Unitary highest weight representations of noncompact supergroups

Murat Günaydin
Physics Department, Penn State University, University Park, Pennsylvania 16802

(Received 1 December 1987; accepted for publication 17 February 1988)

The oscillator method for the construction of unitary highest (or lowest) weight representations of noncompact groups and supergroups is generalized. Within this generalization, the method yields unitary highest weight representations of all simple supergroups whose even subgroups are in the form of a direct product of a compact group with a simple noncompact group. The method is illustrated by studying in detail the unitary highest weight representations of the supergroup OSp $(2n + 1/2m,R)$. The generalized supercoherent states associated with these unitary representations are also defined.

I. INTRODUCTION

The Lie superalgebras have come to play an increasingly important role in theoretical physics. They appear as symmetry algebras of supersymmetrical theories such as superstring[1] and supergravity theories.[2] They also appear as effective dynamical symmetry algebras in the study of nuclear states.[3] For physical theories in which the supersymmetry enters at a fundamental level the relevant supergroups are, in general, noncompact. (We define a noncompact supergroup as one whose even subgroup is noncompact.) For example, the space-time superalgebras are all noncompact. One can expect noncompact supergroups to be also relevant as dynamical symmetry groups in various branches of physics.

Motivated by the possible applications to supergravity theories, a general method for the construction of oscillatorlike unitary representations of noncompact Lie groups was developed in Refs. 4 and 5. This method was then extended to the construction of oscillatorlike unitary representations of noncompact supergroups.[6] Later, this method was further developed and applied to supergravity and superstring theories.[7-11] The oscillator method developed in the above cited references yields the (oscillatorlike) unitary irreducible representations (UIR's) of the lowest weight type of noncompact groups and supergroups that have a three-graded (Jordan) structure with respect to a maximal compact subgroup or subsupergroup. For noncompact Lie groups the oscillator method is of complete generality for the construction of unitary representations of the highest weight type (or equivalently lowest weight type). This is a consequence of the fact that the noncompact simple Lie groups that admit UIR's of the highest weight type all have a Jordan structure with respect to their maximal compact subgroups. However, this is not the case for noncompact supergroups. There is a large class of noncompact supergroups that admits highest weight UIR's but does not have a Jordan structure with respect to a maximal compact subsupergroup. Therefore for noncompact Lie supergroups the oscillator method needs to be generalized for it to be of complete generality. In this paper we give such a generalization. This generalization allows one to construct unitary highest weight representations of all simple noncompact supergroups whose even subgroups are direct products of a compact group with a simple noncompact group.

II. THE OSCILLATOR METHOD AND THE UNITARY HIGHEST WEIGHT REPRESENTATIONS OF NONCOMPACT LIE GROUPS

A simple connected noncompact Lie group G with a maximal compact subgroup K has unitary representations of the highest weight type (or, equivalently, of the lowest weight type) if the quotient space G/K is an Hermitian symmetric space.[12] (If G has a center Z then K/Z is assumed to be compact.) A unitary representation of the highest weight (or lowest weight) type is defined as a unitary representation for which some generator belonging to the Lie algebra L of G has a spectrum bounded from above (or from below). Thus from the list of Hermitian symmetric spaces[13] one can easily read off the simple noncompact groups that admit unitary representations of the highest weight (or lowest weight) type. In Table I we give a complete list of such groups and their maximal compact subgroups. We should note that unitary representations of the highest weight and the lowest weight type are related by a simple involution. In the mathematical literature the use of the term "highest weight representations" has been traditional. In most physical problems the term "lowest weight representation" is more appropriate since the operator whose spectrum is bounded is often the energy operator or the particle number operator.

The Lie algebra L of a noncompact group G that posseses unitary highest weight (UHW) representations has a Jordan decomposition (three-grading) (Ref. 5) with respect to the Lie algebra L_0 of the maximal compact subgroup K, i.e., it can be decomposed in a split basis as a vector space direct sum

TABLE I. The first column gives a complete list of simple noncompact Lie groups G that possess unitary representations of the highest weight type. The second column gives their maximal compact subgroups.

G	K
$SU(n,m)$	$SU(n)\times SU(m)\times U(1)$
$SO(n,2)$	$SO(n)\times SO(2)$
$Sp(2n,R)$	$U(n)$
$SO^*(2n)$	$U(n)$
$E_{6(-14)}$	$SO(10)\times SO(2)$
$E_{7(-25)}$	$E_6\times U(1)$

 0022-2488/88/061275-08$02.50

$$L = L_{-1} \oplus L_0 \oplus L_{+1}. \tag{2.1}$$

The elements of L_m satisfy the formal commutations relations

$$[L_m, L_n] \subseteq L_{m+n}, \quad m,n = -1,0,1, \tag{2.2}$$

where $L_{m+n} = 0$ for $|m+n| \geqslant 2$. In Refs. 4 and 5, a simple method was developed for constructing the oscillatorlike unitary irreducible representations (UIR) of noncompact Lie groups G that admit a Jordan decomposition with respect to their maximal compact subgroup K. In this method, which we refer to as the oscillator method, the elements of L are realized as bilinears of an arbitrary number of bosonic oscillators (annihilation and creation operators) satisfying the usual canonical commutation relations. In the Fock space of these oscillators one chooses a set of states, denoted as $|\Omega\rangle$ and referred to as the lowest weight state, that is annihilated by all the operators belonging to the L_{-1} space and that transforms in a definite representation of the maximal compact subgroup K. By acting on $|\Omega\rangle$ with the operators belonging to the L_{+1} space repeatedly one generates an infinite set of states that forms the basis of a unitary representation of G. The resulting unitary representation of G is irreducible if the lowest weight state $|\Omega\rangle$ transforms in an irreducible representation of the maximal compact subgroup K. Clearly, these oscillatorlike unitary representations are all of the lowest weight type and the generator whose spectrum is bounded from below corresponds to the number operator for the bosonic oscillators.

Subsequently, a classification of the unitary highest weight representations of simple noncompact groups was given in the mathematical literature.[14] A comparison of the oscillatorlike representations with the classification of Ref. 14 shows that the oscillatorlike UIR's exhaust the list of UIR's of the lowest weight type for the groups $SU(n,m)$, $Sp(2n,R)$, and $SO^*(2n)$. A straightforward application of the oscillator method to the noncompact groups $SO(n,2)$, $E_{6(-14)}$, and $E_{7(-25)}$ gives, in general, reducible unitary representations. A simple systematic algorithm for extracting the irreducible unitary representations of the groups $SO(n,2)$, $E_{6(-14)}$, and $E_{7(-25)}$ within the framework of the oscillator approach is yet to be developed.

We should note that the bilinear operators belonging to the L_{-1} and L_{+1} spaces of the noncompact groups G listed in Table I involve di-annihilation and di-creation operators. The repeated application of the bosonic di-creation operators on the lowest weight state $|\Omega\rangle$ generates an infinite set of states forming the basis of the UIR of G. If one replaces the di-creation and di-annihilation operators by bilinears involving one creation and one annihilation operator then the oscillator method yields the UIR's of the compact forms of the groups listed in Table I. If we use fermionic oscillators instead of the bosonic ones the method yields the representations of the compact forms of these groups as well. To construct UIR's of noncompact groups with fermionic oscillators we need an infinite set of them. Such realizations do occur in string theories[1] and will be the subject of a separate study.

III. UNITARY REPRESENTATIONS OF NONCOMPACT SUPERGROUPS WITH A JORDAN STRUCTURE

The oscillator method for the construction of the unitary representations of noncompact groups[4,5] has been extended to the construction of the UIR's of noncompact supergroups.[6] Later this method was further developed and applied to supergravity and superstring theories.[7-11] The oscillator method developed in these works yields the UIR's of the highest (or lowest) weight type for those noncompact supergroups G that have a (super)-Jordan decomposition with respect to a maximal compact subsupergroup K, i.e., the Lie superalgebra L of G has a three-graded structure with respect to the Lie superalgebra L_0 of a maximal compact subsupergroup K:

$$L = L_{-1} \oplus L_0 \oplus L_{+1}. \tag{3.1}$$

In this case the identity (2.2) holds in a graded sense, i.e.,

$$[L_m, L_n\} \subseteq L_{m+n}, \tag{3.2}$$

with $L_{m+n} = 0$ for $|m+n| \geqslant 2$. The bracket $[\ ,\ \}$ means an anticommutator among any two odd generators of L and a commutator otherwise. The elements of L are now realized as bilinears of bosonic and fermionic oscillators. The even generators correspond to bilinears in purely bosonic or purely fermionic oscillators while the odd generators are realized as bilinears involving one bosonic and one fermionic oscillator. To construct the oscillatorlike unitary representations of the noncompact supergroup G one proceeds as in the case of ordinary Lie groups and considers a set of states $|\Omega\rangle$ in the super-Fock space transforming in a definite representation of the maximal compact subsupergroup K and is annihilated by all the bilinear operators in the L_{-1} space of L. Acting on $|\Omega\rangle$ with the operators belonging to the L_{+1} space repeatedly one generates an infinite set of states that forms the basis of a unitary representation of G. Again the resulting representation is irreducible if and only if the state $|\Omega\rangle$ transforms irreducibly under K. Clearly these representations are of the lowest weight type.

The application of this method to the noncompact Lie superalgebras $SU(n,m/p)$, $OSp(2n/2m,R)$, and $OSp(2n^*/2m)$ (Refs. 6–10) yields readily the UIR's of the lowest weight type of the respective superalgebras. Now the oscillator method can be used to construct all the UIR's of the lowest weight type of the even subgroups of these supergroups. Therefore we expect the oscillatorlike UIR's of the supergroups $SU(n,m/p)$, $OSp(2n/2m,R)$, and $OSp(2n^*/2m)$ to exhaust the list of their UIR's of the lowest weight type. For the other noncompact Lie superalgebras that have a Jordan decomposition with respect to a maximal compact subsuperalgebra the unitary representations obtained by the oscillator method are, in general, reducible. A simple method for decomposing these reducible representations into irreducible pieces has not yet been developed.

IV. JORDAN AND KANTOR DECOMPOSITIONS OF LIE ALGEBRAS AND LIE SUPERALGEBRAS

With the exception of G_2, F_4, and E_8 all finite dimensional simple Lie algebras have a Jordan decomposition[5,15-18] with respect to some subalgebra. Furthermore all

Lie algebras with a Jordan decomposition can be constructed from Jordan triple systems via the Tits–Koecher construction.[15–18] In this construction the elements of the Lie algebra L that belong to the L_{+1} space are labeled by the elements of some Jordan triple system. The commutation relations of the elements of L can all be expressed in terms of the Jordan triple product and the Jacobi identities follow from the two defining identities of Jordan triple systems.[15–18] The Tits–Koecher construction has been generalized to the construction of Lie superalgebras from super Jordan triple systems.[17,18] With the exception of G(3), all "classical" Lie superalgebras as classified by Kac[19] (which include the exceptional and strange superalgebras) have a Jordan decomposition and can be constructed from super Jordan triple systems. A complete list of Lie superalgebras that can be constructed from Jordan superalgebras via the generalized Tits–Koecher method can be found in Ref. 18.

The Tits–Koecher method has been extended by Kantor[16] to give a construction of all finite dimensional simple Lie algebras, including G_2, F_4, and E_8, from more general triple systems. These generalized triple systems, which we shall refer to as the Kantor triple systems, are defined by two identities and include the Jordan triple systems as a subclass. The Kantor construction has also been generalized to give a unified construction of Lie algebras and Lie superalgebras from (super-) Kantor triple systems.[17] In the Kantor construction and its generalization one makes crucial use of a five-dimensional graded decomposition (Kantor structure) of a Lie algebra or a Lie superalgebra L with respect to a subalgebra L_0 of maximal rank

$$L = L_{-1} \oplus L_{-1/2} \oplus L_0 \oplus L_{+1/2} \oplus L_{+1}. \tag{4.1}$$

The elements of L belonging to various subspaces L_r satisfy the (super) commutation relations

$$[L_r, L_s\} \subseteq L_{r+s}, \quad r,s = 0, \mp\tfrac{1}{2}, \mp 1, \tag{4.2}$$

and $L_{r+s} = 0$ if $|r+s| > 1$. The elements belonging to the $L_{1/2}$ space can be labeled by the elements of the underlying (super-) triple system and all the (super) commutators in L can be expressed in terms of the (super-) Kantor triple product. The Jacobi identities follow from the two defining identities of the (super-) Kantor triple systems.[17] All simple Lie algebras and classical Lie superalgebras have a Kantor decomposition with respect to a subalgebra of maximal rank and can be constructed in this manner. Clearly, the subspace of L consisting of L_{-1}, L_0, and L_{+1} spaces form a subalgebra with a Jordan structure.

