-(-1)^{(v)}\langle v|E_i w\rangle.$$

Then if $v\in V_0(\Lambda)$ has weight $\nu\in\Pi_0(\Lambda)$ we have

$$\begin{aligned}0\leqslant\langle E_i v|E_i v\rangle&=\langle v|E^iE_i v\rangle\\&=\langle v|(E^i_i+\Omega)v\rangle\\&=(\nu,\epsilon_i-\delta_1)\langle v|v\rangle.\end{aligned}\tag{12}$$

It follows that

$$(\nu,\alpha)\geqslant 0,\quad\forall\nu\in\Pi_0(\Lambda),\ \alpha\in\Phi_1^+$$

and in particular, $(\Lambda,\alpha_S)\geqslant 0$. Thus from Theorem 1, $V(\Lambda)$ is also a type (1) * module, so that we do not obtain any new type (1) grade * modules which are already * modules.

Proceeding further, we recall that Proposition 3 implies

$$(\nu,\alpha)\geqslant 0,\quad\forall\nu\in\Pi(\Lambda),\ \alpha\in\Phi_1^+.$$

Given a vector $w\in V_1(\Lambda)$ of weight $\mu\in\Pi_1(\Lambda)$, we now wish to demonstrate that for each i, $E_i w=0$. There are two cases to consider.

(i) $(\mu,\epsilon_i-\delta_1)=0$. In such a case we set $v=E^i w\in V_0(\Lambda)$, so that v has weight $\nu=\mu+\epsilon_i-\delta_1$ satisfying $(\nu,\epsilon_i-\delta_1)=0$. Equation (12) then implies

$$\langle E_i v|E_i v\rangle=0,$$

so that $E_i v=0$, i.e.,

$$E_iE^i w=-E^iE_i w=0.$$

Hence, we have

$$0=\langle w|E_iE^i w\rangle=-\langle w|E^iE_i w\rangle$$

or

$$0=\langle E^i w|E^i w\rangle=\langle E_i w|E_i w\rangle,$$

which implies $E^i w=E_i w=0$.

(ii) $(\mu,\epsilon_i-\delta_1)>0$. In this case we set $v=E^iE_i w$ and note that

$$\langle E_i v|E_i v\rangle=-\langle v|E^iE_i v\rangle=-(\mu,\epsilon_i-\delta_1)\langle v|v\rangle,$$

which can only occur if $E_i v=0$. Thus

$$0=E_i v=E_iE^iE_i w=(\mu,\epsilon_i-\delta_1)E_i w,$$

which implies $E_i w=0$.

Thus in either case (i) or (ii), we deduce $E_i w=0$, from which we obtain

$$E_i w=0,\quad\forall w\in V_1(\Lambda),\ 1\leqslant i\leqslant n.$$

It therefore follows that $V(\Lambda)=V_0(\Lambda)\oplus V_1(\Lambda)$ can have at most two **Z**-graded levels, i.e., the level index d_Λ satisfies $d_\Lambda=0$ or 1. This latter requirement, together with the restriction $(\Lambda,\alpha_S)\geqslant 0$, Λ real, imposes stringent conditions on the allowed highest weights Λ. In fact (see Appendix A), Λ must have the special form

$$\Lambda=(\tau,-\omega,-\omega,\ldots,-\omega|\omega),\quad\omega\text{ real},\quad\tau+\omega\in\mathbf{Z}^+.$$

Conversely, such a two-level module $V(\Lambda)$ is easily seen to be both type (1) star and type (1) grade star since, with the above inner product, we have

$$(E^i)^\ddagger=(E^i)^\dagger=E_i,\quad(E_i)^\ddagger=-E_i^\dagger=-E^i.$$

We thus arrive at the following theorem.

Theorem 3: For $\Lambda\in D_+$, $V(\Lambda)$ is an irreducible type (1) grade * gl(n|1) module if and only if Λ has the form

$$\Lambda=(\tau,-\omega,\ldots,-\omega|\omega),\quad\omega\text{ real},\quad\tau+\omega\in\mathbf{Z}^+.$$

In such a case $V(\Lambda)$ admits at most two **Z**-graded levels and is also a type (1) * module. □

We remark that the only one-level type 1 grade * module $V(\Lambda)$ is given by the case $\tau=-\omega$ (i.e., $\tau+\omega=0$) above. For this case $V(\Lambda)$, $\Lambda=(-\omega,\ldots,-\omega|\omega)$, $\omega\in\mathbf{R}$, is a trivial one-dimensional module and is both a type (1) and type (2) * and grade * module.

For the remaining cases $\tau+\omega\in N$ (natural numbers), $V(\Lambda)$ gives rise to a two-level type (1) * and grade * module. In view of Proposition 1 and 2 and Theorem 3, $V^*(\Lambda)$ then gives rise to a two-level type (2) * module which is also a grade * module of type (1) but whose maximal **Z**-graded component is regarded as *odd*: i.e., $V^*(\Lambda)$ gives rise to a type (2) grade * module with the standard choice of $\mathbf{Z}_2$ grading. The minimal weight of $V(\Lambda)$ is easily seen to be[17]

$$(-\omega,-\omega,\ldots,\tau-1|\omega+1),\quad\tau+\omega\in N,$$

from which it follows that $V^*(\Lambda)$ has the highest weight

$$(\omega,\omega,\ldots,1-\tau|-\omega-1),\quad\tau+\omega\in N.$$

Thus if Λ is of the above form or equivalently,

$$\Lambda=(-(\omega+1),\ldots,-(\omega+1),\tau-1|\omega),\quad\tau+\omega\in\mathbf{Z}^-,\tag{13}$$

then $V(\Lambda)$ is a type (2) * and grade * module with two levels.

Conversely, if $V(\Lambda)$ is a two-level type 2 grade module, then $V^*(\Lambda)$ gives rise to a two-level type (1) grade * module whose highest weight must satisfy the conditions of Theorem 3; it is easily seen that this occurs if and only if Λ is of the form given in Eq. (13). We thus arrive at the following proposition.

Proposition 7: For $\Lambda\in D_+$, the following conditions are equivalent: (i) $V(\Lambda)$ is a type (2) grade * gl$(n|1)$ module having at most two levels, (ii) Λ is of the form $\Lambda=(-\omega,...,-\omega|\omega)$ or $\Lambda=(-(\omega+1),...,-(\omega+1),\tau-1|\omega)$, $\omega\in\mathbb{R}$, $\tau+\omega\in\mathbb{Z}^-$.

Corollary: All irreducible type 2 grade * modules with at most two levels are also type (2) * modules. □

The above result characterizes all irreducible type (2) grade * modules having at most two $\mathbb{Z}$-graded levels. To complete the classification we now assume that $V(\Lambda)$ is a type (2) grade * module with at *least* three levels, so that Λ is real and $V(\Lambda)$ is equipped with a (positive definite) inner product $\langle|\rangle$ which is invariant of type $\theta=\epsilon=1$, viz. for all $v, w\in V(\Lambda)$,

$$\langle E_i v|w\rangle = -(-1)^{(v)}\langle v|E^i w\rangle,$$
$$\langle E^i v|w\rangle = (-1)^{(v)}\langle v|E_i w\rangle.$$

Furthermore, for v^Λ, the highest weight vector of $V(\Lambda)$, we must have

$$E_1 v^\Lambda \neq 0, \quad E_1E_2v^\Lambda\neq 0;$$

otherwise, we would have $E_iE_jv^\Lambda=0 (1\leqslant i,j\leqslant n)$, contradicting the assumption that $V(\Lambda)$ has at least three $\mathbb{Z}$-graded levels. Equivalently, we have

$$\langle E_1v^\Lambda|E_1v^\Lambda\rangle>0,\quad \langle E_1E_2v^\Lambda|E_1E_2v^\Lambda\rangle>0. \tag{14}$$

Now from Eqs. (14) we obtain

$$0<\langle E_1v^\Lambda|E_1v^\Lambda\rangle = -\langle v^\Lambda|E^1E_1v^\Lambda\rangle$$
$$-(\Lambda,\epsilon_1-\delta_1)\langle v^\Lambda|v^\Lambda\rangle,$$

from which we deduce $(\Lambda,\epsilon_1-\delta_1)<0$ and hence,

$$(\Lambda,\epsilon_i-\delta_1)<0,\quad 1\leqslant i\leqslant n. \tag{15}$$

Note that since

$$\langle E_iv^\Lambda|E_iv^\Lambda\rangle = -\langle v^\Lambda|E^iE_iv^\Lambda\rangle$$
$$= -(\Lambda,\epsilon_i-\delta_1)\langle v^\Lambda|v^\Lambda\rangle,$$

Eq. (15) implies that

$$E_iv^\Lambda\neq 0,\quad 1\leqslant i\leqslant n. \tag{16}$$

Also, we have

$$0\leqslant\langle E_{n-1}E_nv^\Lambda|E_{n-1}E_nv^\Lambda\rangle$$
$$=\langle E_nv^\Lambda|E^{n-1}E_{n-1}E_nv^\Lambda\rangle$$
$$=[(\Lambda,\epsilon_{n-1}-\delta_1)+1]\langle E_nv^\Lambda\langle E_nv^\Lambda\rangle.$$

However, Eq. (16) implies that $\langle E_nv^\Lambda|E_nv^\Lambda\rangle>0$, from which we deduce $(\Lambda,\epsilon_{n-1}-\delta_1)\geqslant-1$ and hence,

$$(\Lambda,\epsilon_i-\delta_1)\geqslant-1,\quad 1\leqslant i<n. \tag{17}$$

We are now in a position to show that inequality (17) implies the $V(\Lambda)$ must have *exactly* three levels. For this it suffices to demonstrate that $E_1E_2E_3v^\Lambda=0$: Note that for $n\leqslant 2$ there is nothing to prove since all irreps have no more than three levels. Now for $n>2$, $E_1E_2E_3v^\Lambda\neq 0$ implies

$$0<\langle E_1E_2E_3v^\Lambda|E_1E_2E_3v^\Lambda\rangle$$
$$=-[(\Lambda,\epsilon_1-\delta_1)+2]\langle E_2E_3v^\Lambda|E_2E_3v^\Lambda\rangle,$$

from which we deduce $(\Lambda,\epsilon_1-\delta_1)<-2$, in contradiction to Eq. (17). Thus we must have $E_1E_2E_3v^\Lambda=0$, from which it follows that $V(\Lambda)$ can have no more than three levels and hence, under our assumptions, exactly three levels.

We have thus shown that an irreducible type (2) grade * module $V(\Lambda)$ with greater than two levels must have exactly three levels, with highest weight Λ satisfying

$$-1\leqslant(\Lambda,\epsilon_i-\delta_1)<0,\quad 1\leqslant i<n,\quad (\Lambda,\epsilon_n-\delta_1)<0.$$

Note that since the components of Λ differ by integers, the last inequality implies

$$\lambda_1=\lambda_2=\cdots=\lambda_{n-1},\quad \Lambda=(\lambda|\omega).$$

Moreover, from Eq. (14) we have

$$0<\langle E_1E_2v^\Lambda|E_1E_2v^\Lambda\rangle$$
$$=[(\Lambda,\epsilon_1-\delta_1)+1]\langle E_2v^\Lambda|E_2v^\Lambda\rangle,$$

from which it follows that $(\Lambda,\epsilon_1-\delta_1)>-1$. We thereby deduce that Λ must satisfy

$$-1<(\Lambda,\epsilon_i-\delta_1)<0,\quad 1\leqslant i<n,\quad (\Lambda,\epsilon_n-\delta_1)<0.$$

The above equation demonstrates that Λ must be typical and since three level typical irreducible gl$(n|1)$ modules exist only for $n=2$, it follows that $V(\Lambda)$ must be a three-level typical type (2) grade * gl$(2|1)$ module with the highest weight $\Lambda=(\lambda_1,\lambda_2|\omega)$ satisfying

$$-1<\lambda_1+\omega<0,\quad \lambda_2+\omega<0. \tag{18}$$

On the other hand, the above implies that $V^*(\Lambda)$ must also give rise to a three-level typical type (2) grade * module whose highest weight $\Lambda^*=(1-\lambda_2,1-\lambda_1|-w-2)$ must have components which also satisfy Eq. (18): This then yields the additional constraints

$$-1<\lambda_2+\omega<0,\quad \lambda_1+\omega<0,$$

from which we deduce that $\Lambda=(\lambda_1,\lambda_2|\omega)$ must satisfy

$$-1<\lambda_i+\omega<0,\quad i=1,2.$$

Since the components λ_i differ by integers we must have $\lambda_1=\lambda_2=\tau$ and Λ has the special form

$$\Lambda=(\tau,\tau|\omega),\quad \omega\in\mathbb{R},\quad -1<\tau+\omega<0.$$

Conversely, it can be shown (see Appendix B) that with Λ as above, $V(\Lambda)$ indeed gives rise to a (four-dimensional) typical type (2) grade *gl$(2|1)$ module: We note that such a module $V(\Lambda)$ is not a * module. We have thus proved the following.

Proposition 8: Type (2) grade * irreps with greater than two levels exist only for gl$(2|1)$. Such irreps have highest weights of the form

$$\Lambda=(\tau,\tau|\omega),\quad \omega\in\mathbb{R},\quad -1<\tau+\omega<0$$

and are four-dimensional and typical. These are the only grade * irreps that are not * representations. □

Propositions 7 and 8 thus yield the following classification scheme.

Theorem 4: $V(\Lambda)$ is an irreducible type (2) grade * gl$(n|1)$ module if and only if Λ has one of the following special forms: (i) $\Lambda=(-\omega,...,-\omega|\omega),\omega\in\mathbb{R}$; (ii) $\Lambda=(-(\omega+1),...,-(\omega+1),\tau-1|\omega),\omega\in\mathbb{R},\tau+\omega\in\mathbb{Z}^-$; (iii) for $n=2$, $\Lambda=(\tau,\tau|\omega)$, $\omega\in\mathbb{R}$, $-1<\tau+\omega<0$.

Corollary 1: For $n \neq 2$, $V(\Lambda)$ is a type (2) grade * module if and only if $V^*(\Lambda)$ is grade * of type (1).

Corollary 2: Typical grade * irreps exist only for $gl(1|1)$ and $gl(2|1)$.

VI. CONCLUSIONS

We have obtained a complete classification of the * and grade * irreps of $gl(n|1)$. The main results are summarized in Theorems 1–4, which give a classification, in terms of highest weights, of the type (1) and (2) * and grade * irreps, respectively. In particular, it was shown that an irrep of $gl(n|1)$ with the highest weight Λ is a type (1) * if and only if Λ is real and $(\Lambda,\alpha_S) \geqslant 0$ and that type (2) * irreps are duals to type 1 * irreps. It follows that for $gl(n|1)$, there exists a large class of * irreps. The situation with grade * irreps is quite different and for $n \neq 2$, only a small class of grade * irreps exist: Such irreps have at most two $\mathbb{Z}$-graded levels and are also * irreps. The only exception is $gl(2|1)$, which admits an additional two-parameter family of four-dimensional typical grade * irreps which are not * irreps. It is interesting that the example of a $gl(2|1)$ grade * irrep considered in the work of Scheunert *et al.*[20] belongs to this class.