Of noncompact simple Lie groups only those that have a Jordan structure with respect to their maximal compact subgroups have unitary representations of the lowest (or highest) weight type.[14] However, this is not true for noncompact Lie supergroups. As we shall discuss in detail in Sec. VI, the noncompact Lie superalgebra $OSp(2n + 1/2m,R)$ has unitary representations of the lowest weight type but does not have a Jordan structure with respect to a compact subsuperalgebra of maximal rank. This can easily be seen from the fact that the even subgroup $SO(2n+1) \times Sp(2m,R)$ has a Jordan structure with respect to its subgroup $SO(2n-1) \times SO(2) \times U(m)$. However, $OSp(2n + 1/2m,R)$ does not, in general, have a subsupergroup whose even subgroup is $SO(2n-1) \times SO(2) \times U(m)$ except for some special values of n and m. Therefore to be able to construct the UIR's of the lowest weight type of all the noncompact Lie superalgebras of the classical type the oscillator method must be generalized to Lie superalgebras that have a Kantor structure with respect to a compact subsuperalgebra.

V. THE UNITARY REPRESENTATIONS OF THE LOWEST WEIGHT TYPE OF NONCOMPACT LIE SUPERGROUPS WITH A KANTOR STRUCTURE

Consider a noncompact Lie superalgebra L with a Kantor decomposition with respect to a suitable compact subalgebra L_0 of maximal rank:

$$L = L_{-1} \oplus L_{-1/2} \oplus L_0 \oplus L_{+1/2} \oplus L_{+1}. \tag{5.1}$$

The subalgebra L_0 contains the generator N of an Abelian U(1) subgroup that gives the grading, i.e.,

$$[N, L_r] = rL_r, \quad r = 0, \mp\tfrac{1}{2}, \mp 1. \tag{5.2}$$

Assume that L is realized in terms of bosonic and fermionic oscillators. Now choose a set of states $|\Omega\rangle$ in the super-Fock space of these oscillators that transforms in a definite representation of the subsuperalgebra L_0 and is annihilated by all the operators belonging to the $L_{-1/2}$ space

$$L_{-1/2}|\Omega\rangle = 0 \tag{5.3}$$

(which implies that $L_{-1}|\Omega\rangle = 0$). Then the set of states generated by the repeated application of the operators belonging to the $L_{+1/2}$ space on the state $|\Omega\rangle$ forms the basis of a representation of L. This representation will be a unitary one for a certain real form of L. The easiest way to determine the real form for which the above representation is unitary is to go to a super-Hermitian basis of L.[6] In the super-Hermitian basis the even generators are Hermitian operators and the odd generators are anti-Hermitian. Furthermore the structure constants are pure imaginary for commutator products and pure real for anticommutators. For details on this point we refer the reader to Ref. 6. The Kantor decomposition of L corresponds to a split basis, in general, and to go to a super-Hermitian basis one needs to take definite linear combinations of the operator belonging to the subspaces L_0, $L_{\mp 1/2}$, and $L_{\mp 1}$. We shall implicitly assume that we are dealing with that real form of the supergroup in question for which the oscillator method as outlined above yields unitary representations.

If the lowest weight state $|\Omega\rangle$ transforms in an irreducible representation of L_0 then the unitary representation of L obtained by repeated application of the operators belonging to the $L_{+1/2}$ space will also be irreducible. To prove this we need to show that all the Casimir operators of L are diagonalized by all the states that form the basis of the representation. Now the Casimir operators are polynomial functions of the generators of L. Each term of the Casimir operator has a vanishing U(1) charge generated by N. Therefore by (super) commutation of the generators one can bring all the Casimir operators to a form such that the operators belonging to $L_{-1/2}$ and L_{-1} spaces are to the right of the operators belonging to L_0, $L_{+1/2}$, and L_{+1} spaces in each term. For example, symbolically, a term of the form $L_{-1}L_0L_{1/2}L_{1/2}$ can be brought to the form $(L_0L_{1/2}L_{1/2}L_{-1} + L_0L_{1/2}L_{-1/2}$

$+L_0^2+L_{1/2}L_{-1/2}+L_{1/2}L_{1/2}L_{-1}$). Then acting on the lowest weight state $|\Omega\rangle$ with the Casimir operators we find that only those terms involving the generators of L_0 alone give a nonvanishing contribution. Since $|\Omega\rangle$ transforms irreducibly under L_0 all the Casimir operators will simultaneously be diagonalized by it. Similarly, the states generated by the action of the operators in $L_{+1/2}$ space on $|\Omega\rangle$ will diagonalize all the Casimir operators since the operators in $L_{+1/2}$ and L_{+1} spaces commute with the Casimir operators of L. This proves that if the lowest weight vector $|\Omega\rangle$ is an irreducible representation of L_0 then the states generated by the action of $L_{1/2}$ on $|\Omega\rangle$ will form the basis of a UIR of L. In some very special cases the eigenvalues of the Casimir operators of L do not uniquely label the UIR's. In such cases one can use the L_0 labels of $|\Omega\rangle$ to uniquely label the UIR's of L.

VI. THE LOWEST WEIGHT UNITARY IRREDUCIBLE REPRESENTATIONS OF OSp(2n+1/2m,R)

In this section we shall apply the method outlined above to construct UIR's of the lowest weight type of the noncompact Lie supergroup $OSp(2n+1/2m,R)$ whose even subgroup is $SO(2n+1)\times Sp(2m,R)$. The group $SO(2n+1)$ has a Kantor structure with respect to its subgroup $U(n)$. We can decompose the generators of $SO(2n+1)$ in a split basis as

$$\begin{aligned} &L=L_{-1}\oplus L_{-1/2}\oplus L_0\oplus L_{+1/2}\oplus L_{+1},\\ &L=A_{\mu\nu}\oplus K_\mu\oplus I^\mu_\nu\oplus K^\mu\oplus A^{\mu\nu}, \end{aligned} \tag{6.1}$$

where the operators belonging to various subspaces satisfy the commutation relations

$$\begin{aligned} &[I^\mu_\nu,I^\rho_\sigma]=\delta^\rho_\nu I^\mu_\sigma-\delta^\mu_\sigma I^\rho_\nu,\quad [K_\mu,K_\nu]=A_{\mu\nu},\\ &[K_\mu,K^\nu]=-I^\nu_\mu,\quad [K_\mu,A^{\nu\lambda}]=\delta^\nu_\mu K^\lambda-\delta^\lambda_\mu K^\nu,\\ &[I^\mu_\nu,K_\lambda]=-\delta^\mu_\lambda K_\nu,\quad [I^\mu_\nu,K^\lambda]=\delta^\lambda_\nu K^\mu,\\ &[I^\mu_\nu,A_{\lambda\rho}]=\delta^\mu_\lambda A_{\rho\nu}-\delta^\mu_\rho A_{\lambda\nu},\\ &[I^\mu_\nu,A^{\lambda\rho}]=\delta^\rho_\nu A^{\lambda\mu}-\delta^\lambda_\nu A^{\rho\mu},\\ &[A_{\mu\nu},A^{\lambda\rho}]=-\delta^\rho_\mu I^\lambda_\nu+\delta^\lambda_\mu I^\rho_\nu-\delta^\lambda_\nu I^\rho_\mu+\delta^\rho_\nu I^\lambda_\mu,\\ &[K^\mu,A_{\lambda\rho}]=\delta^\mu_\lambda K_\rho-\delta^\mu_\rho K_\lambda. \end{aligned} \tag{6.2}$$

One can realize the generator of $SO(2n+1)$ in terms of a single set of fermionic annihilation and creation operators transforming covariantly and contravariantly, respectively, under the $U(n)$ subgroup:

$$\{\xi_\mu,\xi^\nu\}=\delta^\nu_\mu,\quad \{\xi_\mu,\xi_\nu\}=\{\xi^\nu,\xi^\mu\}=0,$$
$$\mu,\nu,\ldots=1,2,\ldots,n. \tag{6.3}$$

The generators of $SO(2n+1)$ in a Kantor basis are then given as

$$\begin{aligned} &A_{\mu\nu}=\xi_\mu\xi_\nu,\quad K_\mu=(1/\sqrt{2})\xi_\mu,\\ &I^\mu_\nu=\tfrac{1}{2}(\xi^\mu\xi_\nu-\xi_\nu\xi^\mu),\\ &K^\mu=(1/\sqrt{2})\xi^\mu,\quad A^{\mu\nu}=\xi^\mu\xi^\nu. \end{aligned} \tag{6.4}$$

The only state in the fermionic Fock space that is annihilated by the operators K_μ belonging to the $L_{-1/2}$ space and transforms irreducibly under $U(n)$ generated by $I_\nu{}^\mu$ is the Fock vacuum $|0\rangle$:

$$K_\mu|0\rangle=\xi_\mu|0\rangle=0. \tag{6.5}$$

Choosing the vacuum $|0\rangle$ as the lowest weight state and acting on it by the operators K^μ belonging to the $L_{+1/2}$ space we generate the basis of a UIR of $SO(2n+1)$:

$$\begin{aligned} &|0\rangle,\\ &K^\mu|0\rangle=\xi^\mu|0\rangle,\\ &K^\mu K^\nu|9\rangle=\xi^\mu\xi^\nu|0\rangle\\ &\underbrace{K^\mu K^\nu\cdots K^\lambda}_{n\text{ times}}|0\rangle=\xi^\mu\xi^\nu\cdots\xi^\lambda|0\rangle. \end{aligned} \tag{6.6}$$

The resulting representation is simply the spinor representation of $SO(2n+1)$. Since it is constructed out of a single set of oscillators we refer to it as the singleton representation[7-10] of $SO(2n+1)$.

One can also realize the $SO(2n+1)$ Lie algebra using a pair of fermionic oscillators α_μ,β_μ carrying the $U(n)$ index and a single fermionic oscillator ψ,

$$\begin{aligned} &\{\alpha_\mu,\alpha^\nu\}=\delta^\nu_\mu=\{\beta_\mu,\beta^\nu\},\\ &\{\alpha_\mu,\alpha_\nu\}=\{\beta_\mu,\beta_\nu\}=\{\alpha_\mu,\beta_\nu\}=\{\alpha_\mu,\beta^\nu\}=0,\\ &\{\psi,\psi^\dagger\}=1,\\ &\{\alpha_\mu,\psi\}=\{\alpha_\mu,\psi^\dagger\}=\{\beta_\mu,\psi\}=\{\beta_\mu,\psi^\dagger\}=0. \end{aligned} \tag{6.7}$$

The generators of $SO(2n+1)$ in a Kantor basis are then given by

$$\begin{aligned} &A_{\mu\nu}=\alpha_\mu\beta_\nu-\alpha_\nu\beta_\mu,\quad K_\mu=\psi\beta_\mu+\alpha_\mu\psi^\dagger,\\ &I^\mu_\nu=\alpha^\mu\alpha_\nu-\beta_\nu\beta^\mu,\\ &K^\mu=\psi\alpha^\mu+\beta^\mu\psi^\dagger,\quad A^{\mu\nu}=\alpha^\mu\beta^\nu-\alpha^\nu\beta^\mu. \end{aligned} \tag{6.8}$$

The vacuum state $|0\rangle$ in the Fock space of all the fermionic oscillators is annihilated by all the annihilation operators

$$\alpha_\mu|0\rangle=\beta_\mu|0\rangle=\psi|0\rangle=0.$$

Any set of states of the form

$$\beta^\mu\beta^\nu\cdots\beta^\lambda|0\rangle \tag{6.9}$$

or of the form

$$\psi^\dagger\alpha^\mu\alpha^\nu\cdots\alpha^\lambda|0\rangle \tag{6.10}$$

is annihilated by the operators K_μ belonging to the $L_{-1/2}$ space. Any subset of these states that transforms irreducibly under the $U(n)$ subgroup generated by L_0 can be used as a lowest weight state for the construction of a UIR of $SO(2n+1)$. The set of states of the form (6.9) and of the form (6.10) lead to the same set of UIR's of $SO(2n+1)$ since they have identical $U(n)$ transformation properties. With one set of the oscillators α_μ, β_μ, and ψ one can construct $(n+1)$ inequivalent irreducible lowest weight states, thus leading to $(n+1)$ inequivalent representations of $SO(2n+1)$. One can construct all the UIR's of $SO(2n+1)$ by the oscillator method. In the general case, one realizes the generators of $SO(2n+1)$ in terms of an arbitrary number of oscillators as follows:

$$\begin{aligned} A_{\mu\nu}&=\boldsymbol{\alpha}_\mu\cdot\boldsymbol{\beta}_\nu-\boldsymbol{\alpha}_\nu\cdot\boldsymbol{\beta}_\mu+\epsilon\xi_\mu\xi_\nu\\ &\equiv\sum_{r=1}^{p}(\alpha_\mu(r)\beta_\nu(r)-\alpha_\nu(r)\beta_\mu(r))+\epsilon\xi_\mu\xi_\nu, \end{aligned}$$

$$K_\mu = \psi\cdot\beta_\mu + \alpha_\mu\cdot\psi^\dagger + (\epsilon/\sqrt{2})\xi_\mu$$
$$\equiv \sum_{k=1}^{p} (\psi(r)\beta_\mu(r) + \alpha_\mu(r)\psi^\dagger(r)) + \frac{\epsilon}{\sqrt{2}}\xi_\mu,$$

$$I_\nu^\mu = \alpha^\mu\cdot\alpha_\nu - \beta_\nu\cdot\beta^\mu + (\epsilon/2)(\xi^\mu\xi_\nu - \xi_\nu\xi^\mu)$$
$$\equiv \sum_{r=1}^{p} (\alpha^\mu(r)\alpha_\nu(r) - \beta_\nu(r)\beta^\mu(r))$$
$$+ (\epsilon/2)(\xi^\mu\xi_\nu - \xi_\nu\xi^\mu), \tag{6.11}$$

$$K^\mu = \psi\cdot\alpha^\mu + \beta^\mu\cdot\psi^\dagger + (\epsilon/\sqrt{2})\,\xi^\mu$$
$$\equiv \sum_{r=1}^{p} (\psi(r)\alpha^\mu(r) + \beta^\mu(r)\psi^\dagger(r)) + \frac{\epsilon}{\sqrt{2}}\xi^\mu,$$

$$A^{\mu\nu} = \alpha^\mu\beta^\nu - \alpha^\nu\cdot\beta^\mu + \epsilon\xi^\mu\xi^\nu$$
$$\equiv \sum_{r=1}^{p} (\alpha^\mu(r)\beta^\nu(r) - \alpha^\nu(r)\beta^\mu(r)) + \epsilon\xi^\mu\xi^\nu,$$

where $\epsilon = 0,1$ and the oscillators $\alpha_\mu(r), \beta_\mu(r)$, and $\psi(r)$ satisfy the canonical anticommutation relations (CAR's)

$$\begin{aligned}
&\{\alpha_\mu(r),\alpha^\nu(s)\} = \delta_\mu^\nu\delta_{rs}, \quad r,s = 1,\ldots,p,\\
&\{\beta_\mu(r),\beta^\nu(s)\} = \delta_\mu^\nu\delta_{rs}, \quad \mu,\nu = 1,\ldots,n,\\
&\{\alpha_\mu(r),\alpha_\nu(s)\} = \{\beta_\mu(r),\beta_\nu(s)\} = \{\alpha_\mu(r),\beta^\nu(s)\} = 0,\\
&\{\psi(r),\psi^\dagger(s)\} = \delta_{rs},\\
&\{\psi(r),\psi(s)\} = \{\psi(r),\alpha_\mu(s)\} = \{\psi(r),\beta_\mu(s)\}\\
&\qquad = \{\psi(r),\alpha^\mu(s)\} = \{\psi(r),\beta^\mu(s)\} = 0.
\end{aligned} \tag{6.12}$$

Interestingly enough, for $\epsilon = 0$ the oscillator method yields vectorial or tensorial type representations of $SO(2n+1)$. While for $\epsilon = 1$ we obtain in general spinorial representations of $SO(2n+1)$. In the general case the possible lowest weight states of $SO(2n+1)$ are of the form

$$[\xi^\mu]^s\psi^\dagger(1)[\alpha^\mu(1)]^{n_1}\psi^\dagger(2)[\alpha^\nu(2)]^{n_2}$$
$$\times\cdots\psi^\dagger(p)[\alpha^\lambda(p)]^{n_p}|0\rangle \tag{6.13}$$

or of the form

$$[\xi^\mu]^s[\beta^\mu(1)]^{n_1}\cdots[\beta^\lambda(p)]^{n_p}|0\rangle, \tag{6.14}$$

where $s = 0,1$ and $[\alpha^\mu(k)]^{n_k}|0\rangle$ stands for a state of the form $\alpha^{\mu_1}(k)\alpha^{\mu_2}(k)\cdots\alpha^{\mu_{n_k}}(k)|0\rangle$ and similarly for $[\beta^\mu(k)]^{n_k}|0\rangle$ ($n_k \leqslant n$). Every irreducible representation of $SO(2n+1)$ can be constructed by choosing a suitable lowest weight state of the form (6.13) or (6.14) that transforms irreducibly under $U(n)$ generated by I_ν^μ and acting on it by the operators K^μ repeatedly.

The noncompact subgroup $Sp(2m,R)$ of $OSp(2n+1/2m,R)$ has a Jordan structure with respect to its maximal compact subgroup $U(m)$.[5,8] The simplest realization of the Lie algebra of $Sp(2m,R)$ is in terms of a single set of m bosonic annihilation and creation operators satisfying

$$[c_i,c^j] = \delta_i^j, \quad [c_i,c_j] = 0, \quad i,j = 1,2,\ldots,m. \tag{6.15}$$

Then the generators of $Sp(2m,R)$ in a Jordan basis read as

$$L = L_{-1}\oplus L_0\oplus L_{+1} = S_{ij}\oplus M_j^i\oplus S^{ij}, \tag{6.16}$$

where

$$S_{ij} = c_ic_j, \quad S^{ij} = c^ic^j, \quad M_j^i = \tfrac{1}{2}(c^ic_j + c_jc^i). \tag{6.17}$$

In the corresponding bosonic Fock space there exist two inequivalent lowest weight states of $Sp(2m,R)$ annihilated by the operators S_{ij} belonging to the L_{-1} space and that transforms irreducibly under the maximal compact subgroup $U(m)$ generated by M_j^i. They are the vacuum $|0\rangle$ which is an $SU(m)$ singlet and the one-particle state $c^i|0\rangle$ that transforms in the fundamental (contravariant) representation of $U(m)$. Acting on $|0\rangle$ or on $c^i|0\rangle$ repeatedly by the operators S^{ij} we generate an infinite set of states that form the basis of a UIR of $Sp(2m,R)$. These two UIR's are referred to as singleton representations.[8] Again the oscillator method yields all the UIR's of the lowest weight type of $Sp(2m,R)$. To construct these UIR's we need to realize the generators of $Sp(2m,R)$ in terms of an arbitrary set of bosonic oscillators:

$$\begin{aligned}
S_{ij} &= \mathbf{a}_i\cdot\mathbf{b}_j + \mathbf{a}_j\cdot\mathbf{b}_i + \epsilon c_ic_j\\
&\equiv \sum_{r=1}^{p} (a_i(r)b_j(r) + a_j(r)b_i(r)) + \epsilon c_ic_j,\\
S^{ij} &= \mathbf{a}^i\cdot\mathbf{b}^j + \mathbf{a}^j\cdot\mathbf{b}^i + \epsilon c^ic^j\\
&\equiv \sum_{r=1}^{p} (a^i(r)b^j(r) + a^j(r)b^i(r)) + \epsilon c^ic^j,\\
M_j^i &= \mathbf{a}^i\cdot\mathbf{a}_j + \mathbf{b}_j\cdot\mathbf{b}^i + (\epsilon/2)\,(c^ic_j + c_jc^i)\\
&\equiv \sum_{r=1}^{p} (a^i(r)a_j(r) + b_j(r)b^i(r))\\
&\quad + (\epsilon/2)(c^ic_j + c_jc^i),
\end{aligned} \tag{6.18}$$

where the bosonic oscillators satisfy

$$\begin{aligned}
&[a_i(r),a^j(s)] = \delta_i^j\delta_{rs},\\
&[b_i(r),b^j(s)] = \delta_i^j\delta_{rs},\\
&[a_i(r),b_j(s)] = [a_i(r),b^j(s)] = [a_i(r),a_j(s)]\\
&\qquad = [b_i(r),b_j(s)] = 0.
\end{aligned} \tag{6.19}$$

The parameter ϵ takes on the values 0 or 1, depending on whether we have an even ($2p$) or odd ($2p+1$) number of bosonic oscillators carrying $U(m)$ indices. The representations of $Sp(2m,R)$ with $\epsilon = 1$ are the noncompact analog of the spinorial representations of the orthogonal group $SO(n)$. The lowest weight states for the construction of a unitary representation of $Sp(2m,R)$ are, in general, tensor products of the states of the form

$$a^{[i}(r)b^{j]}(r)|0\rangle \equiv (a^i(r)b^j(r) - a^j(r)b^i(r))|0\rangle \tag{6.20}$$

tensored with those of the form

$$[c^i]^s[a^j(1)]^{n_1}\cdots[a^k(r)]^{n_r}$$
$$\times[b^l(r+1)]^{n_{r+1}}\cdots[b^m(p)]^{n_p}|0\rangle, \tag{6.21}$$

where $s = 0,1$ and $n_1,\ldots,n_p$ are arbitrary non-negative integers. Here it is important to make sure that family indices $(r,s,\ldots)$ of the states of the form (6.20) do not coincide with each other or with those of (6.21). Again the expression $[a^j(k)]^{n_k}$ represents

$$a^{j_1}(k)a^{j_2}(k)\cdots a^{j_{n_k}}(k).$$

Choosing a subset of these states that transforms irreducibly under $U(m)$ and acting on it with S^{ij} repeatedly we generate the basis of a UIR of $Sp(2m,R)$. These UIR's are uniquely labeled by their lowest weight states.

The Kantor decomposition of $SO(2n+1)$ and the Jordan decomposition of $Sp(2m,R)$ can be embedded in a Kantor decomposition of the noncompact Lie superalgebra $OSp(2n+1/2m,R)$. To do this supercovariantly we combine the bosonic and fermionic oscillators used in the construction of $Sp(2m,R)$ and $SO(2n+1)$ into superoscillators whose first m components are bosonic and remaining n components are fermionic, i.e.,

$$\zeta_A = \begin{pmatrix} a_i \\ \alpha_\mu \end{pmatrix}, \quad \eta_A = \begin{pmatrix} b_i \\ \beta_\mu \end{pmatrix},$$
$$\zeta^A = \begin{pmatrix} a^i \\ \alpha^\mu \end{pmatrix}, \quad \eta^A = \begin{pmatrix} b^i \\ \beta^\mu \end{pmatrix}. \tag{6.22}$$

The Lie superalgebra L of $OSp(2n+1/2m,R)$ can be realized as bilinears of these superoscillators in a Kantor basis as

$$L = L_{-1} \oplus L_{-1/2} \oplus L_0 \oplus L_{+1/2} \oplus L_{+1},$$
$$L = L_{AB} \oplus L_A \oplus L^A{}_B \oplus L^A \oplus L^{AB}, \tag{6.23}$$

where

$$L_A = \psi \cdot \eta_A - \psi^\dagger \cdot \zeta_A \equiv \sum_{r=1}^{p} (\psi(r)\eta_A(r) - \psi^\dagger(r)\zeta_A(r)),$$
$$L_{AB} = \zeta_A \cdot \eta_B + \zeta_B \cdot \eta_A \equiv \sum_{r=1}^{p} (\zeta_A(r)\eta_B(r) + \zeta_B(r)\eta_A(r)),$$
$$L^A{}_B = \zeta^A \cdot \zeta_B + (-1)^{\deg A \deg B} \eta_A \cdot \eta^B$$
$$\equiv \sum_{r=1}^{p} (\zeta^A(r)\zeta_B(r) + (-1)^{\deg A \deg B}\eta_A(r)\eta^B(r)), \tag{6.24}$$
$$L^A = \eta^A \cdot \psi^\dagger - \zeta^A \cdot \psi \equiv \sum_{r=1}^{p} (\eta^A(r)\psi^\dagger(r) - \zeta^A(r)\psi(r)),$$
$$L^{AB} = \zeta^A \cdot \eta^B + \zeta^B \cdot \eta^A \equiv \sum_{r=1}^{p} (\zeta^A(r)\eta^B(r) + \zeta^B(r)\eta^A(r)),$$

where deg A is zero or unity depending on whether the index A refers to a bosonic or fermionic oscillator, respectively. The operators L_A and L^A belonging to the $L_{-1/2}$ and $L_{+1/2}$ subspaces transform in the covariant and contravariant fundamental representation of the compact subsuperalgebra $U(m/n)$ generated by $L_B{}^A$. The states in the super-Fock space that are annihilated by operators belonging to $L_{-1/2}$ space are, in general, linear combinations of the states of the form

$$[\zeta^a(1)]^{r_1}[\zeta^B(2)]^{r_2}\cdots[\zeta^C(p)]^{r_p}\psi^\dagger(1)\cdots\psi^\dagger(p)|0\rangle \tag{6.25}$$

or of the form

$$[\eta^A(1)]^{r_1}\cdots[\eta^C(p)]^{r_p}|0\rangle. \tag{6.26}$$

Choosing a subset $|\Omega\rangle$ of these states that transforms irreducibly under the compact subsuperalgebra $U(m/n)$ and acting on it repeatedly with the operators K^A belonging to the $K_{1/2}$ space we generate an infinite set of states that forms the basis of a UIR or $OSp(2n+1/2m,R)$. Since the method as explained above yields all the irreducible representations of $SO(2n+1)$ and all the lowest weight UIR's of $Sp(2m,R)$ we believe that the UIR's obtained by this method exhausts the list of all lowest weight UIR's of $OSp(2n+1/2m,R)$.