The results of this paper indicate that grade * irreps are relatively rare and as such are unlikely to be of importance in physical applications. It would be of great interest to determine if this situation prevails for all the simple basic classical Lie superalgebras. It would also be of interest to extend the approach of this paper to investigate * and grade * irreps arising from noncompact real forms of $gl(n|1)$. We expect that such irreps will be infinite-dimensional, in which case the infinitesimal character of an irrep may be used in place of a highest weight label.

It can be shown[19] that unlike the grade * case, the tensor product of two * modules is again a * module, which opens up the interesting possibility of determining tensor product branching rules and the corresponding Wigner coefficients for $gl(n|1)$. However, it is important to note that while the tensor product of two irreducible type (1) [or (2)] * modules is again a * module of type (1) [resp. (2)], the tensor product of a type (1) * module with a type (2) * module does *not* yield a * module. This restriction may therefore limit the use of Young diagram methods for * modules.

Finally, it would be of interest to extend the results of this paper to the Lie superalgebras $gl(m|n)$ and $osp(m|n)$ and investigate character formulas, branching rules, matrix elements, etc. for star and grade star modules. In the second paper of this series a start in this direction is made with the classification of the star and grade star irreps of the Lie superalgebra $C(n) = osp(2|2n)$.

ACKNOWLEDGMENTS

The authors take great pleasure in thanking Dr. A. J. Bracken for his continued encouragement and many hours of in depth discussions. We also gratefully acknowledged helpful discussions with Dr. P. D. Jarvis, Professor H. S. Green, and Professor T. D. Palev.

The support of an ARC research grant, without which this work would not have been completed, is also gratefully acknowledged.

APPENDIX A: CLASSIFICATION OF TWO-LEVEL MODULES

Here we classify all irreducible $gl(n|1)$ modules $V(\Lambda)$ having at most two levels. The $\mathbb{Z}$ gradation of $V(\Lambda)$ may be written as[17]

$$V(\Lambda) = \bigoplus_{k=0}^{d_\Lambda} V_k(\Lambda),$$

where $d_\Lambda = n$ for Λ typical; when $(\Lambda+\rho,\epsilon_i - \delta_1) = 0$, we have $d_\Lambda = |I_\Lambda|$, where

$$I_\Lambda = \{j > i\} \cup \{j < i | (\Lambda, \epsilon_j - \epsilon_i) > 0\}.$$

Now $V(\Lambda)$ has exactly one level when $d_\Lambda = 0$, which can only occur when $(\Lambda+\rho,\epsilon_n - \delta_1) = 0$ and $(\Lambda,\epsilon_i - \epsilon_n) = 0$ for all $i < n$: For $\Lambda = (\lambda|\omega)$ this implies

$$\lambda_1 = \lambda_2 = \cdots = \lambda_n = -\omega,$$

so that the one-level irreps of $gl(n|1)$ have highest weights of the special form

$$\Lambda = (-\omega,\ldots,-\omega|\omega).$$

For two-level irreps we have $d_\Lambda = 1$ and there are only two possibilities:

(i) $(\Lambda+\rho,\epsilon_n - \delta_1) = 0, \quad (\Lambda,\epsilon_i - \epsilon_n) = 0,$

$1 < i < n, (\Lambda,\epsilon_1 - \epsilon_n) > 0;$

(ii) $(\Lambda+\rho,\epsilon_{n-1} - \delta_1) = 0, \quad (\Lambda,\epsilon_i - \epsilon_{n-1}) = 0,$

$1 \leqslant i < n-1.$

In case (i) $\Lambda = (\lambda|\omega)$ has the special form

$$\Lambda = (\tau, -\omega,\ldots,-\omega|\omega), \quad \tau + \omega \in N$$

and in case (ii) Λ has the special form

$$\Lambda = (-(\omega+1),\ldots,-(\omega+1),\tau|\omega), \quad -(\tau+\omega) \in N.$$

It follows that $V(\Lambda)$ is an irreducible $gl(n|1)$ module having at most two levels if and only if Λ has one of the following forms: (i) $\Lambda = (\tau,-\omega,\ldots,-\omega|\omega)$, $\tau+\omega \in \mathbb{Z}^+$; (ii) $\Lambda = (-(\omega+1),\ldots,-(\omega+1),\tau-1|\omega)$, $\tau+\omega \in \mathbb{Z}^-$. We note from Theorem 3 that the first gives rise to type (1) * and grade * modules, while Proposition 7 implies that the second class gives rise to type (2) * and grade * modules.

APPENDIX B: TYPICAL GRADE * IRREPS OF gl(2|1)

Here we investigate the irreducible typical type (2) grade * $gl(2|1)$ modules $V(\tau,\tau|\omega)$, $0 > \tau + \omega < -1$ explicitly: We set $\alpha = -(\tau+\omega)$, so that $1 > \alpha > 0$. We have a $gl(2) \oplus gl(1)$ module decomposition

$$V(\tau,\tau|\omega) = V_0(\tau,\tau|\omega) \oplus V_0(\tau,\tau-1|\omega+1) \oplus V_0(\tau-1,\tau-1|\omega+2).$$

The top level is a one-dimensional $gl(2) \oplus gl(1)$ module, with the basis vector e_0 satisfying

$$E^i_j e_0 = \tau \delta^i_j e_0, \quad \Omega e_0 = \omega e_0.$$

The next level gives a two-dimensional module, with the (normalized) basis vectors

$$\bar{e}_i = \alpha^{-1/2} E_i e_0, \quad i = 1,2.$$

The bottom level $V_0(\tau - 1, \tau - 1|\omega + 2)$ is again one-dimensional, with the (normalized) basis vector

$$\bar{\bar{e}}_0 = [\alpha(1-\alpha)]^{-1/2} E_1 E_2 e_0.$$

Thus we have a four-dimensional module with the basis vector $e_0, \bar{e}_1, \bar{e}_2, \bar{\bar{e}}_0$, where $e_0, \bar{\bar{e}}_0$ are *even* vectors and $\bar{e}_1, \bar{e}_2$ are *odd* vectors. The action of our gl(2|1) generators on these basis states are given by

$$E^i_j \bar{e}_k = \tau \delta^i_j \bar{e}_k - \delta^i_k \bar{e}_j, \quad \Omega \bar{e}_k = (\omega + 1)\bar{e}_k, \quad k = 1,2;$$
$$E^i_j \bar{\bar{e}}_0 = (\tau - 1)\delta^i_j \bar{\bar{e}}_0, \quad \Omega \bar{\bar{e}}_0 = (\omega + 2)\bar{\bar{e}}_0;$$
$$E_i e_0 = \alpha^{1/2} \bar{e}_i, \quad E_1 \bar{e}_2 = (1-\alpha)^{1/2} \bar{\bar{e}}_0,$$
$$E_2 \bar{e}_1 = -(1-\alpha)^{1/2} \bar{\bar{e}}_0;$$
$$E_1 \bar{e}_1 = E_2 \bar{e}_2 = E_1 \bar{\bar{e}}_0 = E_2 \bar{\bar{e}}_0 = 0;$$
$$E^i e_0 = E^1 \bar{e}_2 = E^2 \bar{e}_1 = 0, \quad E^1 \bar{e}_1 = -\alpha^{1/2} e_0,$$
$$E^2 \bar{e}_2 = -\alpha^{1/2} e_0;$$
$$E^2 \bar{\bar{e}}_0 = -(1-\alpha)^{1/2} \bar{e}_1, \quad E^1 \bar{\bar{e}}_0 = (1-\alpha)^{1/2} \bar{e}_2.$$

Thus in the basis $e_0, \bar{e}_1, \bar{e}_2, \bar{\bar{e}}_0$ we have the following 4×4 matrix representations for the odd generators:

$$\pi(E_1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \alpha^{1/2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & (1-\alpha)^{1/2} & 0 \end{pmatrix},$$

$$\pi(E_2) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \alpha^{1/2} & 0 & 0 & 0 \\ 0 & (1-\alpha)^{1/2} & 0 & 0 \end{pmatrix},$$

$$\pi(E^1) = \begin{pmatrix} 0 & -\alpha^{1/2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-\alpha)^{1/2} \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\pi(E^2) = \begin{pmatrix} 0 & 0 & -\alpha^{1/2} & 0 \\ 0 & 0 & 0 & -(1-\alpha)^{1/2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrices of the even generators may be similarly determined; it is easily verified that the above matrices indeed give rise to an (irreducible) representation.

Since the above matrices are real we note that their grade adjoint corresponds with their super transpose. Therefore, in order to check that we have a type (2) grade * irrep it remains to show that

$$\pi(E_i)^T = \pi(E^i), \quad \pi(E^i)^T = -\pi(E_i), \quad 1 \leqslant i \leqslant 2,$$

where

$$\pi(E)^T_{\alpha\beta} = (-1)^{(\beta)} \pi(E)_{\beta\alpha}, \quad 1 \leqslant \alpha, \beta \leqslant 4,$$

where we define $(\beta) = 0$ for $\beta = 1,4$ and $(\beta) = 1$ for $\beta = 2,3$, in accordance with our prescription that $e_0, \bar{\bar{e}}_0$ are even and $\bar{e}_1, \bar{e}_2$ are odd vectors. By direct calculation we have

$$\pi(E_1)^T = \begin{pmatrix} 0 & -\alpha^{1/2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-\alpha)^{1/2} \\ 0 & 0 & 0 & 0 \end{pmatrix} = \pi(E^1),$$

$$\pi(E_2)^T = \begin{pmatrix} 0 & 0 & -\alpha^{1/2} & 0 \\ 0 & 0 & 0 & -(1-\alpha)^{1/2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \pi(E^2);$$

similarly, $\pi(E^i)^T = -\pi(E_i)$, $i = 1,2$. Thus the above indeed gives rise to a type (2) grade * irrep of gl(2|1), as required. Note, however, that π is not a * irrep.

[1] J. Wess, and B. Zumino, Nucl. Phys. B **70**, 39 (1974).
[2] F. Iachello, Phys. Rev. Lett. **44**, 772 (1980).
[3] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[4] Y. Nambu, Physica D **15**, 147 (1985).
[5] A. Montorsi, M. Rasetti, and A. I. Solomon, Phys. Rev. Lett. **59**, 2243 (1987).
[6] V. A. Kostelecky and D. K. Campbell, Eds., Physica D **15**, 3–294 (1985).
[7] V. G. Kac, Adv. Math. **26**, 8 (1977).
[8] V. G. Kac, *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), p. 597.
[9] P. H. Dondi and P. D. Jarvis, J. Phys. A: Math. Gen. **14**, 547 (1981).
[10] R. J. Farmer and P. D. Jarvis, J. Phys. A: Math. Gen. **16**, 473 (1983); **17**, 2365 (1984).
[11] A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1810 (1981); **23**, 1239 (1982).
[12] R. C. King, *Lecture Notes in Physics*, Vol. 180 (Springer, Berlin, 1983), p. 41.
[13] J. P. Hurni, J. Phys. A: Math. Gen. **20**, 5755 (1987).
[14] J. W. B. Hughes, J. Math, Phys. **22**, 245 (1981).
[15] J. Van der Jeugt, J. Phys. A: Math. Gen. **20**, 209 (1987).
[16] J. P. Hurni and B. Morel, J. Math. Phys. **22**, 2236 (1982); **24**, 157 (1983).
[17] M. D. Gould, J. Phys. A: Math. Gen. **22**, 1209 (1989); M. D. Gould, A. J. Bracken, and J. W. B. Hughes, J. Phys. A: Math. Gen. **22**, 2879 (1989).
[18] J. W. B. Hughes and R. C. King, J. Phys. A: Math. Gen. **20**, L1047 (1987).
[19] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[20] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[21] J. Van der Jeugt, J. Math. Phys. **26**, 913 (1985).
[22] W. Heidenreich, Phys. Lett. B **110**, 461 (1982).
[23] J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984).
[24] M. De Crombrugghe and V. Rittenberg, J. Math Phys. **18**, 146 (1977).
[25] J. Van der Jeugt, J. Math. Phys. **28**, 758 (1987).
[26] J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, Berlin, 1972).
[27] W. Nahm and M. Scheunert, J. Math. Phys. **6**, 868 (1976).

Representations of the diffeomorphism group describing an infinite Bose gas in the presence of ideal vortex filaments

Uwe K. Albertin and Harry L. Morrison
Department of Physics, University of California, Berkeley, California 94720

(Received 26 September 1989; accepted for publication 24 January 1990)

A number of representations of the group of volume preserving diffeomorphisms of $M = S^3$ in the finite-volume case and $M = R^3$ in the infinite-volume case that describe a Bose gas in the presence of a vortex filament are examined. The core of the vortex is taken to lie along a curve γ. Let Dif $f_\mu^\gamma(M)$ be the group of volume preserving diffeomorphisms that map γ onto itself, M_γ the manifold M with γ removed, and Map(M_γ,T) the space of smooth maps of M_γ into the complex numbers of modulus unity. It is shown, using the formalism of Klauder [J. Math. Phys. **11**, 233 (1970)], that an infinite number of inequivalent representations of Dif $f_\mu^\gamma(M)$ exist in the infinite-volume limit using cyclic vectors, which are coherent states based on elements in Map(M_γ,T). It is found, however, that these states have questionable physical significance since they appear to violate the continuity equation for quantum probability flow. This leads to the postulation that Dif $f_\mu^\gamma(M)$ acts as gauge group. The representations of Dif $f_\mu^\gamma(M)$ are then used to construct induced representations of the full group volume preserving diffeomorphisms of M, which are realized in the Hilbert space over the space of unparametrized loops in M in a natural way. These latter representations are found to be essentially the ones under current study in quantum hydrodynamics [G. A. Goldin, R. Menikoff, and D. H. Sharp, Phys. Rev. Lett. **58**, 2162 (1987)].

I. INTRODUCTION

The nonrelativistic current algebra associated with global gauge symmetry has for some time been used to describe a number of physical systems.[1–6] Since the algebra is a homomorphism on the algebra of smooth vector fields over the manifold of interest, one may instead examine representations of the group of diffeomorphisms of the manifold in order to describe nonrelativistic quantum mechanics. The derivative of an irreducible representation of the group of diffeomorphisms yields a representation of the algebra of smooth vector fields, the operators of which form a complete set of observables for the system. Hence, one has the identification of distinct irreducible representations of the group of diffeomorphisms and distinguishable nonrelativistic quantum systems.

In this paper we restrict attention to the subgroup of volume preserving diffeomorphisms and a simple set of representations of this group that describe the flow surrounding an ideal vortex filament in a Bose gas at zero temperature. The Bose gas shall be considered to be in $M = R^3$ or $M = S^3$ depending on whether a finite or infinite volume is being considered. Let γ denote the curve representing the core of the vortex and let M_γ be the manifold M with the curve γ removed. We take the space of smooth maps from M_γ to the complex numbers of modulus unity, denoted Map(M_γ, T), as a space of possible order parameters describing the condensate. Let Dif $f_\mu^\gamma(M)$ denote the subgroup of volume preserving diffeomorphisms that leave the position of the γ curve unchanged. We first examine regular representations of Dif $f_\mu^\gamma(M)$ that utilize elements of Map(M_γ, T) as cyclic vectors in $L^2(M)$, where the volume of M is finite: These representations are found to be equivalent. However, in the infinite-volume limit we examine exponential representations of Dif $f_\mu^\gamma(M)$ utilizing coherent states based on elements of Map(M_γ, T) as cyclic vectors and find, using the formalism of Klauder,[7] an infinite number of inequivalent representations. We derive a simple criterion that distinguishes these representations.

We then examine the generating functional for these representations, which is essentially the expectation value of the operators of the representation in the cyclic vector, and derive a functional differential equation for the generating functional. The generating functional may be used to calculate current–current correlation functions for the Bose gas in the presence of a vortex filament, and we illustrate the relation of this functional to the path integral.

Next, we examine the elements $e^{i\theta(x)}$ of Map(M_γ, T) and note the interesting result that if they are used as single-particle wavefunctions, one finds that they do not conserve probability flow unless the multiply valued function θ has a gradient that is divergence-free. This restricts the possible elements of Map(M_γ, T) that may be used as physical order parameters, and leads to the conclusion that the action of Dif $f_\mu^\gamma(M)$ should be similar to that of a gauge group. One is then led to examine operators for describing the vortex that depend only on the position of the core and not on the flow surrounding it; these are the vortex operators of Rasetti and Regge[8] and we give a coordinate-free derivation of these operators.

Finally, noting that the action of Dif $f_\mu^\gamma(M)$ is similar to that of a gauge group, we are led to examine induced representations of the full group Dif$f_\mu(M)$ induced from Dif $f_\mu^\gamma(M)$, whose carrier space becomes the space of square integrable functions over the space of unparametrized loops. These representations are found to closely resemble those recently treated in Ref. 4.

The paper is organized as follows. In Sec. II, we intro-

 0022-2488/90/061535-09$03.00

duce the current algebra and representations of Dif $f_\mu^\gamma(M)$ in finite volume. We then take the infinite-volume limit in Sec. III and discuss the generating functionals in Sec. IV. The quantum vortex operators are derived in Sec. V and the induced representations of Dif$f_\mu(M)$ induced from representations of Dif $f_\mu^\gamma(M)$ are obtained in Sec VI.

II. EXPONENTIATION OF THE ALGEBRA AND THE DIFFEOMORPHISM GROUP

We shall restrict attention in this section to a Bose gas initially in $M = S^3$ and endowed with a positive metric such that its volume is finite. We shall further treat the current operators as one-forms given by

$$J(x) = \frac{1}{2i}\left[\phi^\dagger \frac{\partial}{\partial x^i}\phi - \frac{\partial}{\partial x^i}\phi^\dagger\phi\right] dx^i . \tag{2.1}$$

Here, the fields $\phi(x)$ are either boson or fermion fields satisfying canonical commutation or anticommutation relations, respectively. We shall restrict attention to representations that describe the Bose gas. The $J(x)$ are to be considered as operator valued one-forms acting in some Hilbert space. We let Ξ denote the algebra of smooth vector fields on S^3 and let Ξ_δ denote the subalgebra of smooth divergence-free vector fields. We consider the smeared current operators, which may be viewed as operators obtained when pairing $J(x)$ with an element g in Ξ and integrating with regard to the volume form $d\mu$. We shall denote the resulting oeprator by $J(g)$. Hence, more precisely, J is an element of the space of linear functionals on Ξ taking values in the space of operators of some Hilbert space, and thus is a generalized one-form in the dual to the Lie algebra. In local coordinates we may think of $J(x)$ as a one-form whose components are distributions. With this definition, $J(x)$ may also be regarded as an operator valued current in the De Rham sense.

Note that since there is a metric on M, there is a natural way to assign to every vector field $g \in \Xi$ a one-form via the metric. We shall denote this one-form by the same letter g. Then we may write the operator $J(g)$ as

$$J(g) = \int_M J\wedge *g = \int_M J\wedge i_g\, d\mu , \tag{2.2}$$

where g is considered a one-form in the first expression; $*$ denotes the Hodge dual; and g denotes a vector field in the second expression, where i_g is interior product by the vector field g.

The operators $J(g)$ satisfy the commutation relations

$$[J(g),J(f)] = iJ([g,f]) , \tag{2.3}$$

where $[g,f]$ denotes the bracket on vector fields. The exponentiated operators $e^{itJ(g)} = U(\psi_t)$ then form representations of one-parameter groups of diffeomorphisms generated by g. The set of all such operators as g ranges over Ξ_δ form a representation of the group of volume preserving diffeomorphisms Dif$f_\mu(M)$.

We have the regular representation $U(\psi)$ of Dif$f_\mu(M)$ in $L^2(M)$ given by

$$[U(\psi)f](x) = f(\psi^{-1}x) . \tag{2.4}$$

If ψ_t is a one-parameter group of volume preserving diffeomorphisms generated by g, we have immediately that the action of $J(g)$ in $L^2(M)$ is just by Lie differentiation with respect to the vector field $-g$. This corresponds to the action of the quantum current operator $J(x)$ in the single-particle Hilbert space.

Recall that an ideal Bose gas at zero temperature has all its particles in a single quantum state $f(x)$, which is taken as the order parameter for the system when the number of particles becomes sufficiently large. Suppose an ideal vortex exists in this system, whose core lies along a smooth curve γ. Then the order parameter may be written as $e^{i\theta(x)}$ if the density is taken as unity and the flow of the condensate is given by the gradient of $\theta(x)$. The curl of this flow field then has support of γ and vanishes elsewhere.

Before proceeding to the case where the number of particles becomes large, it is simpler first to discuss the single-particle case and consider $e^{i\theta}$ as a state in the single-particle space. We fix the volume of $M = S^3$ to be unity, so that the density is unity. For simplicity, we shall consider the case where γ is a closed unknotted curve in M. Denote by M_γ the manifold M with the curve γ removed. The quantum state $f(x) = e^{i\theta(x)}$ should be considered as a function on M_γ since θ is multiply valued on M_γ and ill defined on γ itself. Hence, $f(x)$ may be considered as an element of the set Map(M_γ,T). To each element $e^{i\theta(x)}$ in Map(M_γ,T) we may associate a global one-form α_θ, given in local coordinates by $\alpha_\theta = [(\partial/\partial x^i)\theta]dx^i$. Then, for any curve γ' that links γ exactly once, we have $\int_{\gamma'}\alpha_\theta = 2\pi n$ and $d\alpha_\theta = 0$, so that α_θ is in the nth De Rham cohomology class in $H^1(M,R)$. Notice that the expectation value of the operator $J(x)$ in the state $e^{i\theta}$ is just the one-form α_θ. The space Map(M_γ,T) forms a group under pointwise multiplication and is disconnected, its various components Map$_n(M_\gamma,T)$ being labeled by the above winding number. Since each De Rham cohomology class has a unique harmonic representative, there exists a unique element $e^{i\theta_n}$ in Map$_n(M_\gamma,T)$ such that α_n is harmonic. These states may be considered the most "vortexlike" single-particle states since they yield a probability flow [which is the expectation value of $J(x)$] that is divergence-free and whose curl has support on γ. Of course, it should be remembered that at this stage we are only considering a single particle.

Having isolated a closed curve γ in M, there are several possible subgroups of Dif$f(M)$ that are of interest. First there is the subgroup of diffeomorphisms Dif$f^\gamma(M)$ that map γ onto itself. Second, there is the subgroup Dif$f^{\gamma 0}(M)$ that leaves every point of γ unchanged. Finally, there is the subgroup Dif$f(M_\gamma)$, which are the diffeomorphisms of M_γ. Of course, corresponding to each of these groups is the corresponding subgroup of volume preserving diffeomorphisms, which we shall augment with the subscript μ. We note that Dif $f_\mu^\gamma(M)$ is a subgroup of Dif$f_\mu(M)$, but it is not a normal subgroup. We note further that Dif $f_\mu^{\gamma 0}(M)$ is a normal subgroup of Dif $f_\mu^\gamma(M)$. In fact, since an arbitrary diffeomorphism in Dif $f_\mu^\gamma(M)$ can yield an arbitrary smooth invertible mapping of γ onto itself when we restrict it to γ, we see that the quotient group Dif $f_\mu^\gamma(M)$/Dif $f_M^{\gamma 0}(M)$ is isomorphic to Dif $f(S^1)$, the diffeomorphisms of the circle. We also make the following remark. Consider the one-forms α_n that are harmonic on M_γ and the vector fields g_n associated to the α_n via the metric. Then, the one-parameter groups of diffeomorphisms generated by the vector fields g_n lie in Dif$f_\mu(M_\gamma)$. However, they do not lie in Dif $f_\mu^\gamma(M)$ for $n\neq 0$ since the vector fields g_n are no longer regular on the curve γ. To see this, it is easiest to consider $M = R^3$ momentarily endowed with the Euclidean metric and consider the case of the rectilinear vortex. We choose polar coordinates (r,χ,z) and let the vortex lie along the z axis. Then $\alpha_n = (n/r)d\chi$ and $g_n = (n/r)(\partial/\partial\chi)$, which is singular for $r=0$. An arbitrary element of Dif$f_\mu(M_\gamma)$ may be approximated arbitrarily closely by elements in Dif $f_\mu^\gamma(M)$. For example, the one-parameter groups generated by the g_n may be approximated by one-parameter groups generated by $b(x)g_n$, where $b(x)$ is a smooth function that vanishes on γ and rises quickly to unity within a small tubular neighborhood of γ.

We shall make use of the following theorem due to Vershik *et al.*[9] Let Y be an open connected submanifold of M with compact closure. Suppose there exists a smooth positive measure on M that is ergodic under the action of Dif $f_\mu(M)$. Let $\widetilde{L}^2(Y)$ be the space of complex square integrable functions with support in Y orthogonal to constants. Then $\widetilde{L}^2[Y]$ is irreducible under the representation $[U(\phi)f](x) = f(\phi^{-1}x)$ of volume preserving diffeomorphisms.

Since the Riemannian measure on M_γ is ergodic with respect to Dif$f_\mu(M_\gamma)$, we have in particular that the regular representation of Dif$f_\mu(M_\gamma)$ is irreducible on $L^2(M_\gamma)$ and, hence, on $L^2(M)$ since M and M_γ differ by a set of zero measure. One might try to find an invariant subspace of functions with nonzero support on γ itself, but since elements of an L^2 space are really equivalence classes of functions that agree up to sets of zero measure, these functions are all in the zero equivalence class. Further, since any element of Dif$f_\mu(M_\gamma)$ may be approximated arbitrarily closely by elements in Dif $f_\mu^\gamma(M)$, we conclude that the regular representation of Dif $f_\mu^\gamma(M)$ on $L^2(M)$ is irreducible as well.

Suppose we now consider the Hilbert space H_n given by choosing $e^{i\theta_n}$ as a cyclic vector for a representation U_n of Dif $f_\mu^\gamma(M)$, i.e., set $H_n =$ completion of span$[U_n(\psi)e^{i\theta_n}]$ in the L^2 norm, where $U_n(\psi)\exp[i\theta_n(x)] = \exp[i\theta_n(\psi^{-1}x)]$. Here, θ_n are the elements of Map$_n(M_\gamma,T)$ whose associated one-forms are harmonic. Since the cyclic vector $e^{i\theta_n}$ is in $L^2(M)$, however, and from the above result we know that the corresponding representations are irreducible, we have the interesting result that the Hilbert spaces H_n coincide with $L^2(M)$ and the representations U_n are unitarily equivalent to the representation U given in Eq. (2.4).

These results are easily generalized to the N-particle Hilbert space by taking tensor products of the above representations. The cyclic vector is taken as

$$\Omega = \prod_i \exp(i\theta_n(x_i)) \tag{2.5}$$

and the representation acts as

$$[U(\psi)f](x_1,\dots,x_n) = f(\psi^{-1}x_1,\dots,\psi^{-1}x_n)\,. \tag{2.6}$$

Once again the various representations coincide.

Hence, we see that if we use the states $e^{i\theta_n}\in$ Map$_n(M_\gamma,T)$ as cyclic vectors for representations of Dif $f_\mu^\gamma(M)$, all the representations are unitarily equivalent as long as we are considering a finite number of particles in a finite volume. However, we should remark that this does not mean that the states $e^{i\theta_n}$ that describe a vortexlike state are indistinguishable. In fact, they have quite different energies because the expectation value of the current operator in these states is just the one-form α_n and the energy is thus essentially the hydrodynamic energy $E_n = (\frac{1}{2})\int_M \alpha_n \wedge *\alpha_n$. This occurs because the unitary operator that relates the various representations does not commute with the Hamiltonian. Note that as long as M_γ has finite volume, the energy E_n is finite since the space of one-forms is a Hilber space, where the norm is precisely given by the expression $\|\alpha\|^2 = \int_M \alpha\wedge *\alpha$.

This situation changes in the infinite-volume limit, where the energies diverge. In this case, one finds representations of the diffeomorphism group that are inequivalent for differing vortex strengths, as we illustrate in Sec. III.

III. REPRESENTATIONS IN THE INFINITE-VOLUME LIMIT AND COHERENT STATES

We now turn to the study of representations of volume preserving diffeomorphisms describing a vortex filament in M in the infinite-volume limit. These representations are of the exponential type and may be conveniently obtained using the exponential Hilbert space formalism of Klauder,[7] which we briefly review.