To illustrate how the method works let us consider the example of $OSp(7/4,R)$ which is simply the $N=7$ extended anti-de Sitter superalgebra in four space-time dimensions. [Recall that $Sp(4,R)$ is the twofold covering group of $SO(3,2)$.] For reasons that will become clear shortly we shall take one set of the oscillators ζ_A, η_A, and ψ (i.e., $p=1$) and choose as our lowest weight state the Fock vacuum $|0\rangle$, which is a singlet of the compact subsuperalgebra $SU(2/3)$ with a definite $U(1)$ charge. By acting on $|0\rangle$ with the operators L^A we generate states that transform in a definite representation of $U(2/3)$. Using supertableaux techniques[7,8,20] we can designate the transformation properties of these states under $U(2/3)$:

$$|0\rangle,$$
$$L^A|0\rangle = |\boxed{}\rangle,$$
$$L^A L^B|0\rangle = |\boxed{}\rangle \oplus |\begin{array}{l}\boxed{}\\ \boxed{}\end{array}\rangle, \tag{6.27}$$
$$L^A L^B L^C|0\rangle = |\boxed{}\rangle \oplus |\begin{array}{l}\boxed{}\\ \boxed{}\end{array}\rangle,$$
$$\vdots \qquad\qquad \vdots,$$

where the supertableaux occurring on the RHS have at most two rows.

The above decomposition of the UIR of $OSp(7/4,R)$ is with respect to its compact subsupergroup $U(2/3)$. For most physical applications, such as to supergravity, it turns out to be more useful to decompose the UIR of a noncompact supergroup with respect to its even subgroup. In our example, the even subgroup is $SO(7)\times Sp(4,R)$. The infinite set of states in (6.27) forming a UIR of $OSp(7/4,R)$ can be combined into certain UIR's of $Sp(4,R)$ that have definite $SO(7)$ transformation properties. Since the UIR's of $Sp(4,R)$ that occur are of the lowest weight type they can be labeled by the quantum numbers of the maximal compact subgroup $SU(2)\times U(1)$ of $Sp(4,R)$ which are spin s and anti-de Sitter energy E_0.[8,21] Therefore our task is simply to determine all the states in (6.27) that are simultaneously lowest weight states for both $SO(7)$ and $Sp(4,R)$. These lowest states are transformed into each other by the action of the odd supersymmetry generators K^i and $S^{i\mu}$. The $SO(7)\times Sp(4,R)$ content of the UIR of $OSp(7/4,R)$ with lowest weight vector $|0\rangle$ and $p=1$ is given in Table II.

The UIR's of $Sp(4,R)$ can be identified with the Fourier modes of fields defined in four-dimensional anti-de Sitter space.[8,22] The fields associated with the UIR listed in Table II are all massless[8,22] and the full unitary supermultiplet has the same field content as the massless $N=8$ anti-de Sitter supermultiplet in $d=4$ whose highest spin field is the graviton. This is simply an extension to anti-de Sitter space of the well-known result from Poincaré supersymmetry.

Before concluding this section we should point out that the Lie superalgebra $OSp(1/2m,R)$ has a special realization in terms of bosonic oscillators alone. The generators of $OSp(1/2m,R)$ in a Kantor basis take the simple form

$$L_{ij} = a_i a_j, \quad L_i = a_i, \quad L^i_j = a^i a_j - a_j a^i,$$
$$L^i = a^i, \quad L^{ij} = a^i a^j, \quad i,j = 1,\ldots,m, \tag{6.28}$$

TABLE II. We give the SO(7)×Sp(4,R) decomposition of the unitary irreducible supermultiplet of OSp(7/4,R) with lowest weight vector $|0\rangle$ and $p=1$. The first column lists the states that contain simultaneous lowest weight states of both SO(7) and Sp(4,R). The second and third columns give the spin [SU(2)] and anti-de Sitter energy E_0(U(1)) of these lowest weight states. The last column gives the SO(7) representations of these states with definite spin and energy E_0.

Lowest states	Spin	E_0	SO(7)
$\|0\rangle$	0	1	35
$K^iK^j\|0\rangle$	0	2	35
$K^i\|0\rangle$	$\frac{1}{2}$	$\frac{3}{2}$	35
$S^{i\mu}\|0\rangle$	$\frac{1}{2}$	$\frac{3}{2}$	21
$S^{i\mu}S^{j\nu}\|0\rangle$	1	2	7
$S^{i\mu}K^j\|0\rangle$	1	2	21
$S^{j\mu}S^{k\nu}K^i\|0\rangle$	$\frac{3}{2}$	$\frac{5}{2}$	7
$S^{i\mu}S^{j\nu}S^{k\rho}\|0\rangle$	$\frac{3}{2}$	$\frac{5}{2}$	1
$S^{j\mu}S^{k\nu}S^{l\rho}K^i\|0\rangle$	2	3	1

where the bosonic oscillators satisfy the usual canonical commutation relations

$$[a_i,a^j]=\delta_i^j, \quad [a_i,a_j]=[a^i,a^j]=0. \tag{6.29}$$

The only state in the bosonic Fock space annihilated by the operator a_i belonging to the $L_{-1/2}$ space that transforms irreducibly under the U(m) subgroup generated by $L^i{}_j$ is the Fock vacuum $|0\rangle$. Acting on $|0\rangle$ repeatedly by the operators L^i we generate the basis of a UIR of OSp(1/2m,R). This shows that the full spectrum of states in the Fock space of m bosonic oscillators forms a single UIR of OSp(1/2m,R). Note that the even subgroup Sp(2m,R) transforms states with even (or odd) number of bosonic excitations into states with even (or odd) numbers of bosonic excitations. The odd generators belonging to the $L_{-1/2}$ and $L_{+1/2}$ spaces mix states having an even number of excitations with states having an odd number of excitations. In other words, the odd supersymmetry generators of OSp(1/2m,R) interpolate between the two singleton representations of Sp(2m,R).

VII. SUPERCOHERENT STATES FOR THE UNITARY HIGHEST WEIGHT REPRESENTATIONS

In Ref. 6 the concept of supercoherent states was first introduced and the coherent state basis for the highest weight UIR's of the noncompact supergroup SU(m,p/n) was studied. The definition of supercoherent states given for SU(m,p/n) extends in a straightforward manner to all noncompact supergroups with a Jordan structure with respect to a maximal compact subsupergroup. The coherent states of such a noncompact supergroup g with a maximal compact subsupergroup K can be labeled by the complex supercoordinates of the quotient space G/K. More specifically if the Lie superalgebra L of G has a Jordan decomposition with respect to the Lie superalgebra L_0 of K, i.e.,

$$L=L_{-1}\oplus L_0\oplus L_{+1},$$

then the coherent state associated with the UIR of lowest weight $|\Omega\rangle$ is defined as

$$|\Omega;Z\rangle=e^{L_{+1}\cdot Z}|\Omega\rangle, \tag{7.1}$$

where $L_{+1}\cdot Z$ represents the operators in the L_{+1} space multiplied with their corresponding complex parameters. For even generators these parameters are complex numbers and for odd genrators they are complex anticommuting Grassmann numbers. Under the unitary action U(g) of the supergroup corresponding to the element $g\in G$, the coherent states with label Z undergo transformations of the form

$$U(g)|\Omega;Z\rangle=|(\gamma Z+\delta)^{-1}\Omega;(\alpha Z+\beta)(\gamma Z+\delta)^{-1}\rangle, \tag{7.2}$$

where

$$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}=g\in G$$

and $|(\gamma Z+\delta)^{-1}\Omega\rangle$ represents the state obtained from the lowest weight state $|\Omega\rangle$ by the induced action of the maximal compact subsupergroup. The matrices α,β,γ, and δ involve ordinary as well as Grassmann parameters. For further details we refer the reader to Ref. 6.

Our aim here is to generalize the concept of supercoherent states to those noncompact supergroups G that have a Kantor structure with respect to a compact subsupergroup K of maximal rank. Now the Lie superalgebra L of G decomposes with respect to the Lie superalgebra L_0 of K as

$$L=L_{-1}\oplus L_{-1/2}\oplus L_0\oplus L_{+1/2}\oplus L_{+1}.$$

Given a lowest weight UIR with the lowest weight state $|\Omega\rangle$ we define the corresponding coherent state $|\Omega;Y,Z\rangle$ as

$$|\Omega;Y,Z\rangle\equiv e^{L_{+1/2}\cdot Y+L_{+1}\cdot Z}|\Omega\rangle, \tag{7.3}$$

where Y and Z are the superparameters associated with the generators belonging to the $L_{+1/2}$ and L_{+1} space and $L_{+1/2}\cdot Y$ and $L_{+1}\cdot Z$ represent the generators multiplied with their parameters. Since the subsupergroup generated by L_{-1}, L_0, and L_{+1} has a Jordan structure with respect to the compact subsupergroup K the label Z will be the coherent state label of this subsupergroup as discussed above.

For the noncompact supergroup OSp(2n + 1/2m,R) the supercoherent states defined by the lowest weight vector $|\Omega\rangle$ of a UIR has the form

$$e^{L^AY_A+L^{AB}Z_{AB}}|\Omega\rangle=|\Omega;Y,Z\rangle. \tag{7.4}$$

To illustrate the structure of these supercoherent states consider the supergroup OSp(1/2m,R). The supercoherent state associated with the singleton UIR of OSp(1/2m,R) has a very simple decomposition

$$|0;\theta_i,Z_{ij}\rangle\equiv e^{\theta_ic^i+Z_{ij}c^ic^j}|0\rangle=(1+\theta_ic^i)e^{Z_{ij}c^ic^j}|0\rangle$$
$$=|0;Z_{ij}\rangle+\theta_kc^k|0;Z_{ij}\rangle, \tag{7.5}$$

where the θ_i are complex anticommuting parameters and the Z_{ij} are the complex parameters labeling the usual coherent states of Sp(2m,R). The coherent states $|0;Z_{ij}\rangle$ and $c^k|0;Z_{ij}\rangle$ are the coherent states associated with the two singleton UIR's of Sp(2m,R) with lowest weight states $|0\rangle$ and $c^i|0\rangle$. Under the action of the odd generators of OSp(1/2m,R) these two coherent states get transformed into each other.

As mentioned above the labels of the super coherent states can be identified with the complex (super) coordinates of the (super) coset space G/K. For OSp(1/2m,R) the

complex coordinates θ_i and Z_{ij} can be chosen as the coordinates of the quotient space $\mathrm{OSp}(1/2m,R)/U(m)$. For the simplest case of $m = 1$, the $\mathrm{OSp}(1/2,R)$ invariant volume element $d\mu$ on the quotient space $\mathrm{OSp}(1/2,R)/U(1)$ with coordinates θ and Z can be obtained simply using the results of Ref. 23:

$$d\mu = [(1 - \tfrac{1}{4}\bar{\theta}\theta)/(1 - \bar{z}z)^2]dz\, d\bar{z}\, d\bar{\theta}\, d\theta. \tag{7.6}$$

This can be generalized in a straightforward manner to general m. The study of the coherent states of $\mathrm{OSp}(2n + 1/2m,R)$ as defined above requires a detailed knowledge of the quotient space $\mathrm{SO}(2n + 1)/\mathrm{U}(n)$. We hope to return to this and related issues in a separate study.