Let $\mathbf{h}$ be a Hilbert space. We shall denote the vectors in $\mathbf{h}$ by lower-case letters. We define the exponential Hilbert space $\mathbf{H} = \mathbf{EXP}(\mathbf{h})$ as the completion of the span of vectors of the form

$$\Psi = 1 \oplus \psi \oplus (\tfrac{1}{2})(\psi\otimes\psi) \oplus (\tfrac{1}{3!})(\psi\otimes\psi\otimes\psi) \oplus \cdots\,. \tag{3.1}$$

It is known that $\mathbf{H}$ corresponds to the Fock space based on $\mathbf{h}$. The above states are nothing but the coherent states, i.e., eigenstates of the annihilation operator. Elements of $\mathbf{H}$ will be labeled by upper-case letters. The inner product in $\mathbf{H}$ is given by

$$(\Psi,\Psi') = \exp[-(\tfrac{1}{2})\|\psi\|^2 - (\tfrac{1}{2})\|\psi'\|^2 + (\psi',\psi)] \tag{3.2}$$

$$= NN'\exp[\psi',\psi)]\,, \tag{3.3}$$

with

$$N' = \exp[-(\tfrac{1}{2})\|\psi'\|^2]\,,\quad N = \exp[-(\tfrac{1}{2})\|\psi\|^2]\,. \tag{3.4}$$

The annihilation operators are defined as follows. For every ψ in h, we have an operator $\phi(\psi)$ on $\mathbf{H}$ defined by $\phi(\lambda)\Psi' = (\lambda,\psi')\Psi'$: These satisfy the commutation relations

$$[\phi(\lambda),\phi(\lambda')^\dagger] = (\lambda,\lambda')\,. \tag{3.5}$$

For any bounded operator b on $\mathbf{h}$, one can associate an operator B on $\mathbf{H}$ by

$$(\Psi',B\Psi) = NN' \exp(\psi',b\psi) . \tag{3.6}$$

Hence, B may be identified with

$$B = \oplus_{n=0}^{\infty} \otimes_{s=1}^{n} b . \tag{3.7}$$

If $B = \exp(-iWt)$ and $b = \exp(-iwt)$, where w is self-adjoint, so that B and b are unitary, we then have

$$(\Psi',W\psi) = (\psi',w\psi)(\Psi',\Psi) . \tag{3.8}$$

Let λ_n be an orthonormal basis of the Hilbert space $\mathbf{h}$. Then, from Eq. (3.8), we find that W is bilinear in the annihilation operators as

$$W = \sum_{nm} \phi(\lambda_n)(\lambda_n,w\lambda_m)\phi(\lambda_m) . \tag{3.9}$$

We use Klauder's shorthand notation of setting $W = (\phi,w\phi)$.

Let l be an element of an index set such that $\psi[l] \in h$ and let the set of such $\psi[l]$ form a total set. Consider two such sets of labeled vectors $\psi'[l]$ and $\psi[l]$ such that

$$\Psi'[l] = \exp[-i\,\mathrm{Im}(\xi,\psi)]\Psi[l] , \tag{3.10}$$

where ξ is in $\mathbf{h}$. Computing the inner product of $\Psi'[l_1]$ with $\Psi'[l_2]$, one finds

$$\begin{aligned}(\Psi'[l_1],\Psi'[l_2]) = &\exp[(-\tfrac{1}{2})\|\psi[l_1]-\xi\|^2] \\ &\times\exp[(-\tfrac{1}{2})\|\psi[l_2]-\xi\|^2] \\ &\times\exp[(\psi[l_1]-\xi,\psi[l_2]-\xi)] . \end{aligned} \tag{3.11}$$

Identifying the inner product with the expression in Eq. (3.3), we have the following identity that identifies $\psi'[l]$ as a translated set:

$$\psi'[l] = \psi[l] - \xi . \tag{3.12}$$

Note that $\mathrm{Im}(\xi,\psi[l]) = \mathrm{Im}(\xi,\psi'[l])$, so that either $\psi[l]$ or $\psi'[l]$ can be used in defining the phase in Eq. (3.12). We see that translated sets in $\mathbf{h}$ correspond to phase shifted sets in $\mathbf{H}$. From the action of the annihilation operator $\phi'(\lambda)$ on $\Psi'[l]$, we have the result

$$\phi'(\lambda) = \phi(\lambda) - (\lambda,\xi) , \tag{3.13}$$

which demonstrates the shift on the annihilation operator. Since ξ is in $\mathbf{h}$, we have the unitary equivalence between ϕ' and ϕ, which is accomplished using the unitary operator

$$V = \exp[\phi(\xi)^\dagger - \phi(\xi)] .$$

Inequivalent representations of the annihilation and creation operators are obtained when ξ is not in $\mathbf{h}$. This usually occurs when the norm of ξ becomes infinite as, for example, if ξ is the limit of a sequence of vectors in $\mathbf{h}$ whose norm goes to infinity. The translation in this case is termed improper. Note that even though ψ and ξ are no longer in $\mathbf{h}$, their difference still is by assumption. Thus, given a Fock representation $\phi(\lambda)$ of the annihilation and creation operators and two translated representations

$$\phi_1(\lambda) = \phi'(\lambda) + (\lambda,\xi_1) , \quad \phi_2(\lambda) = \phi'(\lambda) + (\lambda,\xi_2)$$

defined for a set λ dense in $\mathbf{h}$, then ϕ_1 and ϕ_2 are unitarily equivalent if and only if $(\xi_1 - \xi_2)$ is in $\mathbf{h}$.

Let G be a Lie group. Let $u(g)$ be a unitarily irreducible representation of G in $\mathbf{h}$ and let Ψ_0 be the vector in $\mathbf{H}$ based on the vector ψ_0. Then we have a continuous family of states in $\mathbf{H}$ given by $\Psi(g) = U(g)\Psi_0$. Here, $U(g)$ is related to $u(g)$ by Eq. (3.6). If w_i is the set of self-adjoint operators that generate the representation u, then we have the accompanying generators $W_i = (\phi,w_i\phi)$ that generate U with $\phi(\lambda)\Phi(g) = (\lambda,\phi(g))\Phi(g)$, as shown previously.

Suppose we now define a translated set of vectors as

$$\Psi'(g) = \exp[-i\,\mathrm{Im}(\xi,\psi'(g))]\Psi(g) , \tag{3.14}$$

where we take $\Psi'(g) = \exp[\psi'(g)]$. Then we have a new set of generators $W_i' = (\phi',w_i\phi')$ for a new representation U' of G, where $\phi'(\lambda)\Psi'(g) = (\lambda,\psi'(g))\Psi'(g)$. The annihilation operators ϕ and ϕ' are related by a translation, as in Eq. (3.13).

As shown previously, we arrive at consistency conditions for ξ that have two solutions. The first occurs for $\xi = 0$ and $\psi'(g) = u(g)\psi_0$. Then we have that the representations U and U' coincide and U is reducible. In particular, if u is irreducible, the representation U is the direct sum of irreducible representations on each n-particle subspace. This may be seen because the generators of the representation commute with the number operator. Hence, subspaces corresponding to a particular number are invariant and the representation is reducible. We note that the generating functional for the representation U may be written as

$$(\Psi_0,U(g)\Psi_0) = \exp(\psi_0,(u(g)-1)\psi_0) , \tag{3.15}$$

indicating the exponential nature of the representation.

The second solution occurs when $\xi = \psi_0$ and $\psi'(g) = (u(g)-1)\psi_0$. This corresponds to a translated representation U of G. As long as ψ_0 lies in $\mathbf{h}$, the representations U and U' will be equivalent since the annihilation operators that comprise the generators are equivalent. However, if ψ_0 lies outside of $\mathbf{h}$, the two representations U and U' will be inequivalent. These will be called representations obtained by improper translation.

Suppose that we are given two such representations U_1 and U_2 corresponding to different choices of ψ_0 labeled ξ_1 and ξ_2, respectively. Then these two representations are unitarily equivalent to one another if and only if $\xi_1 - e^{ic}\xi_2$ is in $\mathbf{h}$, i.e., if ϕ_1 is unitarily equivalent to $e^{ic}\phi_2$ for some constant c. The extra phase occurs here because the generators of the representations are bilinear in the annihilation and creation operators.

If the representation u is irreducible and the representation U is obtained by an improper translation, u is irreducible in general. This is because, on the carrier space for these representations, the number operator no longer exists. All the vectors in these spaces involve infinite superpositions of states with all possible particle numbers. What would be required for reducibility would be an operator $Y = (\phi,y\phi)$ that commutes with $W(X) = (\phi,w(X)\phi)$; instead, one finds that $[W(X),Y] = (\phi,[w(X),Y]\phi)$ and thus this operator does not exist as a result of the irreducibility of $u(g)$.

This concludes a brief summary of the Klauder treatment. We now construct representations of Dif $f_\mu^\gamma(M)$ in the infinite-volume limit using the above formalism. We shall set $M = R^3$ with the Euclidean metric, so that M has infinite volume. We set $\mathbf{h} = L^2(M)$ and use the cyclic vector $\Psi_0 = \exp(\psi_0)$, $\psi_0 \in \mathrm{Map}(M_\gamma,T)$. Let g be an element of

Dif $f_\mu^\gamma(M)$. The representation acts on Ψ_0 as in Eq. (3.14), yielding

$$U(g)\Psi_0 = \exp(+i(\psi_0,\psi'(g))\Psi'(g)) \tag{3.16}$$

and $\psi'(g) = (u(g)-1)\psi_0$. The representation U of Dif $f_\mu^\gamma(M)$ is obtained by an improper translation using ψ_0 since elements in Map(M_γ,T) are no longer in $L^2(M)$: It is irreducible since Dif $f_\mu^\gamma(M)$ acts irreducibly in $L^2(M)$. Recall that two such representations corresponding to ψ_0 and ψ_0' are equivalent if and only if $\psi_0 - e^{ic}\psi_0'$ is in $L^2(M)$. We write this condition as

$$\int d\mu |\psi_0 - e^{ic}\psi_0'|^2 < \infty . \tag{3.17}$$

We are considering the case where ψ_0 and ψ_0' are in Map(M_γ,T); thus we write $\psi_0 = e^{i\theta(x)}$, $\psi_0' = e^{i\theta'(x)}$ and set $f(x) = \theta(x) - \theta'(x) - c$. Then the requirement for equivalence of the representations becomes

$$\int d\mu \sin^2\left[\frac{f(x)}{2}\right] < \infty . \tag{3.18}$$

This gives a general condition for two exponential representations of Dif $f_\mu^\gamma(M)$ to be equivalent.

Let us return to the case of a rectilinear vortex and suppose that $f(x)$ depends only upon the polar angle χ. Then the requirement is that the integral

$$\int_0^{2\pi} d\chi \sin^2(f(\chi)) \tag{3.19}$$

is nonvanishing. This is true for any smooth $f(\chi)$ that is nonvanishing. Hence, we have an infinite number of inequivalent representations, each of which is labeled by various functions $f(\chi)$. In particular, the representations based on the canonical choices $e^{in\chi}$ in Map$_n(M_\gamma,T)$ are all inequivalent and these are based upon cyclic vectors that are coherent states describing the quantized vortex flow. Note that the expectation value of the current operator in the coherent state $\exp(e^{i\theta})$ is again the one-form α_n, as it was in the single-particle case. We remark that the statement of inequivalence of these representations is equivalent to the statement of symmetry breaking since the representations obtained by improper translation the annihilation operator has obtained a nonvanishing expectation value in the ground state, as seen in Eq. (3.13). We also remark that the representations found here are very similar to those discussed in Ref. 10.

IV. GENERATING FUNCTIONALS AND BROKEN SYMMETRY

Many representations of the diffeomorphism group are best characterized in terms of their generating functionals. In the case of the group of volume preserving diffeomorphisms, the generating functional is essentially the generator of correlation functions for the current operators in the cyclic vector. Let $L(\psi)$ be a complex valued functional on a Lie group G. Then there exists a Hilbert space: a continuous unitary representation $U(\psi)$ of G, with the cyclic vector Ω such that $L(\psi) = (\Omega, U(\psi)\Omega)$ if and only if L satisfies the conditions (i) L is continuous with respect to the topology of the group; (ii) $L(1) = 1$; and (iii) L is positive, i.e., the matrix given by $\|L(\psi_i^{-1}\psi_j)\|$, $i,j = 1,\ldots,V$, is positive definite.

For the regular representation of Dif $f_\mu^\gamma(M)$ on $L^2(M)$ with cyclic vector $e^{i\theta}$, the generating functional is of the form

$$L_1(\psi) = \int d\mu\, e^{-i\theta(x)} e^{i\theta(\psi^{-1}x)} , \tag{4.1}$$

where $\psi \in$ Dif $f_\mu^\gamma(M)$. Note that the integrand may be considered as a cocyle $\chi_\psi(x)$ satisfying the cocycle condition

$$\chi_\psi(x)\chi_{\psi'}(\psi^{-1}x) = \chi_{\psi\psi'}(x) . \tag{4.2}$$

However, we remark that since the integrand is expressed as a product, these cocycles are trivial in the sense of cocycles that appear in the theory of induced representations.[11] In fact, the trivial nature of these cocycles is an alternative proof that the representations U_n introduced in Sec. III are equivalent when M has finite volume. It is readily verified that the functional $L_1(\psi)$ satisfies conditions (i)–(iii). In general, two representations of G will be unitarily equivalent if their generating functionals are equal, but the converse is not necessarily true, as the representations U_n illustrate.

For the N-particle Hilbert space we have the generating functional given by

$$L_N(\psi) = (1/V)^N (L_1(\psi))^N , \tag{4.3}$$

where we have explicitly restored the volume factor. Taking the infinite-volume limit, we find the generating functional

$$L(\psi) = \exp\left[\int d\mu \{e^{-i\theta(x)} e^{i\theta(\psi^{-1}x)} - 1\}\right] . \tag{4.4}$$

Here, we have set the average density in the thermodynamic limit to be unity. It is readily verified that the generating functional (4.4) is the expectation value $(\Omega, U(\psi)\Omega)$ in the cyclic vector $\exp(e^{i\theta})$ and thus corresponds to the generating functional for an exponential representation, as may be seen from Eq. (3.15). We remark that L satisfies the functional differential equation

$$\frac{\delta L(\psi)}{\delta g(x)} = \alpha_\theta(x) e^{-i\theta(\psi x)} e^{i\theta(x)} L(\psi) , \tag{4.5}$$

where $\alpha_\theta(x)$ is the one-form corresponding to $e^{i\theta}$. We consider the functional derivative in this case as a one-form implicity defined by

$$\int_M \frac{\delta}{i\,\delta g(x)} L(\psi) \wedge *g = \frac{d}{dt} L(\psi_t \psi)|_{t=0} , \tag{4.6}$$

where $g(x)$ generates the one-parameter group of diffeomorphisms ψ_t when considered as a vector field. The derivatives of the functional $L(\psi)$ yield correlation functions for the current operators, as in Ref. 5. The functional differential equation (4.5) demonstrates the exponential nature of the generating functional. We further remark that as shown previously, the representations described by the functional L break topological symmetry in the following sense. By taking a functional derivative, we have immediately the initial condition $\langle J \rangle = \alpha_\theta$, so that the line integral of $\langle J(x) \rangle$ over γ' yields $2\pi n$, as required for a quantized vortex.