VIII. CONCLUSION

In this paper we gave a generalization of the oscillator method so as to be able to construct lowest weight UIR's of noncompact Lie superalgebras that have a Kantor structure with respect to a compact subsuperalgebra of maximal rank. With this generalization the oscillator method can be used to construct the lowest weight unitary representations of certain real forms of *all* the classical simple Lie superalgebras,[19] which include the exceptional and strange superalgebras. The relevant real form of the classical supergroups are those for which the even subgroup is a direct product of a compact group with a simple noncompact (or compact) group. As mentioned above in certain cases the oscillator method does not lead directly to irreducible unitary representations. A simple algorithm for projecting out the irreducible representations in these instances is yet to be developed. We have also given a definition of generalized supercoherent states associated with the lowest weight UIR's of noncompact supergroups with a Kantor structure with respect to a compact subsupergroup of maximal rank. The extension of the method to construct unitary representations of noncompact supergroups whose even subgroups contain direct products of two simple noncompact groups will be the subject of a separate study.

[1]For an excellent book on the subject see the book by M. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge U. P., Cambridge, 1987), Vols. I and II.

[2]For a review of supergravity theories see P. van Nieuwenhuizen, Phys. Rep. **68**, 189 (1981).

[3]A. B. Balantekin, I. Bars, and F. Iachello, Nucl. Phys. A **370**, 284 (1981).

[4]M. Günaydin and C. Saclioglu, Phys. Lett. B **108**, 169 (1982); M. Günaydin, "Unitary realization of the non-compact symmetry groups of supergravity," in *Proceedings of the 2nd Europhysics Study Conference on Unification of the Fundamental Interactions*, edited by J. Ellis and S. Ferrara (Plenum, New York, 1983), p. 193.

[5]M. Günaydin and C. Saclioglu, Commun. Math. Phys. **87**, 159 (1982).

[6]I. Bars and M. Günaydin, Commun. Math. Phys. **91**, 21 (1983); M. Günaydin, "Oscillator-like unitary representations of non-compact groups and supergroups and extended supergravity theories," in *Proceedings of the XIth International Colloquium on Group Theoretical Methods in Physics*, edited by M. Serdaroğlu and E. Inönü, *Lecture Notes in Physics*, Vol. 180 (Springer, Berlin, 1982), p. 192.

[7]M. Günaydin, P. van Vieuwenhuizen, and N. P. Warner, Nucl. Phys. B **255**, 63 (1985).

[8]M. Günaydin and N. P. Warner, Nucl. Phys. B **272**, 99 (1986).

[9]M. Günaydin and N. Marcus, Class. Quantum Gravity **2**, L11, L19 (1985).

[10]M. Günaydin, G. Sierra, and P. K. Townsend, Nucl. Phys. B **274**, 429 (1986); M. Günaydin, B. E. W. Nilsson, G. Sierra, and P. K. Townsend, Phys. Lett. B **171**, 182 (1986).

[11]M. Günaydin, L. J. Romans, and N. P. Warner, Nucl. Phys. B **272**, 598 (1986).

[12]Harish-Chandra, Am. J. Math. **77**, 743 (1955); **78**, 1 (1956).

[13]J. A. Wolf, J. Math. Mech. **13**, 489 (1964).

[14]T. Enright, R. Howe, and W. Wallach, "A classification of unitary highest weight modules," in *Representations of Reductive Groups*, edited by P. Trombi (Birkhäuser, Boston, 1983), pp. 97–143.

[15]J. Tits, Nederl. Akad. Wetensch. **65**, 530 (1962); M. Koecher, Am. J. Math. **89**, 787 (1967).

[16]I. L. Kantor, Trudy Sem. Vector. Anal. **16**, 407 (1972) (in Russian).

[17]I. Bars and M. Günaydin, J. Math. Phys. **20** (1979) 1977.

[18]M. Günaydin, in *Proceedings of the 8th International Colloquium on Group Theoretical Methods in Physics*, Ann. Israel Phys. Soc. **3**, 279 (1980).

[19]V. G. Kac, Adv. Math. **26**, 8 (1977).

[20]A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1149, 1810 (1981); **23**, 1239 (1982); J. Q. Chen, X. G. Chen, and M. G. Gao, J. Phys. A **16**, 1361 (1983).

[21]C. Fronsdal, Rev. Mod. Phys. **37**, 221 (1965); Phys. Rev. D **10**, 589 (1974); **12**, 3819 (1975); C. Fronsdal and R. B. Hagen, *ibid.* **12**, 3810 (1975).

[22]D. Z. Freedman and H. Nicolai, Nucl. Phys. B **237**, 342 (1984).

[23]F. A. Berezin and V. N. Tolstoy, Commun. Math. Phys. **78**, 409 (1981).

A Grassmann calculus for infinite spin groups

John Palmer
Mathematics Department, University of Arizona, Tucson, Arizona 85721

(Received 5 August 1987; accepted for publication 3 February 1988)

The correlation functions of the two-dimensional Ising model satisfy nonlinear equations both on the lattice (discovered by McCoy and Wu [Nucl. Phys. B **180**, 89 (1981)] and Perk [Phys. Lett. A1 **79**, 3 (1980)]) and in a continuum limit (discovered by Wu, McCoy, Tracy, and Barouch and generalized by Sato, Miwa, and Jimbo [Publ. RIMS **14**, 223 (1978)]). In this paper results are presented for infinite spin groups that extend results of Sato, Miwa, and Jimbo [Publ. RIMS **14**, 223 (1978)] and lead to normal ordered product formulas for the difference of adjacent Ising spin fields. These product formulas are shown to lead directly to the lattice difference equations for the correlations and are also a key ingredient in the scaling limit analysis.

I. INTRODUCTION

Let W denote an infinite-dimensional Hilbert space with a distinguished conjugation and associated bilinear form. In this paper we generalize some of the results proved in Ref. 1 for the finite-dimensional spin groups to the infinite-dimensional complex spin groups, $\hat{\mathrm{Spin}}_Q(W)$, introduced in Ref. 2. In Ref. 3 some such infinite-dimensional generalizations were already given. In that paper, however, the structure of the group $\hat{\mathrm{Spin}}_Q(W)$ was incompletely understood and the results in Ref. 3 are consequently limited to a subsemigroup of $\hat{\mathrm{Spin}}_Q(W)$.

The group $\hat{\mathrm{Spin}}_Q(W)$ acts by linear transformation on a dense invariant subspace $\mathscr{D}$ of an appropriate Fock space. Let $L(\mathscr{D})$ denote the space of linear maps from $\mathscr{D}$ into $\mathscr{D}$. The Fock representation of $\hat{\mathrm{Spin}}_Q(W)$ is a homomorphism $\hat{\Gamma}: \hat{\mathrm{Spin}}_Q(W) \to L(\mathscr{D})$ the image of which consists of transformations that implement automorphisms of the Clifford algebra $\mathscr{C}(W)$ associated with elements of the restricted orthogonal group $\mathrm{SO}_{\mathrm{res}}(W)$.[2] As is explained in Sec. II one may extend the action of $\hat{\mathrm{Spin}}_Q(W)$ by the invertible elements in $W \subseteq \mathscr{C}(W)$ to get the group $\hat{\mathrm{Pin}}_Q(W)$. One obtains in this fashion the largest group of transformations implementing automorphisms of $\mathscr{C}(W)$ induced by complex orthogonals on W that can be defined on a domain containing the vacuum vector in Fock space. The association of $\hat{\mathrm{Pin}}_Q(W)$ with a "maximal" family of Bogoliubov transformations is one of the principal motivations for the study of this group.

Let $A_0(W)$ denote the algebraic alternating tensor algebra over W consisting of finite sums of finite products from W. In Sec. III of this paper we introduce a map $N: A_0(W) \times \hat{\mathrm{Spin}}_Q(W) \to L(\mathscr{D})$ that extends the notion of "normal" ordering familiar in the physics literature. One of the principal results of this section is a "structure" theorem (Theorem 3.0) for the image of $\hat{\mathrm{Pin}}_Q(W)$ under $\hat{\Gamma}$ [this result has implications for a cell decomposition of $\hat{\mathrm{Pin}}_Q(W)$ but this is not pursued here]. A second application of the normal ordering map is a "perturbation" result (Theorem 3.3) that expresses $\hat{\Gamma}(g_2)$ in terms of $\hat{\Gamma}(g_1)$ when the elements g_1 and g_2 in $\hat{\mathrm{Spin}}_Q(W)$ have induced rotations that differ by a finite rank operator.

In Sec. IV we prove a generalized version of Wick's theorem for $\hat{\mathrm{Spin}}_Q(W)$ (Theorem 4.2). The proof we give is a simplification (and generalization) of the proof in Ref. 3 and is based on an idea in Ref. 1. It avoids the rather heavy product deformation formalism of Sato, Miwa, and Jimbo (SMJ) that was used in the proof in Ref. 3. If one combines Theorem 3.0 with Theorem 4.2 one obtains a generic "reduction" formula for the vacuum expectation of a product of elements from $\hat{\mathrm{Pin}}_Q(W)$. Formulas for the vacuum expectations of the "elementary types" to which the reduction formula leads are not given here but see Refs. 1 and 3.

In the final section of this paper we give an application of the results of Secs. III and IV to the correlations of the two-dimensional Ising model. The results of Sec. III lead to "perturbation formulas" for the Ising field $\sigma(a+u)$ incremented by a unit lattice vector u in terms of $\sigma(a)$ (Theorem 5.1). A much less elegant version of these local difference identities can be found in Theorem 2.1 of Ref. 4. There they were a principal ingredient in the SMJ analysis of the scaled correlations. The derivation of the local difference identities presented in Sec. V of this paper is not only more direct than that given in Ref. 4 but it has the virtue of suggesting the role of the "disorder variables"[5] that arise as a matter of course in the analysis. This is very useful when considering generalizations to monodromy fields[6] where there is not yet an independent indentification of the appropriate disorder variables. In a sequel to this paper we will derive analogs of the local difference identities for monodromy fields. These are used in the thesis of Davey to establish the SMJ anaysis of the scaled monodromy correlations. We conclude this paper by showing that the difference identitites of Theorem 5.1 also lead directly to the McCoy, Wu, and Perk (MWP)[7,8] difference relations for the Ising correlations. The connection is via the generalized version of Wick's theorem and no doubt must be closely related to Perk's original derivation of these identities[9] which also relies on the "thermodynamic" version of Wick's theorem. That the local difference identities (5.15) lead to the MWP difference identities and are a crucial ingredient in the SMJ analysis of the scaled correlations reveals their significance for the "deeper" analysis of the Ising correlations.

What is new in this paper is mostly the formulation and calculation with Ising model correlations directly in the infinite volume limit. The Ising fields are, in this limit, elements

 0022-2488/88/061283-17$02.50

of an infinite-dimensional spin group that in general consists of *unbounded* operators acting on Fock space. The approximation of these group elements by algebraic elements in the Clifford algebra is somewhat subtle but one of the main points of this paper is that in spite of the fact that the representation of the complex spin group does not respect the Hilbert space structure of Fock space many useful tools from the algebraic finite-dimensional setting carry over to infinite spin groups if they are properly formulated. There is, for example, no useful normal ordered product formula for the Ising field operators themselves but the notion of "relative normal ordered products" introduced in Sec. III leads to a fruitful generalization, Theorem 3.3, of an obvious finite-dimensional result that may then be applied to the Ising field to get the new result (5.15). This result may then be used as the cornerstone of both the McCoy, Wu, and Perk analysis and the Sato, Miwa, and Jimbo analysis of the Ising model correlations. It is the author's hope that the additional complications of the infinite volume analysis will eventually be compensated by conceptual simplicity. If one recalls that the study of phase transitions is framed in the infinite volume limit precisely to obtain the conceptual simplicity of sharp discontinuities this hope may not seem completely misplaced.

II. THE GROUP $\hat{\mathrm{Spin}}_Q(W)$

In this section we first review the setting and some of the results for infinite spin groups that we will need. More details and proofs can be found in Ref. 2. Let W denote an infinite-dimensional complex Hilbert space with a distinguished congugation P. We write $\langle\cdot,\cdot\rangle$ for the Hermitian inner product on W and $(\cdot,\cdot)=\langle\cdot,P\cdot\rangle$ for the associated symmetric bilinear form. The algebraic Clifford algebra $\mathscr{C}_0(W)$ is an associative algebra with identity e generated by the elements of W (and e) subject to the relations

$$xy+yx=(x,y)e, \quad x,y\in W.$$

The algebra $\mathscr{C}_0(W)$ consists of finite sums of finite products from W and multiples of e. A subspace V of W is said to be isotropic if the bilinear form $(\cdot,\cdot)$ vanishes identically on V. Each splitting of $W=W_+\oplus W_-$ into subspaces $W_\pm$ that are isotropic and orthogonal to one another with respect to the Hermitian inner product gives rise to a representation of $\mathscr{C}_0(W)$ on the complex alternating tensor algebra $A(W_+)$ that we will now describe. The space

$$A(W_+)=\mathbf{C}\oplus\sum_{k=1}^{\infty}\Lambda^k(W_+),$$

where $\Lambda^k(W_+)$ is the Hilbert space completion of the alternating k tensors over W_+. We write $1=1\oplus 0\oplus 0\cdots$ for the "vacuum" vector in $A(W_+)$. For each $x\in W_+$ one may define a creation operator $a^*(x)$ and its Hermitian adjoint $a(x)$ acting on $A(W_+)$ so that (i) $a(x)1=0$, for all $x\in W_+$; (ii) $a^*(x)a(y)+a(y)a^*(x)=\langle x,y\rangle I$; (iii) $a^*(x)a^*(y)+a^*(y)a^*(x)=0$. This construction is described in more detail in Ref. 2. In terms of these annihilation and creation operators one constructs a representation of $\mathscr{C}_0(W)$ as follows. The identity e is mapped to I acting on $A(W_+)$ and the generator $x\in W\subseteq\mathscr{C}_0(W)$ is mapped to

$$F(x)\overset{\text{def}}{=}a^*(x_+)+a(\bar{x}_-),$$

where $x=x_++x_-$, $x_\pm\in W_\pm$ and $\bar{x}_-=Px_-$. This representation is called a Fock representation of $\mathscr{C}_0(W)$.

There are many isotropic splittings of W and it is convenient to parametrize them by the involution Q that is Id on vectors in W_+ and $-\mathrm{Id}$ on vectors in W_-. Let $*$ denote the Hermitian adjoint for operators on W and let τ denote the transpose for operators on W relative to the symmetric bilinear form on W. Then a linear involution Q will have ± 1 eigenspaces which are isotropic iff $Q^\tau=-Q$ and these eigenspaces will be Hermitian orthogonal iff $Q^*=Q$. Unless otherwise indicated the involutions we consider will satisfy both these conditions. We write $Q_\pm=(I\pm Q)/2$ for the projections on $W_\pm$. When we wish to emphasize the choice of Q we will write $F_Q(\cdot)$ for the Fock representation of $\mathscr{C}_0(W)$ based on the decomposition $W=Q_+W\oplus Q_-W$. The vacuum vector in $A(W_+)(W_+=Q_+W)$ will be denoted by 1_Q and we let

$$\mathscr{C}_0(W)\ni X\to\langle X\rangle_Q\overset{\text{def}}{=}\langle F_Q(X)1_Q,1_Q\rangle$$

denote the associated state on the Clifford algebra.

Let $G_0(W)$ denote the group of invertible elements $g\in\mathscr{C}_0(W)$ such that $gxg^{-1}=Gx$ for $x\in W\subseteq\mathscr{C}_0(W)$ and G a complex orthogonal on W. We will say that a map G on W is complex orthogonal if G is invertible and G preserves the symmetric bilinear form $(\cdot,\cdot)$ on W. Thus G is complex orthogonal if and only if $G^\tau G=I$. The elements $g\in G_0(W)$ are either even or odd in the Clifford algebra[2] and we define

$$gxg^{-1}=\begin{cases}T(g)x, & \text{for } g \text{ even in } \mathscr{C}_0(W),\\ -T(g)x, & \text{for } g \text{ odd in } \mathscr{C}_0(W).\end{cases}\tag{2.1}$$

The map $T\colon G_0(W)\to O_0(W)$ is a surjective homomorphism onto the group of complex orthogonals, $O(W)$, which consists of finite rank perturbations of the identity on W. The sign change in (2.1) is introduced precisely so this last statement is true. We will next describe the subgroup $\mathrm{Spin}_0(W)$ of $G_0(W)$. The identity on W extends uniquely to a linear involution τ on $\mathscr{C}_0(W)$ such that $(XY)^\tau=Y^\tau X^\tau$ for $X,Y\in\mathscr{C}_0(W)$. The map $G_0(W)\ni g\to g^\tau g$ is a homomorphism from $G_0(W)$ into $\mathbf{C}^*$ known as the spinor norm. We write $\mathrm{nr}(g)=g^\tau g$. The kernel for the homomorphism nr is a subgroup of $G_0(W)$ we denote by $\mathrm{Pin}_0(W)$. The group

$$\mathrm{Spin}_0(W)=\{g\in\mathrm{Pin}_0(W)\,|\,T(g)\in\mathrm{SO}_0(W)\},$$

where $\mathrm{SO}_0(W)$ is the connected component of the identity in $O_0(W)$. An element $G\in O_0(W)$ is in $\mathrm{SO}_0(W)$ if and only if $\det G=1$. One has the exact sequence

$$\mathbf{Z}/2\mathbf{Z}\to\mathrm{Spin}_0(W)\overset{T}{\to}\mathrm{SO}_0(W)\to 0\tag{2.2}$$

(see Ref. 2).

For G a linear transformation on W we write

$$G=\begin{bmatrix}A(G) & B(G)\\ C(G) & D(G)\end{bmatrix}$$

for the matrix of G relative to the $W_+\oplus W_-$ decomposition of W [or perhaps $A_Q(G)$, etc., if we wish to identify the dependence on the choice of isotropic splitting for W]. If $g\in G_0(W)$ and $G=T(g)$ then we also write

$$G=\begin{bmatrix}A(g) & B(g)\\ C(g) & D(g)\end{bmatrix}$$

for this same matrix to avoid the clumsy notation $A(T(g))$, etc. We now define a metric d_Q on $\mathrm{Spin}_0(W)$ as follows:

$$d_Q(g_1,g_2) = \|A(g_1) - A(g_2)\|_1 + \|D(g_1) - D(g_2)\|_1 + \|B(g_1) - B(g_2)\|_2 + \|C(g_1) - C(g_2)\|_2 + \|F_Q(g_1)1_Q - F_Q(g_2)1_Q\|,$$

where $\|\cdot\|_1$ is the trace norm, $\|\cdot\|_2$ is the Schmidt norm, and $\|\cdot\|$ is the Hilbert space norm on $A(W_+)$. We have also written $A(g_i) = A_Q(g_i)$, etc., for brevity. One of the principal results of Ref. 2 is that the sequential closure of $\mathrm{Spin}_0(W)$ in the metric d_Q is a *continuous* topological group $\mathrm{Spin}_Q(W)$ in the induced metric topology. Let $\mathrm{O}_Q(W)$ denote the group of invertible complex orthogonals on W that have a diagonal element $\left[\begin{smallmatrix} A & 0 \\ 0 & D \end{smallmatrix}\right]$ that is a trace class perturbation of the identity on W and an off diagonal element $\left[\begin{smallmatrix} 0 & B \\ C & 0 \end{smallmatrix}\right]$ that is a Schmidt class operator. Let $\mathrm{SO}_Q(W)$ denote the connected component of the identity in $\mathrm{O}_Q(W)$. The exact sequence (2.2) survives under completion to give an exact sequence:

$$\mathbf{Z}/2\mathbf{Z} \to \mathrm{Spin}_Q(W) \xrightarrow{T} \mathrm{SO}_Q(W) \to 0. \tag{2.3}$$

Furthermore, there is a dense domain $\mathscr{D} \subseteq A(W_+)$ that contains 1_Q and is invariant under the Fock representation of $\mathscr{C}_0(W)$ such that the Fock representation of $\mathrm{Spin}_0(W)$ extends to a strongly continuous representation Γ_Q: $\mathrm{Spin}_Q(W) \to L(\mathscr{D})$. Here $L(\mathscr{D})$ denotes the collection of the linear maps from $\mathscr{D}$ into $\mathscr{D}$. What we mean by strong continuity is that the map $\mathrm{Spin}_Q(W) \ni g \to \Gamma_Q(g)v \in \mathscr{D}$ defined for each $v \in \mathscr{D}$ is continuous from $\mathrm{Spin}_Q(W)$ into $A(W_+)$. The representation Γ_Q acts on the Fock representation of $\mathscr{C}_0(W)$ as follows:

$$\Gamma_Q(g)F_Q(x)\Gamma_Q(g)^{-1} = F_Q(T(g)x), \tag{2.4}$$

for $g \in \mathrm{Spin}_Q(W)$, $x \in W$, and both sides are understood as elements of $L(\mathscr{D})$.

The elements of the group of invertible complex orthogonals on W that commute with Q have diagonal matrices $A \oplus D$ with $D = A^{-\tau}$. We may thus identify this group with $\mathrm{GL}(W_+)$. For $A \in \mathrm{GL}(W_+)$ the map $W \ni x \to A \oplus A^{-\tau}$ extends to an automorphism of $\mathscr{C}_0(W)$. It is shown in Ref. 2 that this induces a continuous automorphism $\alpha(A)$ acting on $\mathrm{Spin}_Q(W)$. The map $A \to \alpha(A)$ is a homomorphism and

$$T(\alpha(A)g) = (A \oplus A^{-\tau})T(g)(A \oplus A^{-\tau})^{-1}.$$

There is a representation Γ: $\mathrm{GL}(W_+) \to L(\mathscr{D})$ that fits together with Γ_Q in the following manner. Let $\mathrm{Spin}_Q(W) \times_\alpha \mathrm{GL}(W_+)$ denote the semidirect product with composition rule: $g_1 \times A_1 \cdot g_2 \times A_2 = g_1\alpha(A_1)g_2 \times A_1A_2$. Then $\Gamma_Q \times \Gamma(g \times A) = \Gamma_Q(g)\Gamma(A)$ gives a representation of $\mathrm{Spin}_Q(W) \times_\alpha \mathrm{GL}(W_+)$ on $L(\mathscr{D})$. Let ker denote the kernel of this representation and define

$$\hat{\mathrm{Spin}}_Q(W) \overset{\mathrm{def}}{=} \mathrm{Spin}_Q(W) \times_\alpha \mathrm{GL}(W_+)/\mathrm{ker}.$$

The map $\Gamma_Q \times \Gamma$ induces a representation $\hat{\Gamma}_Q$: $\hat{\mathrm{Spin}}_Q(W) \to L(\mathscr{D})$. If we let $T(g \times A) = T(g) \cdot (A \oplus A^{-\tau})$ then this map induces a homomorphism T: $\hat{\mathrm{Spin}}_Q(W) \to \mathrm{SO}_{\mathrm{res}}(W)$, where $\mathrm{SO}_{\mathrm{res}}(W)$ is the connected component of the identity in the group $\mathrm{O}_{\mathrm{res}}(W)$ of complex orthogonals on W that have Schmidt class commutators with Q. We have

$$\hat{\Gamma}_Q(g)F_Q(x)\hat{\Gamma}_Q(g)^{-1} = F_Q(T(g)x), \quad g \in \hat{\mathrm{Spin}}_Q(W), \quad x \in W,$$

regarded as an equality of maps in $L(\mathscr{D})$. The group $\hat{\mathrm{Spin}}_Q(W)$ is not a twofold cover of $\mathrm{SO}_{\mathrm{res}}(W)$. Instead one has the exact sequence

$$\mathbf{C}^* \to \hat{\mathrm{Spin}}_Q(W) \xrightarrow{T} \mathrm{SO}_{\mathrm{res}}(W) \to 0 \tag{2.5}$$

so that $\hat{\mathrm{Spin}}_Q(W)$ is an extension of $\mathrm{SO}_{\mathrm{res}}(W)$ by $\mathbf{C}^*$. Interesting central extensions of loop groups (associated with Kac–Moody algebras) and central extensions of the diffeomorphism group of the circle (associated with Virosoro algebras) lie buried in (2.5) for special choices of W and Q.[10]

If we combine the action of $\hat{\mathrm{Spin}}_Q(W)$ with the action of $\mathscr{C}_0(W)$ on $\mathscr{D}$ then we may "cover" the action (by automorphisms) of the restricted orthogonal group $\mathrm{O}_{\mathrm{res}}(W)$ on $\mathscr{C}_0(W)$ in the following sense. Recall that a linear map on a Hilbert space is said to be Fredholm if its kernel and cokernel are finite dimensional. The difference of the dimensions of these two finite-dimensional spaces is called the index. If $G \in \mathrm{O}_{\mathrm{res}}(W)$ then $A(G)$ and $D(G)$ are Fredholm maps with index 0.[2] The map

$$G \to \epsilon(G) \overset{\mathrm{def}}{=} (-1)^{\dim \ker D(G)}$$

is a homomorphism from $\mathrm{O}_{\mathrm{res}}(W)$ to $\{\pm 1\}$.[2] The kernel of this homomorphism is $\mathrm{SO}_{\mathrm{res}}$. Given $G \in \mathrm{O}_{\mathrm{res}}(W)$ we will show how to construct an invertible linear transformation $g \in L(\mathscr{D})$ so that

$$gF_Q(x)g^{-1} = \epsilon(G)F_Q(Gx) \tag{2.6}$$

is an equality on $\mathscr{D}$. When (2.6) is satisfied for $g \in L(\mathscr{D})$ we will say that g implements G in the Fock representation. The sign $\epsilon(G)$ is introduced in (2.6) to make the extension of (2.1) natural. However, the reader should note that the map $-I$ commutes with Q on W and $\Gamma(-I)F_Q(x)\Gamma(-I)^{-1} = -F_Q(x)$. Thus if one can implement the automorphism $F_Q(x) \to -F_Q(Gx)$ as in (2.6) then one can also implement the automorphism $F_Q(x) \to F_Q(Gx)$.