Parallel to this development, we may also write a path integral representation for the current correlation functions as follows. We let $\phi(\tau,x)$ be the thermal fields defined on $M \times S^1$. Then we may write the generating function as

$$L_p(g) = \int D[\phi^*]D[\phi]\exp - \frac{1}{\beta}\left[\int_0^\beta d\tau \int_M \frac{1}{2} d\phi \wedge *d\phi - \int_0^\beta d\tau\, J(g)\right]. \quad (4.7)$$

Here, g is in Ξ and $J(g)$ is as before, but with $J(x)$ written in terms of the thermal fields $\phi(\tau,x)$. We evaluate this by setting $\phi(\tau,x) = \phi_0(x) + \eta(\tau,x)$, where $\phi_0 = e^{i\theta}\in\mathrm{Map}(M_\gamma, T)$ and η represent fluctuations. The zeroth-order contribution in the zero-temperature limit is just

$$L_{p0}(g) = \exp\left[\int_M \alpha_\theta \wedge *g\right]. \quad (4.8)$$

Comparing with Eq. (4.4), we see that $L(\psi_t)$ reduces to $L_{p0}(g)$ when t is sufficiently small. Further, we know that $L_{p0}(g)$ generates normal ordered current–current correlation functions at zero temperature, neglecting quantum fluctuations. Taking two functional derivatives, one verifies immediately that in local coordinates

$$\frac{\delta}{dg^i(x)}\frac{\delta}{\delta g^j(y)} L(\psi)|_{\psi=1} = \partial_i\theta(x)\partial_j\theta(y) + (\partial_i\theta(x))(\partial_j\theta(x))\delta^3(x-y), \quad (4.9)$$

whereas

$$\frac{\delta}{\phi g^i(x)}\frac{\delta}{\delta g^j(y)} L_{p0}(g) = \partial_i\theta(x)\partial_j\theta(y). \quad (4.10)$$

The difference between (4.9) and (4.10) is precisely the contribution due to normal ordering. Hence, we have a direct connection between the generating functional for a representation of Dif $f^\gamma_\mu(M)$ and the path integral. Since only the zeroth-order term is taken into account, we conclude that the generating functionals (4.4) and (4.7) neglect fluctuations entirely, which is not surprising since they are based on coherent states.

V. THE HODGE DECOMPOSITION AND QUANTUM VORTEX OPERATORS

Having constructed some representations of the group Dif $f^\gamma_\mu(M)$ in Sec. IV, we still have the possibility of constructing representations of the full group Dif $f_\mu(M)$. We shall turn to this in Sec. VI, but for now we shall briefly discuss the role of various fluctuations in the system and the action of the various diffeomorphism groups.

From a physical viewpoint, one may postulate two types of fluctuations of the condensate when considering a quantum vortex. First, there are fluctuations that do not move the position of the core, but change the flow surrounding the vortex so that it is no longer harmonic. Such a fluctuation may be viewed as changing the coherent state from one based on the element $e^{i\theta_n(x)}$ in $\mathrm{Map}_n(M_\gamma,T)$, where α_n is harmonic, to one based on some other element $e^{i\theta(x)}$, where α is no longer harmonic and α_n and α are both one-forms on M_γ. It is interesting to note that this is precisely the action of Dif $f^\gamma_\mu(M)$ on these states, even if the diffeomorphism is volume preserving. To illustrate this, we examine the action of Dif $f^\gamma_\mu(M)$ on the vortexlike state $e^{i\theta(x)_n}$. An arbitrary element of Dif $f^\gamma_\mu(M)$ takes this state to the state $e^{i\theta_n(\psi^{-1}x)}$. The one-form α associated with the latter state does not satisfy $\delta\alpha = 0$ and thus does not give a flow that is divergence-free. This occurs because d commutes with pullback, but δ does not, even when ψ is a volume preserving diffeomorphism. We then have the following interesting observation. It then appears that the state $e^{i\theta_n(\psi^{-1}x)}$ violates the continuity equation since it has a time-independent probability density equal to unity everywhere except possibly at the core, but has a probability current that is not divergence-free. We then have two options. We could treat the continuity equation like an equation of motion and treat such fluctuations as going "off shell," as is done in field theory, or we can treat the action of Dif $f^\gamma_\mu(M)$ as something that does not really alter the physics. The former possibility has its drawbacks since the continuity equation arises as a conservation law for an internal symmetry and we expect it to hold as long as the symmetry is good. The second possibility requires that we treat the action of Dif $f^\gamma_\mu(M)$ as the action of a gauge group. This is in accord with the treatment of superfluid dynamics as a gauge theory, where the superfluid velocity is not an observable; only the vorticity is an observable. The analogy is that α would no longer be an observable; only $d\alpha$ would be an observable. Since $d\alpha = 0$ on M_γ, everything depends only upon the choice of γ. The action of Dif $f^\gamma_\mu(M)$ becomes like that of a gauge group if we note the following. The action of Dif $f^\gamma_\mu(M)$ is such that when it acts on $e^{i\theta_n}$ it never takes α_n out of its cohomology class. The resulting state $U(\psi)e^{i\theta_n} = e^{i\theta_n(\psi^{-1}x)}$ has an associated one-form α with a unique Hodge decomposition $\alpha = \alpha_n + df$ for some function f since it still satisfies $d\alpha = 0$ on M_γ. Here, f has compact support since ψ has compact support. Hence, we conclude that $U(\psi)e^{i\theta_n} = e^{if(x)}e^{i\theta_n(x)}$, so that the group Dif $f^\gamma_\mu(M)$ acts like a gauge group and the one-form α transforms like a connection under this action. We are thus led to look for representations of the full group Dif $f_\mu(M)$ that somehow isolate this action of the subgroup Dif $f^\gamma_\mu(M)$. We shall do this in Sec. VI.

Second, there are fluctuations that are fluctuations in the core position, but that do not change the harmonic nature of the flow. It is clear that any operators describing such fluctuations would depend only upon the position of the core of the vortex and not on the condition of the flow surrounding it. A natural candidate for these operators are the vortex operators of Rasetti and Regge,[8] and we now turn to their construction in coordinate-free language, making use of a Hodge decomposition for the operator $J(x)$.

Recall that Hodge theory gives a decomposition for a differential form as

$$\alpha = d\delta G\alpha + \delta dG\alpha + \alpha_{\mathrm{H}}, \quad (5.1)$$

where G is the Green's operator for the manifold. As above, we shall work on S^3 or R^3. In the case of R^3, we shall restrict attention to differential forms with compact support. Again, since the manifolds we are working on have metrics, we shall interchange freely between vector fields and one-forms and denote them by the same letter. We wish to consider a possible decomposition for the current operator J analogous to the Hodge decomposition. This may not be possible for a generalized one-form because of a lack of a suitable Green's operator G on generalized one-forms. Indeed, the space of generalized forms is not a Hilbert space and thus the usual treatments do not apply. Nevertheless, we may still obtain a decomposition of J from the decomposition of g as follows. We define differentiation of generalized one-forms in the distribution sense, so that

$$\delta J(g) = J(dg), \quad dJ(g) = J(\delta g) , \tag{5.2}$$

where δ is the coderivative. Suppose there now exists a generalized one-form K such that

$$K(g) = J(Gg) . \tag{5.3}$$

Then a decomposition for J follows directly. Define $J_{\rm H}(g) = J(Hg)$, where H is the projection onto the harmonic part for a one-form g. The generalized form $J_{\rm H}$ exists since we may take it to be $(1-\Delta)J$, where Δ is the Laplacian. The Laplacian acts on J by virtue of the definitions of differentiation in Eq. (5.2). Then we have

$$\begin{aligned} J(g) &= J(d\delta Gg) + J(\delta dGg) + J(Hg) \\ &= d\delta K(g) + \delta dK(g) + J_{\rm H}(g) , \end{aligned} \tag{5.4}$$

since G commutes with both d and δ. Define $J_d = d\delta K$ and $J_\delta = \delta dK$. Then we have immediately $dJ_d = 0$, $\delta J_\delta = 0$, and $\Delta J_{\rm H} = 0$ in the sense of distributions, so that we have indeed achieved a decomposition of a generalized one-form. The construction depends entirely on the existence of K, which we shall assume for the particular operators of interest.

Then we have the following proposition due to Kirillov.[12] Let $\Xi_{\delta 0}$ denote the subalgebra of volume preserving vector fields g such that $i_g d\mu$ is exact. This subalgebra is an ideal in Ξ_δ and contains the derived subalgebra $[\Xi_\delta,\Xi_\delta]$.

The proof of the above follows by noting that

$$i_{[g,g']}\, d\mu = L_g i_{g'}\, d\mu = d(i_g i_{g'}\, d\mu) , \tag{5.5}$$

which yields the result.

We may now construct the operators of Rasetti and Regge[8] in the following way. If the vortex were classical and ideal and its flow were given by $J(x)$, then we could write, for a vortex with unit circulation,

$$*dJ = \omega,$$

where

$$\omega = \int_\gamma ds \frac{dx_i(s)}{ds} \delta^3(x - x(s)) dx^i , \tag{5.6}$$

i.e., ω is a De Rham current describing the vorticity of the flow J. We shall now restore the operator nature of J using expression (5.6) and restrict ourselves to smearing vector fields in the ideal $\Xi_{\delta 0}$. Then we have

$$\begin{aligned} J(g) &= \int J \wedge i_g\, d\mu = \int J \wedge *(*d\alpha) \\ &= \int \delta * J \wedge * \alpha = \int *dJ \wedge * \alpha , \end{aligned}$$

where $i_g d\mu = d\alpha$. Using the definition for ω, this becomes

$$J(f) = \omega(\alpha) = \int_\gamma \alpha , \tag{5.7}$$

so that these operators are expressible as line integrals and correspond to the operators given by Rasetti and Regge,[8] but put in a slightly different formalism. We require that the operators $J(f)$ still satisfy the same current algebra: Note that these operators depend only upon a line integral over the vortex core and hence, as desired, will not describe fluctuations that may be occurring outside the core. Further, these operators are independent of the parametrization of the curve γ. From our discussion above, we would expect that these operators should form a set of observables that shift the position of the core of the vortex, whereas the $J(g)$ in the representations of Dif $f^\gamma_\mu(M)$ described above produce changes in the flow surrounding the vortex. It is interesting that the appropriate smearing vector fields come from the ideal $\Xi_{\delta 0}$.

We may write an expression for the linking invariant as an operator of the above type. We let γ' be a curve that links γ. Define β as

$$\beta = \int_{\gamma'} ds \frac{dx_i(s)}{ds} \delta^3(x - x(s)) dx^i . \tag{5.8}$$

Then the linking invariant may be expressed as

$$c(\gamma,\gamma') = 4\pi\omega(*dG\beta) , \tag{5.9}$$

where G is the Green's operator on M. Of course, we have neglected the exact operator nature of $c(\gamma,\gamma')$. As pointed out in Ref. 8, this operator should play the role of a Casimir operator for particular representations of the diffeomorphism group. We shall attempt to write some possible representations with this property in Sec. VI.

VI. INDUCED REPRESENTATIONS OF Dif $f_\mu(M)$ FOR THE VORTEX

Recall from the discussion in Sec. V that we needed to look for representations of the full group Dif $f_\mu(M)$ that isolated the action of the group Dif $f^\gamma_\mu(M)$. This is conveniently done using the theory of induced representations and the method of orbits, which we now briefly review.[11]

Given a Lie group G and a subgroup H of G, we may form the left coset space $X = G/H$. Let U be a representation of H in a Hilbert space V. If G may be given a measure that is invariant under the action of G, then we always have the left regular representation of G in $L^2(G,V)$ given by $[\tilde{U}(g)f](g_0) = f(g^{-1}g_0)$. Now consider G as a principal bundle over X and choose a global section s of G that is smooth up to a set of measure zero (recall that smooth global sections of a principal bundle do not exist unless the bundle is trivial). Restrict the left regular representation to the space $L^2(G,H,U)$ of functions in $L^2(G,V)$ that are equivalent with respect to the representation of U of H, i.e., that satisfy

$$f(gh) = U(h^{-1}) f(g) . \tag{6.1}$$

The induced representation of G induced from the subgroup U is then defined to be the left regular representation of G on $L^2(G,H,U)$: It may also be realized in the space of square integrable functions $L^2(X,V)$, with values in V as

$$[\tilde{U}(g)f](x) = A(x,g) f(g^{-1}x) , \tag{6.2}$$

where $A(x,g)$ is a cocycle taking values in the space of operators on V defined as follows. Define an element $h(x,g) \in H$ by $g^{-1}s(x) = s(g^{-1}x)h(x,g)$. Then $A(x,g)$ is taken as $U(h(x,g)^{-1})$. Here, $A(x,g)$ satisfies the cocycle condition

$$A(x,g)A(g^{-1}x,g') = A(x,gg') . \tag{6.3}$$

The measure used to define $L^2(X,V)$ may be obtained by projecting down the measure from G, which may be roughly thought of locally as a product measure over X, with X over the fiber that is isomorphic to H. We shall assume that such a measure exists and that it is invariant under the action of G. Alternatively, the representation may be thought of as acting in the space of sections of some appropriate vector bundle over X with fiber given by V, constructed as follows. Form the cross product $G \times V$ and then mod out by the equivalence relation $(g,v) \sim (gh, U(h^{-1})v)$. The resulting vector bundle is denoted $G \times_U V$. This bundle is a vector bundle over X with fiber V and it is known that the space of sections of $G \times_U V$ is isomorphic to the space of equivariant functions on G given in Eq. (6.1). The induced representation $\widetilde{U}$ then acts as a left regular representation on this space of sections.

The induced representation is often not irreducible. This is most easily seen from a brief discussion of the method of orbits. In this theory one chooses an element F in the dual to the Lie algebra and examines the orbit under the coadjoint action of the group that passes through this element. The orbit is isomorphic to $X = G/G_F$, where G_F is the stabilizer of F, and has a symplectic structure, so that it plays the role of a phase space. The symplectic two-form is then used to obtain a character representation of the stabilizer G_F; this is then used to induce a representation of G on the space of equivariant functions over the orbit, i.e., over phase space, as was described above. One then finds a situation similar to what happens when one looks at the regular representation of the Heisenberg–Weyl group on the space of square integrable functions over phase space. This representation is reducible, but becomes irreducible when restricted to the space of square integrable functions over the position coordinates alone. What is then required is to pick an appropriate polarization of X, i.e., the coordinates that will play the roles of momenta and position. Irreducible representations are then obtained by taking the representation to act on equivariant functions over just the position coordinates. A case of particular interest is when X has a complex structure, in which case it is Kähler. Then the irreducible representations are obtained by taking the representation space as the space of holomorphic sections of a holomorphic vector bundle over X. When G is compact, for example, and H is the maximal torus, then the irreducible representations occur when U is a highest weight character representation of H. Then X is Kähler and the bundle $G \times_U V$ is holomorphic. The representation acts irreducibly in the space of holomorphic sections of $G \times_U V$, which is a statement of the Borel–Weil theorem. More details of this may be found in Ref. 13.