If $G \in \mathrm{SO}_{\mathrm{res}}(W)$ then G is implemented in the Fock representation by an element of $\hat{\mathrm{Spin}}_Q(W)$. Thus in order to prove that a g satisfying (2.6) exists it suffices to consider the case in which $D(G)$ has an odd-dimensional kernel. In this event $\epsilon(G) = -1$. We will reduce this to the $\mathrm{SO}_{\mathrm{res}}(W)$ case by finding a complex orthogonal reflection $\mathrm{O}(w)$ so that $\mathrm{O}(w)G \in \mathrm{SO}_{\mathrm{res}}(W)$. Let $w \in W$ be such that $(w,w) \neq 0$. Then the complex orthogonal reflection in the hyperplane perpendicular to w is given by

$$\mathrm{O}(w)x = x - 2[(x,w)/(w,w)]w.$$

It is straightforward to check that

$$F_Q(w)F_Q(x)F_Q(w)^{-1} = -F_Q(\mathrm{O}(w)x), \quad x \in W.$$

Suppose now that $\mathrm{O}(w)G \in \mathrm{SO}_{\mathrm{res}}(W)$ and choose $g' \in \hat{\mathrm{Spin}}_Q(W)$ so that $\hat{\Gamma}(g')$ implements $\mathrm{O}(w)G$ in the Fock representation. Then $F_Q(w)\hat{\Gamma}(g')$ implements $-\mathrm{O}(w) \cdot \mathrm{O}(w)G = -G = \epsilon(G)G$ in the Fock representation.

It remains to determine $w \in W$ so that $\mathrm{O}(w)G \in \mathrm{SO}_{\mathrm{res}}(W)$. Let $G = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ denote the matrix of G relative

to the $W_+ \oplus W_-$ decomposition of W. Choose w so that $(w,w) = 2$. Then, for $x \in W$,

$$D(\mathrm{O}(w)G)x = Dx - (Gx,w)w_-, \tag{2.7}$$

where $w_- = Q_- w$. We want to choose w so that the null space of $D(\mathrm{O}(w)G)$ is smaller than the null space of $D = D(G)$. Equation (2.7) suggests that one ought to choose w_- in a complement to the range D. If this is done the right-hand side of (2.7) will vanish only if $Dx = 0$ and $(Gx,w) = 0$ separately. For such a choice it is clear that $\ker D(\mathrm{O}(w)G) \subseteq \ker D(G)$. Now suppose $u \in \ker D$. Since G is orthogonal we have $D^\tau B + B^\tau D = 0$. It follows that B maps $\ker D$ into $\ker D^\tau$. Note that the restriction of B to $\ker D$ does not have a null space since G is invertible. Let $w = c(Bu + \overline{Bu})$, where the constant c is chosen so that $(w,w) = 2c^2(Bu, \overline{Bu}) = 2$ (we write $\bar{x} = Px$ as above). Observe now that $Q_- w = c\,\overline{Bu} \in \overline{\ker D^\tau} = \ker D^* =$ orthogonal complement of the range of D. Thus the kernel of $D(\mathrm{O}(w)G)$ is contained in the kernel of $D(G)$. Furthermore,

$$\begin{aligned} D(\mathrm{O}(w)G)u &= Du - (Gu,w)w_- \\ &= -c(Bu, Bu + \overline{Bu})w_- \\ &= -c^{-1}w_- \neq 0. \end{aligned}$$

Thus the kernel of $D(\mathrm{O}(w)G)$ is strictly contained in the kernel of $D(G)$. By choosing a basis for $\ker D$ that contains u and is orthogonal with respect to the inner product $(u,v)_B = (Bu, \overline{Bu})$ $(u,v \in \ker D)$ one sees immediately that $\dim \ker D(\mathrm{O}(w)G) = \dim \ker D(G) - 1$. It follows that $\mathrm{O}(w)G \in \mathrm{SO}_{\mathrm{res}}(W)$. We have finished the proof that $\mathrm{O}_{\mathrm{res}}(W)$ is implementable in the Fock representation.

We may summarize these developments in a slightly more abstract fashion as follows. Let W_{nr} denote the elements of W with spinor norm 1 (i.e., $w \in W_{\mathrm{nr}}$ iff $w^2 = 1$) then $\mathrm{Pin}_0(W)$ is an "extension" of $\mathrm{Spin}_0(W)$ by W_{nr} in the following sense. Each element g' in $\mathrm{Pin}_0(W)$ is either in $\mathrm{Spin}_0(W)$ or there is a $w \in W_{\mathrm{nr}}$ and $g \in \mathrm{Spin}_0(W)$ such that $g' = wg$. We may thus think of $\mathrm{Pin}_0(W)$ as the union of $\mathrm{Spin}_0(W)$ with $W_{\mathrm{nr}} \times \mathrm{Spin}_0(W)/\sim$, where the equivalence relation $\sim$ on $W_{\mathrm{nr}} \times \mathrm{Spin}_0(W)$ is $w \times g \sim w' \times g'$ if and only if $w'^{-1}w = g'g^{-1}$ in $\mathrm{Spin}_0(W)$. The multiplication rules in $\mathrm{Pin}_0(W)$ are

$$\begin{aligned} &w_1 \times g_1 \cdot g_2 = w_1 \times g_1 g_2, \\ &g_1 \cdot w_2 \times g_2 = T(g_1)w_2 \times g_1 g_2, \\ &(w_1 \times g_1)\cdot(w_2 \times g_2) = w_1 T(g_1) w_2 g_1 g_2, \end{aligned} \tag{2.8}$$

where $w_j \in W_{\mathrm{nr}}$, $g_j \in \mathrm{Spin}_0(W)$ $(j = 1,2)$ and we used the fact that $w_1 T(g_1) w_2$ is in $\mathrm{Spin}_0(W)$ when $w_j \in W$. The multiplication rules (2.8) extend to the union of $\mathrm{Spin}_Q(W)$ with $W_{\mathrm{nr}} \times \mathrm{Spin}_Q(W)/\sim$ and makes this union a group we denote by $\mathrm{Pin}_Q(W)$. The map T: $\mathrm{Pin}_Q(W) \to \mathrm{O}_Q(W)$ defined by

$$\begin{aligned} &T(g) = T(g), \quad g \in \mathrm{Spin}_Q(W), \\ &T(w \times g) = \mathrm{O}(w)T(g), \quad w \in W_{\mathrm{nr}}, \quad g \in \mathrm{Spin}_Q(W), \end{aligned}$$

is easily seen to extend the homomorphism T on $\mathrm{Pin}_0(W)$ defined in (2.1). One has the exact sequence:

$$\mathbf{Z}/2\mathbf{Z} \to \mathrm{Pin}_Q(W) \xrightarrow{T} \mathrm{O}_Q(W) \to 0.$$

We may extend the homomorphism Γ_Q: $\mathrm{Spin}_Q(W) \to L(D)$ to $\mathrm{Pin}_Q(W)$ by defining

$$\begin{aligned} &\Gamma_Q(g) = \Gamma_Q(g), \quad g \in \mathrm{Spin}_Q(W), \\ &\Gamma_Q(w \times g) = F_Q(w)\Gamma_Q(g), \quad w \in W_{\mathrm{nr}}, \quad g \in \mathrm{Spin}_Q(W). \end{aligned}$$

The group $\mathrm{GL}(W_+)$ acts on $\mathrm{Pin}_Q(W)$ via the automorphism α defined by

$$\begin{aligned} &\alpha(G)g = \alpha(G)g, \quad g \in \mathrm{Spin}_Q(W), \\ &\alpha(G)w \times g = (G \oplus G^{-\tau})w \times \alpha(G)g, \\ &w \in W_{\mathrm{nr}}, \quad g \in \mathrm{Spin}_Q(W) \end{aligned}$$

[this obviously extends the action of α on $\mathrm{Pin}_0(W)$]. We may thus form the semidirect product $\mathrm{Pin}_Q(W) \times_\alpha \mathrm{GL}(W_+)$. The map $\Gamma_Q \times \Gamma$ is a homomorphism from $\mathrm{Pin}_Q(W) \times_\alpha \mathrm{GL}(W_+)$ into $L(D)$. Define

$$\hat{\mathrm{Pin}}_Q(W) = \mathrm{Pin}_Q(W) \times_\alpha \mathrm{GL}(W_+)/\ker(\Gamma_Q \times \Gamma)$$

and let $\hat{\Gamma}_Q$ denote the induced homomorphism from $\hat{\mathrm{Pin}}_Q(W)$ to $L(D)$. Then $T(g \times G) = T(g)G \oplus G^{-\tau}$ induces a homomorphism from $\hat{\mathrm{Pin}}_Q(W)$ to $\mathrm{O}_{\mathrm{res}}(W)$ for which there is the exact sequence

$$\mathbf{C}^* \to \hat{\mathrm{Pin}}_Q(W) \xrightarrow{T} \mathrm{O}_{\mathrm{res}}(W) \to 0.$$

III. A GRASSMANN CALCULUS FOR $\hat{\mathrm{Spin}}_Q(W)$

In this section we introduce a "relative" Grassmann calculus for $\hat{\mathrm{Spin}}_Q(W)$ inspired by the finite-dimensional calculus in Ref. 1 and the somewhat limited infinite-dimensional extensions presented in Ref. 3. For comparison with Ref. 3 the reader should note that the results in Ref. 3 were established for those $g \in \hat{\mathrm{Spin}}_Q(W)$ with the property that $\hat{\Gamma}_Q(g)$ extends to a bounded linear transformation on $\Lambda(W_+)$.

Suppose $g \in \hat{\mathrm{Spin}}_Q(W)$ and $w_j \in W$ $(j = 1,\ldots,n)$, then we inductively define a linear map $N_Q(w_n \cdots w_1 g) \in L(\mathscr{D})$ as follows:

$$\begin{aligned} &N_Q(g) = \hat{\Gamma}(g), \\ &N_Q(w_n \cdots w_1 g) = F_Q(w_n^+)N_Q(w_{n-1} \cdots w_1 g) \\ &\qquad + (-1)^{n-1} N_Q(w_{n-1} \cdots w_1 g) \\ &\qquad \times F_Q(w_n^-), \end{aligned} \tag{3.1}$$

where $w_n^\pm = Q_\pm w_n$. The motivation for this definition of normal ordering comes from the finite-dimensional situation where every $g \in \mathrm{Spin}_{\mathbf{C}}(W)$ can be expressed as an even normal ordered product.[2] Such an even element does not contribute to the sign change needed to move $F_Q(w_n^-)$ all the way to the right in a normal ordered product.

We now wish to show that the map

$$\Lambda^k(W) \ni w_n \wedge \cdots \wedge w_1 \to N_Q(w_n \cdots w_1 g)$$

is well defined. To show that $N_Q(w_n \cdots w_1 g)$ is antisymmetric in the arguments $w_1, w_2, \ldots, w_n$, it is enough to show that it changes sign when two adjacent w_j are interchanged. By using the reduction formula (3.1) this immediately reduces to the case where w_n and w_{n-1} are interchanged. Two further applications of (3.1) show

$$\begin{aligned} &N_Q(w_n w_{n-1} w_{n-2} \cdots w_1 g) \\ &\quad = -N_Q(w_{n-1} w_n w_{n-2} \cdots w_1 g). \end{aligned}$$

Let $A_0(W)$ denote the algebraic Grassmann algebra over W. The algebra $A_0(W)$ consists of finite sums of products $w_1 \wedge \cdots \wedge w_k$ and multiples of $1 \in \Lambda^0(W) \simeq \mathbf{C}$. Then $N_Q(\cdot g)$ extends by linearity to a map

$$A_0(W) \ni X \to N_Q(Xg) \in L(\mathscr{D}).$$

It is not reasonable to expect to extend this map to a closure of $A_0(W)$ since the domain $\mathscr{D}$ is not closed under infinite sums.

As an illustration of the significance of this notion of normal ordered product we will prove a structure theorem for the maps on $\mathscr{D}$ that implement transformations $G \in O_{res}(W)$ in the Fock representation.