We now wish to see how far we can apply the above theory to the representations we have been discussing. Let us work in S^3 and choose $G = \text{Dif}\, f_\mu(M)$ and $H = \text{Dif}\, f_\mu^\gamma(M)$. Choose as the representation U the regular representation of $\text{Dif}\, f_\mu^\gamma(M)$ on $L^2(M)$ discussed above, so that $V = L^2(M)$. Then we have that $X = \text{Dif}\, f_\mu(M)/\text{Dif}\, f_\mu^\gamma(M)$. Let $\text{Map}(S^1,M)$ be the loop space on M. Here, $\text{Dif}\, f_\mu(M)$ has a natural action on this space by a push forward. Consider two loops in M that lie in different knot classes. In order to get from one knot to another, it is necessary to pass two strands of the knot through one another some number of times. Now imagine that we try to accomplish this using a diffeomorphism of M. At the point where the two strands are crossing, however, the diffeomorprhism would be mapping two separate points in M to the crossing point and, hence, would no longer be invertible. Hence, we conclude that the action of the diffeomorphism group cannot change the knot class of the loop. Using this we see that the action of $\text{Dif}\, f_\mu(M)$ on $\text{Map}(S^1,M)$ yields a foliation of $\text{Map}(S^1,M)$, with each orbit being labeled by the knot class. Let O_γ be one of these orbits based on the loop γ. The action of the subgroup $\text{Dif}\, f_\mu^\gamma(M)$ on γ yields an arbitrary mapping of γ onto itself. This may also be accomplished by a reparametrization of γ. Then we see that the quotient $\text{Dif}\, f_\mu(M)/\text{Dif}\, f_\mu^\gamma(M)$ is actually isomorphic to the orbit O modulo reparametrizations and thus may be thought of as the space of unparametrized loop configurations of a particular knot class or as $\widetilde{O}_\gamma = O_\gamma/\text{Dif}\, f(S^1)$, where $\text{Dif}\, f(S^1)$ acts by reparametrization. The induced representation may then be realized in the space of square integrable functions over the space of unparametrized loops on M that take values in $L^2(M)$.

To construct the representation, all that remains is a choice of cocycle $A(\text{x},\psi)$, which takes values in the operators in $L^2(M)$. This follows from a choice of section in $\text{Dif}\, f_\mu(M)$. A possible choice of section may be as follows. Choose a reference loop γ_0 in $\widetilde{O}_\gamma$ and assign to it the identity in $\text{Dif}\, f_\mu(M)$. Let $\psi \in \text{Dif}\, f_\mu(M)$. Then we obtain a new element $\psi^{-1}\gamma_0$ in $\widetilde{O}_\gamma$ by a push forward. We also have the unique harmonic forms $\alpha_{\psi^{-1}\gamma_0}$ and α_{γ_0} on $M_{\psi^{-1}\gamma_0}$ and M_{γ_0}, respectively, for a fixed De Rham cohomology class. Then, under pullback, $\tilde{\alpha} = \psi^* \alpha_{\gamma_0}$ is a form on $M_{\psi^{-1}\gamma_0}$, which is no longer harmonic. We remark that we should be somewhat reserved in these manipulations because the diffeomorphism ψ changes the mainfold on which the one-forms are defined. Then we have the Hodge decomposition $\tilde{\alpha} = \alpha_{\psi^{-1}}\gamma_0 + df$ for some function f. We conjecture that there exists a diffeomorphism ψ_0 such that $\psi_0^* \tilde{\alpha} = \alpha_{\psi^{-1}}\gamma_0$. Then the resulting diffeomorphism $\tilde{\psi} = \psi_0 \psi$ takes α_{γ_0} to $\alpha_{\psi^{-1}\gamma_0}$ under pullback and thus preserves the harmonic nature of the flow. Hence, it is a harmonic map in the sense of Ref. 14. We conjecture that this diffeomorphism as well as ψ_0 exist and are unique. The existence of $\tilde{\psi}$ is not obvious and gives rise to the following problem: For two arbitrary loop configurations γ and γ' in M, there exists a diffeomorphism $\tilde{\psi}$ of M such that $\tilde{\psi}$ is a harmonic map from M_γ to $M_{\gamma'}$. We remark in this context that in the case of two-dimensional Riemannian manifolds, harmonic maps are known to correspond to conformal transformations,[14] but we do not know the corresponding results for arbitrary manifolds. Once the existence of $\tilde{\psi}$ is established, we may assign to $\psi^{-1}\gamma_0$ the diffeomorphism $(\psi_0\psi)^{-1}$ to give a section of $\text{Dif}\, f_\mu(M)$. The cocycle $A(x,\psi)$ then corresponds essentially to the action of $U(\psi_0)$ in $L^2(M)$. Note that we may also think of the induced representation as acting in the space of sections of an $L^2(M)$ bundle over $X \sim \widetilde{O}_\gamma$. It is interesting to note that if we restrict ourselves to one state in the fiber $V = L^2(M)$, in particular to the state that corresponds to the canonical element in

$\mathrm{Map}_n(M_\gamma, T)$, the bundle over $\widetilde{O}_\gamma$ becomes a line bundle and the resulting representations look very similar to those studied recently in the context of quantum hydrodynamics.[4]

We remark that the above procedure may be generalized to the R^3 case by using the exponential representations of Dif $f_\mu^\gamma(M)$ to induce representations of the full group Dif $f_\mu(M)$.

We further remark that the above orbits are also the coadjoint orbits. This may be seen as follows. Choose $F_\gamma = \omega$ in the dual to the Lie algebra given by the De Rham current in Eq. (5.6). It is clear that the support of the De Rham current ω is just γ itself. Further, the coadjoint action is such that it leaves the form of the functional ω the same, but shifts its support. Hence, the mapping from these currents to their support (modulo reparametrizations) is always one-to-one under the coadjoint action; hence, the coadjoint orbits correspond to the orbits in the space of unparametrized loops. If we were to follow the method of orbits, we would choose a character representation of Dif $f_\mu^\gamma(M)$ given by $e^{i\langle F, Y\rangle}$, where Y generates a one-parameter group in Dif $f_\mu^\gamma(M)$, and use this to induce a representation of Dif $f_\mu(M)$ instead of the representation U used above. Note that since different knot classes correspond to different orbits, we expect to obtain inequivalent representations for different knot classes, which would be labeled by operators such as $c(\gamma,\gamma')$. Different knot classes will lead to different choices of the functional F_γ and, hence, inequivalent character representations of Dif $f_\mu^\gamma(M)$. The process of induction should then yield inequivalent representations of Dif $f_\mu(M)$.

We make one final comment concerning the irreducibility of the representations introduced above. Since the orbit is symplectic, the representations above are representations over the phase space and, hence, we anticipate that they will be reducible. For irreducibility, we need a choice of polarization. This would be greatly facilitated if the orbit $\widetilde{O}_\gamma$ were known to possess a complex structure, although we do not know of any structure. The problem of choosing a polarization is under current study.

ACKNOWLEDGMENTS

The authors would like to thank A. Speliotopuolus, I. Juricic, M. Cleveland, T. Pritchett, and J. Robbins for useful discussions. Part of the material presented here was included as a thesis that U.K.A. submitted to the University of California in partial fulfillment of the Doctor of Philosophy degree.

[1] G. A. Goldin, J. Grodnick, R. T Powers, and D. H. Sharp, J. Math. Phys. **15**, 88 (1974).
[2] G. A. Goldin, R. Menikoff, and D. H. Sharp, J. Math. Phys. **21**, 650 (1980).
[3] G. A. Goldin, R. Menikoff, and D. H. Sharp, J. Math. Phys. **22**, 1664 (1981).
[4] G A. Goldin, R. Menikoff, and D. H. Sharp, Phys. Rev. Lett. **58**, 2162 (1987).
[5] R. Menikoff, J. Math. Phys. **15**, 1138 (1974).
[6] R. Menikoff, J. Math. Phys. **15**, 1394 (1974).
[7] J. R. Klauder, J. Math. Phys. **11**, 233 (1970).
[8] M. Rasetti and T. Regge, Physica A **80**, 217 (1975).
[9] A. M. Vershik, I. M. Gelfand, and M. I. Graev, Russ. Math. Surveys **30**, 1 (1975).
[10] A. M. Vershik, I. M. Gelfand, and M. I. Graev, Funct. Anal. Appl. **8**, 151 (1974).
[11] A. A. Kirillov, *Elements of the Theory of Group Representations* (Springer, New York, 1976).
[12] A. A. Kirillov, Sel. Math. Sov. **1**, 351 (1981).
[13] A. Pressley and G. Segal, *Loop Groups* (Clarendon, Oxford, 1986).
[14] J. Eells and J. H. Sampson. Am. J. Math. **86**, 109 (1964).

Hidden local gauge invariance in the one-dimensional Hubbard model and its equivalent coupled spin model

Huan-Qiang Zhou and Lin-Jie Jiang
Center of Theoretical Physics, CCAST (World Laboratory), Beijing 100080, People's Republic of China and Institute of Solid State Physics, Sichuan Normal University, Chengdu 610066, People's Republic of China
Ping-Feng Wu
Department of Physics, Sichuan Normal University, Chengdu 610066, People's Republic of China and Department of Physics, Jishou University, Jishou, Hunan 416000, People's Republic of China

(Received 15 September 1989; accepted for publication 20 December 1989)

Hidden local gauge invariance in the one-dimensional (1-D) Hubbard model and its equivalent coupled spin model is studied. It is found that Abelian $U(1) \otimes U(1)$ gauge transformations appear in both cases. Furthermore, it is shown that the energy spectrum is gauge invariant whereas the eigenvectors are explicitly gauge dependent. However, this result relies heavily on Shastry's conjecture about the eigenvalue of the transfer matrix for the 1-D Hubbard model. Lastly, there is also a discrete symmetry associated to $Z_2 \otimes Z_2$. Once this symmetry is broken, one immediately obtains another nontrivial solution to the Yang–Baxter relations.

I. INTRODUCTION

Recently much attention has been paid to the study of completely integrable quantum systems. At present, quite a number of quantum systems in $(1+1)$-dimensional field theories and in two-dimensional (2-D) statistical mechanics are shown to be integrable by various techniques developed in different branches of one-dimensional mathematical physics, such as the coordinate Bethe ansatz method,[1–3] Baxter's commuting transfer matrix technique,[4] the construction of an infinite number of conserved currents,[5] and the quantum inverse scattering method (QISM).[6–8] Nevertheless, QISM appears to offer a framework for a unified setting of many of these different techniques. Indeed, the development of this method has led to the important notion of the R matrix and established its crucial role in the theory of quantum completely integrable systems. From these matrices, one may extract the commutation relations between the elements of the quantum global monodromy matrix. In some sense, QISM opens the way for a systematic construction of the families of integrable systems connected with given R matrices.

Along this line, de Vega and Lopes[9] studied an interesting feature of completely integrable quantum systems, i.e., what they referred to as hidden local gauge invariance. They showed that the combination of the quantum integrability, i.e., the existence of the R matrix, with a global gauge transformation group leads to an Abelian local gauge invariance in the Heisenberg XXZ magnetic chain. As a consequence, one may construct a more general family of completely integrable quantum systems. Further, they also showed that in this model the exact energy spectrum turns out to be gauge invariant whereas the eigenvectors are explicitly gauge dependent. In a recent work,[10] we have presented a general formalism for hidden local gauge invariance in completely integrable lattice models of fermions. As an example, we studied a 1-D small polaron model, which can be mapped onto the 1-D Heisenberg XXZ model via the Jordan–Wigner transformation, and reexamined de Vega and Lopes' conclusion. There we also pointed out that Abelian $U(1) \otimes U(1)$ gauge transformations appear in the 1-D Hubbard model.

The purpose of this paper is to give a detailed study of hidden local gauge invariance in the 1-D Hubbard model and its equivalent coupled spin model. We show that in both cases the exact energy spectrum is also gauge invariant whereas the eigenvectors are explicitly gauge dependent as in the cases of the 1-D Heisenberg XXZ model and its equivalent fermion model. However, our result relies heavily on Shastry's conjecture about the eigenvalue of the transfer matrix for the 1-D Hubbard model, and thus is not complete.

The outline of this paper is in the following. In Sec. II, we give a brief review about some basic results for the 1-D Hubbard model and its equivalent coupled spin model. In Sec. III, we construct several lower-order conserved currents from the gauge-transformed transfer matrix for both cases. In Sec. IV, we present an incomplete derivation of the Bethe ansatz equations for the gauge-transformed systems, based on Shastry's conjecture about the eigenvalue of the transfer matrix for the 1-D Hubbard model. Finally, Sec. V is devoted to the conclusions.

II. BASIC NOTATIONS

Let us start from the 1-D Hubbard model described by the Hamiltonian

$$\mathscr{H} = -\sum_{j,s}(a_{js}^+ a_{j-1s} + a_{j-1s}^+ a_{js}) + U\sum_{j=1}^{N}\left(n_{j\uparrow} - \frac{1}{2}\right)\left(n_{j\downarrow} - \frac{1}{2}\right). \quad (1)$$

Here U is the coupling constant describing the Coulomb interaction, and s represents the two components of the fermions ($s = \uparrow$ or $\downarrow$). As usual, a_{js} and a_{js}^+ satisfy

$$\{a_{ir}, a_{js}\} = \{a_{ir}^+, a_{js}^+\} = 0, \quad \{a_{ir}, a_{js}^+\} = \delta_{ij}\delta_{rs}. \quad (2)$$

As was first noted by Shastry,[11] this Hamiltonian may be brought into the form

$$H=\sum_{j=1}^{N}(\sigma_j^+\sigma_{j-1}^- + \sigma_j^-\sigma_{j-1}^+ + \tau_j^+\tau_{j-1}^- + \tau_j^-\tau_{j-1}^+) + \frac{U}{4}\sum_{j=1}^{N}\sigma_j^z\tau_j^z, \tag{3}$$

using the Jordan–Wigner transformation

$$\begin{aligned}
a_{j\uparrow} &= \exp\left(i\pi\sum_{l=1}^{j-1}\sigma_l^-\sigma_l^+\right)\sigma_j^-,\\
a_{j\uparrow}^+ &= \exp\left(i\pi\sum_{l=1}^{j-1}\sigma_l^-\sigma_l^+\right)\sigma_j^+,\\
n_{j\uparrow} &= \tfrac{1}{2}(1+\sigma_j^z),\\
a_{j\downarrow} &= \exp\left(i\pi\sum_{l=1}^{N}\sigma_l^-\sigma_l^+\right)\exp\left(i\pi\sum_{l=1}^{j-1}\tau_l^-\tau_l^+\right)\tau_j^-,\\
a_{j\downarrow}^+ &= \exp\left(i\pi\sum_{l=1}^{N}\sigma_l^-\sigma_l^+\right)\exp\left(i\pi\sum_{l=1}^{j-1}\tau_l^-\tau_l^+\right)\tau_j^+,\\
n_{j\downarrow} &= \tfrac{1}{2}(1+\tau_j^z).
\end{aligned} \tag{4}$$

Here the sum over j is from 1 to N, and the periodic boundary condition is imposed. In Ref. 11, Shastry showed that the Hamiltonian (3), which we shall refer to as a coupled spin model, generates a solution to the Yang–Baxter relations

$$R(\lambda,\mu)L_j(\lambda)\otimes L_j(\mu) = L_j(\mu)\otimes L_j(\lambda)R(\lambda,\mu), \tag{5}$$

where

$$L_j = IL_j^{(\sigma)}\otimes L_j^{(\tau)}I, \tag{6}$$

with

$$I=\begin{pmatrix} e^{h/2} & 0 & 0 & 0\\ 0 & e^{-h/2} & 0 & 0\\ 0 & 0 & e^{-h/2} & 0\\ 0 & 0 & 0 & e^{h/2}\end{pmatrix},$$

$$\sinh 2h = \frac{U}{4}\sin 2\lambda, \tag{7}$$

and

$$L_j^{(\sigma)}(\lambda) = \begin{pmatrix} \cos\lambda\,\sigma_j^+\sigma_j^- + \sin\lambda\,\sigma_j^-\sigma_j^+ & \sigma_j^- \\ \sigma_j^+ & \sin\lambda\,\sigma_j^+\sigma_j^- + \cos\lambda\,\sigma_j^-\sigma_j^+\end{pmatrix}. \tag{8}$$

Note that $L_j^{(\tau)}(\lambda)$ has the same form as $L_j^{(\sigma)}(\lambda)$ with τ's replacing σ's. Here, λ is the spectral parameter. In this case, the corresponding R matrix is

$$R(\lambda,\mu)=\begin{pmatrix}
a_+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & c_+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & c_+ & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & b_+ & 0 & 0 & d & 0 & 0 & d & 0 & 0 & b_+-a_+ & 0 & 0 & 0\\
0 & c_- & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & a_- & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & d & 0 & 0 & b_- & 0 & 0 & b_--a_- & 0 & 0 & d & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & c_- & 0 & 0\\
0 & 0 & c_- & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & d & 0 & 0 & b_--a_- & 0 & 0 & b_- & 0 & 0 & d & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_- & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & c_- & 0\\
0 & 0 & 0 & b_+-a_+ & 0 & 0 & d & 0 & 0 & d & 0 & 0 & b_+ & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & c_+ & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_+ & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_+
\end{pmatrix}, \tag{9}$$

with

$$\begin{aligned}
a_\pm(\lambda,\mu) &= \cosh[h(\lambda)-h(\mu)]\cos(\lambda-\mu) \pm \sinh[h(\lambda)-h(\mu)]\cos(\lambda+\mu),\\
b_\pm(\lambda,\mu) &= \frac{\cosh[h(\lambda)-h(\mu)]\cos(\lambda+\mu) \pm \sinh[h(\lambda)-h(\mu)]\cos(\lambda-\mu)}{\cos^2\lambda-\sin^2\mu},\\
c_\pm(\lambda,\mu) &= -\cosh[h(\lambda)-h(\mu)]\sin(\lambda-\mu) \pm \sinh[h(\lambda)-h(\mu)]\sin(\lambda+\mu),\\
d(\lambda,\mu) &= \frac{e^{-2h(\mu)}\sin 2\lambda - e^{-2h(\lambda)}\sin 2\mu}{2(\cos^2\lambda-\sin^2\mu)}.
\end{aligned} \tag{10}$$

Thus the system under study possesses an infinite number of conserved currents that are in involution to each other. As is well known, the generating functional for those conserved currents is the so-called transfer matrix $\tau(\lambda)$, which is the trace of a global monodromy matrix $T(\lambda)$,

$$\tau(\lambda) = \mathrm{tr}\, T(\lambda), \quad T(\lambda) = L_N(\lambda) L_{N-1}(\lambda) \cdots L_1(\lambda). \tag{11}$$

Indeed, an expansion of the transfer matrix $\tau(\lambda)$ in powers of λ leads to the Hamiltonian (3) as well as higher conserved currents.[11–13]

As for the Hamiltonian (1), the corresponding results have been obtained by Wadati and his co-workers.[13] They showed that this system also generates a solution to the Yang–Baxter relations:

$$\mathscr{R}(\lambda,\mu)\mathscr{L}_j(\lambda) \underset{s}{\otimes} \mathscr{L}_j(\mu) = \mathscr{L}_j(\mu) \underset{s}{\otimes} \mathscr{L}_j(\lambda)\mathscr{R}(\lambda,\mu), \tag{12}$$

where

$$\mathscr{L}_j = I\mathscr{L}_{j\uparrow} \underset{s}{\otimes} \mathscr{L}_{j\downarrow} I, \tag{13}$$

with

$$\mathscr{L}_{j\uparrow}(\lambda) = \begin{pmatrix} i\cos\lambda n_{j\uparrow} - \sin\lambda a_{j\uparrow} a_{j\uparrow}^{+} & i a_{j\uparrow} \\ a_{j\uparrow}^{+} & -i\sin\lambda n_{j\uparrow} - \cos\lambda a_{j\uparrow} a_{j\uparrow}^{+} \end{pmatrix},$$

$$\mathscr{L}_{j\downarrow}(\lambda) = \begin{pmatrix} -i\cos\lambda n_{j\downarrow} + \sin\lambda a_{j\downarrow} a_{j\downarrow}^{+} & a_{j\downarrow} \\ i a_{j\downarrow}^{+} & i\sin\lambda n_{j\downarrow} + \cos\lambda a_{j\downarrow} a_{j\downarrow}^{+} \end{pmatrix}. \tag{14}$$

In this case, the $\mathscr{R}$ matrix has the form

$$\mathscr{R}(\lambda,\mu) = \begin{pmatrix}
a_+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & ic_+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & ic_+ & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & b_+ & 0 & 0 & -id & 0 & 0 & id & 0 & 0 & a_+ - b_+ & 0 & 0 & 0 \\
0 & -ic_- & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & a_- & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & id & 0 & 0 & b_- & 0 & 0 & a_- - b_- & 0 & 0 & -id & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -ic_- & 0 & 0 \\
0 & 0 & -ic_- & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -id & 0 & 0 & a_- - b_- & 0 & 0 & b_- & 0 & 0 & id & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_- & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -ic_- & 0 \\
0 & 0 & 0 & a_+ - b_+ & 0 & 0 & id & 0 & 0 & -id & 0 & 0 & b_+ & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & ic_+ & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ic_+ & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_+
\end{pmatrix}. \tag{15}$$

Here we have used the Grassmann tensor product

$$(\mathscr{A} \underset{s}{\otimes} \mathscr{B})_{ik,jl} = (-1)^{[P(i)+p(j)]p(k)} \mathscr{A}_{ij} \mathscr{B}_{kl}, \tag{16}$$

with $P(1) = P(4) = 0$, and $P(2) = P(3) = 1$.

In exactly the same way as before, one may construct an infinite number of conserved currents for the system (1). The generating functional for those conserved currents is the transfer matrix $\tau(\lambda)$, which is the supertrace of a monodromy matrix $\mathscr{T}(\lambda)$,

$$\tau(\lambda) = \mathrm{str}\, \mathscr{T}(\lambda), \tag{17}$$

with

$$\mathscr{T}(\lambda) = \mathscr{L}_N(\lambda) \mathscr{L}_{N-1}(\lambda) \cdots \mathscr{L}_1(\lambda). \tag{18}$$

Several lower-order conserved currents may be found in Ref. 12.

III. HIDDEN LOCAL GAUGE INVARIANCE

Let us first study hidden local gauge invariance in a coupled spin model (3). Now we notice that the Yang–Baxter relations (5) are invariant under the transformation

$$L_j \to L_j^{(g)} = g(k_j, k_j') L_j g^{-1}(l_j, l_j'), \tag{19}$$

provided the following condition holds:

$$[g \otimes g, R] = 0. \tag{20}$$

In this case, there are three gauge symmetries: A continuous one

$$g(k,k') = \begin{pmatrix} e^{i(k+k')} & 0 & 0 & 0 \\ 0 & e^{i(k-k')} & 0 & 0 \\ 0 & 0 & e^{-i(k-k')} & 0 \\ 0 & 0 & 0 & e^{-i(k+k')} \end{pmatrix}, \quad k,k' \in \mathbf{R}, \tag{21}$$

and two discrete ones

$$g=\begin{pmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{pmatrix}\quad\text{and}\quad g=\begin{pmatrix}0&0&0&-1\\0&0&1&0\\0&1&0&0\\-1&0&0&0\end{pmatrix}. \tag{22}$$

As emphasized in our previous work,[10] these two discrete symmetries impose additional constraints on the R matrix as well as the L_j matrix. Once these symmetries are broken, one may obtain another solution to the Yang–Baxter relations.[12] So we shall analyze here the continuous symmetry (21). It can be verified that an expansion of the logarithm of the gauge-transformed transfer matrix $\tau_{cs}^{(g)}(\lambda)$ through third order in λ is given by

$$\ln \tau_{cs}^{(g)}(\lambda)=\ln \tau_{cs}^{(g)}(0)+H^{(g)}\lambda-\frac{1}{2!}(J^{(g)}+2N)\lambda^2+\frac{1}{3!}\left[K^{(g)}-\left(2+\frac{5U^2}{8}\right)H^{(g)}\right]\lambda^3+\cdots, \tag{23}$$

where

$$H^{(g)}=\sum_{j=1}^{N}\{[\sigma_j^+\sigma_{j-1}^-\exp(2i(k_{j-1}-l_j))+\text{h.c.}]+(\sigma\to\tau,k\to k',l\to l')\}+\frac{U}{4}\sum_{j=1}^{N}\sigma_j^z\tau_j^z, \tag{24}$$

$$\begin{aligned}J^{(g)}=&\sum_{j=1}^{N}\{[\sigma_{j+1}^+\sigma_j^z\sigma_{j-1}^-\exp(2i(k_j-l_{j+1}+k_{j-1}-l_j))\\&-\text{h.c.}]+(\sigma\to\tau,k\to k',l\to l')\}\\&-\frac{U}{2}\sum_{j=1}^{N}\{[\sigma_j^+\sigma_{j-1}^-\exp(2i(k_{j-1}-l_j))-\text{h.c.}]\\&\times(\tau_j^z+\tau_{j-1}^z)+(\sigma\leftrightarrow\tau,k\to k',l\to l')\},\end{aligned} \tag{25}$$

and

$$\begin{aligned}K^{(g)}=&2\sum_{j=1}^{N}\{[\sigma_{j+1}^+\sigma_j^z\sigma_{j-1}^z\sigma_{j-2}^-\exp(2i(k_j-l_{j+1}+k_{j-1}-l_j+k_{j-2}-l_{j-1}))+\text{h.c.}]+(\sigma\to\tau,k\to k',l\to l')\}\\&+U\sum_{j=1}^{N}\{[2\sigma_{j+1}^+\sigma_j^-\exp(2i(k_j-l_{j+1}))+\sigma_j^+\sigma_{j-1}^-\exp(2i(k_{j-1}-l_j))-\text{h.c.}]\\&\times[\tau_j^+\tau_{j-1}^-\exp(2i(k'_{j-1}-l'_j))-\text{h.c.}]\\&-[\sigma_{j+1}^+\sigma_j^z\sigma_{j-1}^-\exp(2i(k_j-l_{j+1}+k_{j-1}-l_j))+\text{h.c.}](\tau_{j+1}^z+\tau_j^z+\tau_{j-1}^z)\\&-\tfrac{1}{4}(2\sigma_{j+1}^z+\sigma_j^z)\tau_j^z+(\sigma\leftrightarrow\tau,k\leftrightarrow k',l\leftrightarrow l')\}\\&+\frac{U^2}{2}\sum_{j=1}^{N}\{[\sigma_j^+\sigma_{j-1}^-\exp(2i(k_{j-1}-l_j))+\text{h.c.}]\tau_j^z\tau_{j-1}^z+(\sigma\leftrightarrow\tau,k\to k',l\to l')\}-\frac{U^3}{8}\sum_{j=1}^{N}\sigma_j^z\tau_j^z.\end{aligned} \tag{26}$$

Here and hereafter, the subscript "cs" denotes that the corresponding quantity is assigned to a coupled spin model (3), while the superscript "g" denotes that the corresponding quantity has been gauge-transformed. From the above results one can see how the Hamiltonian as well as higher conserved currents transform under the gauge transformations given by Eqs. (19) and (21).

A similar analysis can also be carried out for the 1-D Hubbard model (1). In fact, the Yang–Baxter relations (12) are invariant under the gauge transformation

$$\mathscr{L}_j\to\mathscr{L}_j^{(g)}=g(k_j,k'_j)\mathscr{L}_j g^{-1}(l_j,l'_j), \tag{27}$$

provided the following condition holds:

$$\left[g\underset{s}{\otimes}g,\mathscr{R}\right]=0. \tag{28}$$

In this case, there are also three gauge symmetries: A continuous one

$$g(k,k')=\begin{pmatrix}e^{i(k+k')}&0&0&0\\0&e^{i(k-k')}&0&0\\0&0&e^{-i(k-k')}&0\\0&0&0&e^{-i(k+k')}\end{pmatrix},\quad k,k'\in\mathbf{R}, \tag{29}$$

and two discrete ones

$$g=\begin{pmatrix}0&0&0&1\\0&0&-1&0\\0&1&0&0\\-1&0&0&0\end{pmatrix}\quad\text{and}\quad g=\begin{pmatrix}0&0&0&-1\\0&0&-1&0\\0&1&0&0\\1&0&0&0\end{pmatrix}. \tag{30}$$

As in the case of a coupled spin model, these discrete symmetries impose additional constraints on the $\mathscr{R}$ matrix as well as $\mathscr{L}_j$ matrix. When these constraints are removed, one may get another solution to the Yang–Baxter relations (12). Here we restrict ourselves to the continuous symmetry (29). Similarly, one may construct the first few conserved currents from an expansion of the logarithm of the transformed transfer matrix $\tau_H^{(g)}(\lambda)$ through third order in λ,

$$\ln \tau_H^{(g)}(\lambda) = \ln \tau_H^{(g)}(0) + \mathscr{H}^{(g)}\lambda - \frac{1}{2!}(\mathscr{J}^{(g)} + 2N)\lambda^2 + \frac{1}{3!}\left[\mathscr{K}^{(g)} - \left(2 + \frac{9U^2}{8}\right)\mathscr{H}^{(g)} + \frac{NU}{4}(6 + U^2)\right]\lambda^3 + \cdots, \tag{31}$$

where

$$\mathscr{H}^{(g)} = -\sum_{j=1}^{N}\{[a_{j\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_{j-1} - l_j)) + \text{h.c.}] + (\uparrow\to\downarrow, k\to k', l\to l')\} + U\sum_{j=1}^{N}\left(n_{j\uparrow} - \frac{1}{2}\right)\left(n_{j\downarrow} - \frac{1}{2}\right), \tag{32}$$

$$\begin{aligned}\mathscr{J}^{(g)} = &-\sum_{j=1}^{N}\{[a_{j+1\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_j - l_{j+1} + k_{j-1} - l_j)) - \text{h.c.}] + (\uparrow\to\downarrow, k\to k', l\to l')\}\\ &+ U\sum_{j=1}^{N}\{[a_{j\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_{j-1} - l_j)) - \text{h.c.}](n_{j\downarrow} + n_{j-1\downarrow}) - [a_{j\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_{j-1} - l_j)) - \text{h.c.}]\\ &+ (\uparrow\leftrightarrow\downarrow, k\to k', l\to l')\},\end{aligned} \tag{33}$$

and

$$\begin{aligned}\mathscr{K}^{(g)} = &-2\sum_{j=1}^{N}\{[a_{j+1\uparrow}^{+}a_{j-2\uparrow}\exp(2i(k_j - l_{j+1} + k_{j-1} - l_j + k_{j-2} - l_{j-1})) + \text{h.c.}] + (\uparrow\to\downarrow, k\to k', l\to l')\}\\ &+ U\sum_{j=1}^{N}\{[2a_{j+1\uparrow}^{+}a_{j\uparrow}\exp(2i(k_j - l_{j+1})) + a_{j\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_{j-1} - l_j)) - \text{h.c.}]\\ &\times[a_{j\downarrow}^{+}a_{j-1\downarrow}\exp(2i(k'_{j-1} - l'_j)) - \text{h.c.}]\\ &+ 2[a_{j+1\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_j - l_{j+1} + k_{j-1} - l_j)) + \text{h.c.}](n_{j+1\downarrow} + n_{j\downarrow} + n_{j-1\downarrow}) - (2n_{j+1\uparrow} + n_{j\uparrow})n_{j\downarrow} + 3n_{j\uparrow}\\ &- 3[a_{j+1\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_j - l_{j+1} + k_{j-1} - l_j)) + \text{h.c.}] + (\uparrow\leftrightarrow\downarrow, k\leftrightarrow k', l\leftrightarrow l')\}\\ &- U^2\sum_{j=1}^{N}\{2[a_{j\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_{j-1} - l_j)) + \text{h.c.}]n_{j\downarrow}n_{j-1\downarrow} - [a_{j\uparrow}^{+}a_{j-1\uparrow}\exp(2i(k_{j-1} - l_j)) + \text{h.c.}]\\ &\times(n_{j\downarrow} + n_{j-1\downarrow}) + (\uparrow\leftrightarrow\downarrow, k\to k', l\to l')\} - \frac{U^3}{2}\sum_{j=1}^{N}(2n_{j\uparrow}n_{j\downarrow} - n_{j\uparrow} - n_{j\downarrow}).\end{aligned} \tag{34}$$

Here the subscript "H" denotes that the corresponding quantity is assigned to the 1-D Hubbard model (1).

IV. BETHE ANSATZ EQUATIONS

In order to study the effect of the local gauge transformation on the eigenvalues and eigenvectors for the systems under study, we have to derive the so-called Bethe ansatz equations. This problem, however, is nontrivial from either the coordinate Bethe ansatz or the algebraic Bethe ansatz point of view. Here we only give an incomplete derivation of these equations based on Shastry's conjecture[14] about the eigenvalue of the transfer matrix for a coupled spin model (3).

Denoting the matrix elements of $T^{(g)}(\lambda)$ by $T_{ij}^{(g)}(\lambda)$ ($i,j = 1,2,3,4$) and defining the pseudovacuum $|0\rangle$ as the state with all spins up, we have

$$\begin{aligned}&T_{11}^{(g)}(\lambda)|0\rangle = e^{i(Q+Q')}e^{Nh}\cos^{2N}\lambda\,|0\rangle,\\ &T_{22}^{(g)}(\lambda)|0\rangle = e^{i(Q-Q')}e^{-Nh}\sin^{N}\lambda\cos^{N}\lambda\,|0\rangle,\\ &T_{33}^{(g)}(\lambda)|0\rangle = e^{-i(Q-Q')}e^{-Nh}\sin^{N}\lambda\cos^{N}\lambda\,|0\rangle,\\ &T_{44}^{(g)}(\lambda)|0\rangle = e^{-i(Q+Q')}e^{Nh}\sin^{2N}\lambda\,|0\rangle,\\ &Q = \sum_{j=1}^{N}(k_j - l_j),\quad Q' = \sum_{j=1}^{N}(k'_j - l'_j).\end{aligned} \tag{35}$$

Now let us consider a general state with M particles with M-K particles having spin up and K particles having spin down. Taking into account Shastry's conjecture on the eigenvalue of the transfer matrix for (3) and having in mind the fact that the Yang–Baxter relations (5) are invariant under the local gauge transformations, we may write out the eigenvalue of the transformed transfer matrix $\tau_{cs}^{(g)}(\lambda)$:

$$\begin{aligned}\Lambda_{M,K}^{(g)}(\lambda,\{\lambda_n\},\{\mu_m\}) = &\,e^{i(Q+Q')}e^{Nh}\cos^{2N}\lambda\prod_{n=1}^{M}\frac{\tan\lambda + e^{i\lambda_n - 2h}}{1 - e^{i\lambda_n - 2h}\tan\lambda} + (-1)^M e^{-i(Q+Q')}e^{Nh}\sin^{2N}\lambda\prod_{n=1}^{M}\frac{\tan\lambda + e^{i\lambda_n + 2h}}{1 - e^{i\lambda_n + 2h}\tan\lambda}\\ &+ (-1)^{M-K}e^{i(Q-Q')}e^{-Nh}\sin^{N}\lambda\cos^{N}\lambda\prod_{n=1}^{M}\frac{\tan\lambda + e^{i\lambda_n - 2h}}{1 - e^{i\lambda_n - 2h}\tan\lambda}\\ &\times\prod_{m=1}^{K}\frac{e^{2h}/\tan\lambda - e^{-2h}\tan\lambda - 2\mu_m + U/2}{e^{2h}/\tan\lambda - e^{-2h}\tan\lambda - 2\mu_m - U/2} + (-1)^K e^{-i(Q-Q')}e^{-Nh}\sin^{N}\lambda\cos^{N}\lambda\\ &\times\prod_{n=1}^{M}\frac{\tan\lambda + e^{i\lambda_n + 2h}}{1 - e^{i\lambda_n + 2h}\tan\lambda}\prod_{m=1}^{K}\frac{e^{-2h}/\tan\lambda - e^{2h}\tan\lambda - 2\mu_m - U/2}{e^{-2h}/\tan\lambda - e^{2h}\tan\lambda - 2\mu_m + U/2}.\end{aligned} \tag{36}$$

Repeating Shastry's argument,[14] we immediately obtain the Bethe ansatz equations

$$e^{2iQ'}e^{iN\lambda_n} = (-1)^{M-K-1} \prod_{m=1}^{K} \frac{i\sin\lambda_n - \mu_m + U/4}{i\sin\lambda_n - \mu_m - U/4}, \tag{37}$$

$$e^{2i(Q-Q')} \prod_{n\neq m} \frac{\mu_m - \mu_n + U/2}{\mu_m - \mu_n - U/2} = (-1)^M \prod_{n=1}^{M} \frac{i\sin\lambda_n - \mu_m - U/4}{i\sin\lambda_n - \mu_m + U/4}. \tag{38}$$

Since the transformed transfer matrices for different values of the spectral parameter λ commute, the eigenvectors are λ-independent and then the eigenvalues of the Hamiltonian as well as higher conserved currents can be determined from the logarithmic derivative of $\Lambda^{(g)}_{M,K}(\lambda)$. From Eqs. (23) and (36) we have

$$\begin{aligned} P^{(g)}_{cs} &= Q + Q' + \sum_{n=1}^{M} \lambda_n, \\ E^{(g)}_{cs} &= \left(\frac{N}{2} - \frac{M}{4}\right) U + 2 \sum_{n=1}^{M} \cos\lambda_n, \\ J^{(g)}_{cs} &= -2i \sum_{n=1}^{M} \sin 2\lambda_n, \\ K^{(g)}_{cs} &= \left(M - \frac{N}{2}\right) U \left(1 - \frac{3U^2}{16}\right) \\ &\quad + \sum_{n=1}^{M} \Big[4\cos 3\lambda_n - 6U\cos 2\lambda_n \\ &\quad + \left(8 + \frac{11U^2}{4}\right) \cos\lambda_n \Big]. \end{aligned} \tag{39}$$

Here we have introduced the momentum operator $P^{(g)}$ defined by

$$P^{(g)} = -i \ln \tau^{(g)}_{cs}(0), \tag{40}$$

while the values of λ_n and μ_m are obtained by solving Eqs. (37) and (38). From this we see that the eigenvalues of the gauge-transformed transfer matrix $\tau^{(g)}_{cs}(\lambda)$ and hence those of the Hamiltonian as well as higher conserved currents depend on k_j, l_j and k'_j, l'_j only through the sums

$$Q = \sum_{j=1}^{N} (k_j - l_j) \quad \text{and} \quad Q' = \sum_{j=1}^{N} (k'_j - l'_j). \tag{41}$$

In particular, for a one-particle excitation over the pseudovacuum $|0\rangle$, the exact dispersion relation turns out to be gauge-invariant,

$$E^{(g)}_{cs} = \left(\frac{N}{2} - \frac{M}{4}\right) U + 2\cos P^{(g)}_{cs}. \tag{42}$$

Here the momentum $P^{(g)}_{cs}$ has been shifted by $Q + Q'$;

$$P^{(g)}_{cs} \to P^{(g)}_{cs} - (Q + Q').$$

However, the eigenvectors of $\tau^{(g)}_{cs}(\lambda)$ will depend in a detailed form on k_j, l_j, and k'_j, l'_j as in the case of the Heisenberg *XXZ* model.[9]

Analogously, we may write out the eigenvalue of the transformed transfer matrix and the Bethe ansatz equations for the equivalent fermion model (1), from which the same conclusion is recovered.[10]

V. CONCLUSION

In this paper we have studied hidden local gauge invariance in the 1-D Hubbard model and its equivalent coupled spin model. We have shown that Abelian $U(1) \otimes U(1)$ gauge transformations appear in these two equivalent systems. This allows us to construct a new family of completely integrable quantum systems, each of which corresponds to a conserved current obtained from an expansion of the gauge-transformed transfer matrix $\tau^{(g)}(\lambda)$ with respect to the powers of the spectral parameter λ. Further, we have presented an incomplete discussion of the Bethe ansatz equations for a coupled spin model (3), based on Shastry's conjecture about the eigenvalue of the transfer matrix for this system. From this we have concluded that the exact energy spectrum is gauge-invariant as in the case of the Heisenberg *XXZ* model.

In conclusion let us add some comments on the discrete symmetry associated to $Z_2 \otimes Z_2$. As mentioned above, this symmetry imposes an additional constraint on the R matrix as well as L_j matrix. Once this constraint is removed, one may get another solution to the Yang–Baxter relations. For a coupled spin model (3), the corresponding L_j matrix have the form

$$L_j = I L^{(\sigma)}_j \otimes L^{(\tau)}_j I, \tag{43}$$

where

$$L^{(\sigma)}_j(\lambda) = \begin{pmatrix} \xi\cos\lambda\,\sigma^+_j\sigma^-_j + \xi^{-1}\sin\lambda\sigma^-_j\sigma^+_j & \sigma^-_j \\ \sigma^+_j & \xi\sin\lambda\,\sigma^+_j\sigma^-_j + \xi^{-1}\cos\lambda\,\sigma^-_j\sigma^+_j \end{pmatrix}, \tag{44}$$

where

$$\xi = \sec\alpha\cos(\lambda + \alpha)\sec\lambda,$$

and $L^{(\tau)}_j(\lambda)$ has the same form as $L^{(\sigma)}_j(\lambda)$ with τ's replacing σ's. We have shown[12] that this L_j matrix provides a natural description of the 1-D Hubbard model with the chemical potential term.

ACKNOWLEDGMENTS

We are grateful to Professor B. L. Zhang for his interest in this work. We also thank Dr. Jian-Gang Tang, Dr. Yu-Bin Liu, and Dr. Zhuang Xiong for very helpful discussions.

[1] H. Bethe, Z. Phys. **79**, 205 (1931).
[2] C. N. Yang, Phys. Rev. Lett. **19**, 312 (1967).
[3] A. A. Belavin, Phys. Lett. B **87**, 117 (1979); P. B. Wiegmann, Phys. Lett. A **80**, 163 (1980); N. Andrei, K. Furuya, and J. H. Lowenstein, Rev. Mod. Phys. **5**, 331 (1983).
[4] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).
[5] M. Lüscher, Nucl. Phys. B **117**, 475 (1976).
[6] L. D. Faddeev, Sov. Sci. Rev. Math. Phys. C **1**, 107 (1981).
[7] H. B. Thacker, Rev. Mod. Phys. **53**, 253 (1982).
[8] A. G. Izergin and V. E. Korepin, Sov. J. Part. Nucl. **13**, 207 (1982); P. P. Kulish and E. K. Sklyanin, in *Lecture Notes in Physics*, edited by J. Hietarinta and C. Montonen (Springer, Berlin, 1982), Vol. 151.
[9] H. J. de Vega and E. Lopes, Phys. Lett. B **186**, 180 (1987).

[10]H. Q. Zhou, L. J. Jiang, and P. F. Wu, Phys. Lett. A **137**, 244 (1989).
[11]B. S. Shastry, Phys. Rev. Lett. **56**, 1529 (1986); **56**, 2453 (1986).
[12]H. Q. Zhou, Z. Xiong, and L. J. Jiang, J. Phys. A **21**, 3385 (1988); H. Q. Zhou and J. G. Tang, Phys. Rev. B **38**, 11915 (1988); H. Q. Zhou and J. G. Tang, J. Phys. A (to be published).
[13]M. Wadati, E. Olmedilla, and Y. Akutsu, J. Phys. Soc. Jpn. **56**, 1340 (1987); E. Olmedilla, M. Wadati, and Y. Akutsu, *ibid.* **56**, 2298 (1987).
[14]B. S. Shastry, J. Stat. Phys. **50**, 57 (1988).