Theorem 3.1: Suppose that $G \in O_{res}(W)$ and that $\dim \ker D(G) = n$. Then there exist vectors $z_1,\ldots,z_n \in W$ [given by (3.4) below] and $g_0 \in \hat{S}pin_Q(W)$ with $\langle g_0 \rangle_Q \neq 0$ such that

$$g \overset{\text{def}}{=} N_Q(z_n \cdots z_1 g_0)$$

implements G in the Fock representation.

Proof: Let $G = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right]$ be the matrix of G relative to the $W_+ \oplus W_-$ decomposition of W. Choose a basis $u_1,\ldots,u_n$ for $\ker D$ that is orthonormal with respect to the inner product $(u,v)_B = (Bu, \overline{Bv})$, $u,v \in \ker D$. Then let $w_i = Bu_i + \overline{Bu_i}$. Define $G_0 = O(w_1) \cdots O(w_n) G$. Then a straightforward calculation shows that, for $x \in W_-$,

$$D(G_0)x = D(G)x - \sum_{j=1}^{n} (Gx, w_j)\, \overline{Bu_j}, \tag{3.2}$$

since $Q_- w_j = \overline{Bu_j}$. The null space of $D(G_0)$ is contained in the null space of $D(G)$ since the vectors $\overline{Bu_j}$ all live in the orthogonal complement of the range of $D(G)$. However, the vectors

$$D(G_0)u_k = -\sum_{j=1}^{n} (Bu_k, Bu_j + \overline{Bu_j})\, \overline{Bu_j} = -\overline{Bu_k}$$

are linearly independent since the restriction of B to $\ker D$ does not have a kernel. Thus $D(G_0)$ does not have a kernel. Since it is Fredholm of index 0 it follows that $D(G_0)$ is invertible and hence that $G_0 \in SO_{res}(W)$. Thus there exists $g_0 \in \hat{S}pin_Q(W)$ such that $\hat{\Gamma}_Q(g_0)$ implements G_0 in the Fock representation. The condition that $D(G_0)$ is invertible is equivalent to $\langle g_0 \rangle_Q \neq 0$.[3] In the remainder of the proof it will be convenient to unburden the notation writing $\hat{\Gamma}$, F, and N for $\hat{\Gamma}_Q$, F_Q, and N_Q.

Since $G = O(w_n) \cdots O(w_1) G_0$ it follows that $F(w_n) \cdots F(w_1) \hat{\Gamma}(g_0)$ implements the automorphism of $\mathscr{C}_0(W)$ induced by G. We will show that $F(w_n) \cdots F(w_1) \hat{\Gamma}(g_0)$ can be expressed as a normal ordered product $N(z_n \cdots z_1 g_0)$. We begin by showing that $F(w_1) \hat{\Gamma}(g_0)$ is a normal ordered product $N(z_1 g_0)$. Suppose $z_1 \in W$ and write $z_1^{\pm} = Q_{\pm} z_1$. Then

$$\begin{aligned} N(z_1 g_0) &= F(z_1^+) \hat{\Gamma}(g_0) + \hat{\Gamma}(g_0) F(z_1^-) \\ &= F(z_1^+ + G_0 z_1^-) \hat{\Gamma}(g_0). \end{aligned}$$

Thus $F(w_1) \hat{\Gamma}(g_0) = N(z_1 g_0)$ if $(G_0 Q_- + Q_+) z_1 = w_1$. In $W_+ \oplus W_-$ coordinates this becomes

$$\begin{bmatrix} I & B_0 \\ 0 & D_0 \end{bmatrix} \begin{bmatrix} z_1^+ \\ z_1^- \end{bmatrix} = \begin{bmatrix} Bu_1 \\ \overline{Bu_1} \end{bmatrix}, \tag{3.3}$$

where $B_0 = B(G_0)$ and $D_0 = D(G_0)$. But (3.2) implies that $D_0 u_1 = -\overline{Bu_1}$. Thus $z_1 = (-u_1) + (B + B_0)u_1$ solves (3.4). Now let $G_l = O(w_l) \cdots O(w_1) G_0$ and suppose for the sake of an inductive argument that

$$g_l \overset{\text{def}}{=} F(w_l) \cdots F(w_1) \hat{\Gamma}(g_0)$$

is a normal ordered product $N(z_l \cdots z_1 g_0)$. We will show that g_{l+1} is a normal ordered product $N(z_{l+1} z_l \cdots z_1 g_0)$. Consider the normal ordered product

$$\begin{aligned} N(z_{l+1} z_l \cdots z_1 g_0) &= F(z_{l+1}^+) g_l + (-1)^l g_l F(z_{l+1}^-) \\ &= F((G_l Q_- + Q_+) z'_{l+1}) g_l, \end{aligned}$$

where $z'_{l+1} = z_{l+1}^+ + (-1)^l z_{l+1}^-$. Thus it will be possible to write $F(w_{l+1}) g_l$ in normal ordered form if one can solve $(G_l Q_- + Q_+) z'_{l+1} = w_{l+1}$. Let $\left[\begin{smallmatrix} A_l & B_l \\ C_l & D_l \end{smallmatrix}\right]$ denote the matrix of G_l relative to the $W_+ \oplus W_-$ decomposition of W.

In $W_+ \oplus W_-$ components the equation $(G_l Q_- + Q_+) z'_{l+1} = w_{l+1}$ becomes

$$\begin{bmatrix} I & B_l \\ 0 & D_l \end{bmatrix} \begin{bmatrix} z_{l+1}^+ \\ (-1)^l z_{l+1}^- \end{bmatrix} = \begin{bmatrix} Bu_{l+1} \\ \overline{Bu_{l+1}} \end{bmatrix}.$$

But the analog of (3.2) for D_l is

$$D_l x = Dx - \sum_{j=l+1}^{n} (G_l x, w_j)\, \overline{Bu_j}.$$

Thus $D_l u_{l+1} = -\overline{Bu_{l+1}}$ and it follows that we may choose $z_{l+1}^- = (-1)^{l+1} u_{l+1}$ with $z_{l+1}^+ = (B + B_l) u_{l+1}$. Hence

$$z_{l+1} = (-1)^{l+1} u_{l+1} + (B + B_l) u_{l+1} \tag{3.4}$$

solves $(G_l Q_- + Q_+) z'_{l+1} = w_{l+1}$, for $l = 0,\ldots,n-1$. This finishes the proof that given $G \in O_{res}(W)$ there exists $g = N(z_n \cdots z_1 g_0)$ that implements G in the Fock representation. Q.E.D.

We next want to review some of the finite-dimensional results for the Grassmann calculus of the Clifford group developed by Sato, Miwa, and Jimbo in Ref. 1. Suppose W is a finite-dimensional complex vector space with nondegenerate symmetric bilinear form $(\cdot,\cdot)$ and distinguished isotropic splitting $W = W_+ \oplus W_-$. Then there is a linear map θ from the Grassmann algebra $A(W)$ onto the Clifford algebra $\mathscr{C}(W)$ determined by the conditions

$$\begin{aligned} &(1)\quad \theta(1) = 1, \quad \theta(x) = x, \quad x \in W, \\ &(2)\quad \theta(X \wedge Y) = \theta(X)\theta(Y), \quad \text{if } x \in A(W_+) \\ &\qquad\qquad \text{or } Y \in A(W_-). \end{aligned} \tag{3.5}$$

It is not hard to see that θ is surjective and a dimension argument then shows that θ is injective. This map may be used to give a calculus for the Clifford group $G(W)$ that is nicely adapted to the Fock representation. To see how this works suppose that $G \in SO(W)$ and that $D(G)$ is invertible. Define $R(G) = (G - I)(Q_- G + Q_+)^{-1}$, where $Q_{\pm}$ is the pair of projections associated with the splitting $W = W_+ \oplus W_-$ as usual. It is easy to check that $R(G)^T = -R(G)$ and that $G = (1 - RQ_-)^{-1}(1 + RQ_+)$ where we have written $R = R(G)$. Furthermore, one may easily prove that

if R is a skew symmetric ($R^\tau = -R$) map on W such that $(1 - RQ_-)$ is invertible, then

$$G \overset{\text{def}}{=} (1 - RQ_-)^{-1}(1 + RQ_+)$$

is in $\mathrm{SO}(W)$ and $D(G)$ is invertible. Let $\{w_j\}$ denote a basis for W and let $\{w_j^*\}$ denote the dual basis defined by $(w_j, w_k^*) = \delta_{jk}$. If R is a linear map on W with $R^\tau = -R$ then the sum $\Sigma_{j=1}^n Rw_j \wedge w_j^*$ is an element of the Grassmann algebra that does not depend on the choice of basis $\{w_j\}$. To avoid introducing extra notation we will denote this element of $\Lambda^2(W)$ by R. The reader should have no difficulty with this abuse of notation since R will be regarded as an element of $\Lambda^2(W)$ only when it appears as an argument of θ or the normal ordering map N_Q. For $R \in \Lambda^2(W)$ we define

$$\exp\left(\frac{1}{2}R\right) = \sum_{n=0}^{\infty} (2^n n!)^{-1} R^n,$$

where $R^n = R \wedge R \wedge \cdots \wedge R$ (n factors). When W is finite dimensional the sum defining this exponential is of course a finite sum. We summarize the results we will need from Refs. 2 and 3 in the following theorem.

Theorem 3.2: Suppose R: $W \to W$ is skew symmetric ($R^\tau = -R$). Then $\theta(\exp \frac{1}{2}R) \in \mathscr{C}(W)$ is in the Clifford group $G(W)$ if and only if $(I - RQ_-)$ is invertible. If $(1 - RQ_-)$ is invertible then $T(\theta(\exp \frac{1}{2}R)) = (1 - RQ_-)^{-1}(1 + RQ_+)$. Furthermore, every element in $G(W)$ has a representation $\theta(z_n \wedge \cdots \wedge z_1 \wedge \exp \frac{1}{2}R)$ for some choice of $z_j \in W$ and skew symmetric R such that $(1 - RQ_-)$ is invertible.

This is proved in Refs. 2 and 3. The reader should have no difficulty making the connection between the second part of this result and Theorem 3.1 above.

We now return to the consideration of the infinite-dimensional situation. In the remainder of this section W will denote an infinite-dimensional Hilbert space as described at the beginning of Sec. II. We are interested in a relative version of the map θ for the infinite spin group $\hat{\mathrm{Spin}}_Q(W)$. Let $g \in \hat{\mathrm{Spin}}_Q(W)$ such that $D(g)$ is invertible. Define

$$R(g) = (T(g) - I)(Q_- T(g) + Q_+)^{-1}.$$

Let $R_Q(W)$ denote the class of linear maps R on W such that $R^\tau = -R$ and such that

$$R = \begin{bmatrix} -D(R)^\tau & B(R) \\ C(R) & D(R) \end{bmatrix}$$

has matrix element $D(R)$ in the trace class and matrix elements $B(R)$ and $C(R)$ in the Schmidt class. Note that $R \in R_Q(W)$ with $(1 - RQ_-)$ invertible gives rise to an element $G = (1 - RQ_-)^{-1}(1 + RQ_+)$ in $\mathrm{SO}_Q(W)$. The topology on $R_Q(W)$ is given by trace norm convergence on the diagonal and Schmidt norm convergence on the off diagonal.

Definition 3.2: Suppose $g \in \hat{\mathrm{Spin}}_Q(W)$ and $\langle g \rangle_Q \neq 0$. Suppose $R \in R_Q(W)$ and that $(I - R'Q_-)$ is invertible where $R' = R + R(g)$. Then we define $\theta(\exp(\frac{1}{2}R)g)$ to be the element of $\hat{\mathrm{Spin}}_Q(W)$ with induced rotation $G' = (I - R'Q_-)^{-1}(I + R'Q_+)$ and vacuum expectation equal to $\langle g \rangle_Q$.

Remark: At the moment the notation $\theta(\exp(\frac{1}{2}R)g)$ is merely suggestive. However, we shall see that in certain circumstances it does make sense to expand the exponential $\exp(\frac{1}{2}R)$.

It is instructive to consider the case where $g = e =$ the identity in $\hat{\mathrm{Spin}}_Q(W)$. If $R \in R_Q(W)$ and $(I - RQ_-)$ is invertible then $\theta(\exp(\frac{1}{2}R)e)$ is the element in $\hat{\mathrm{Spin}}_Q(W)$ with induced rotation $(I - RQ_-)^{-1}(I + RQ_-)$ and vacuum expectation 1. We would now like to make a connection between (3.5) and Definition 3.2. If $R \in R_Q(W)$ is finite rank (rank K, say) then by choosing a basis $\{w_j\}$ so that $w_{k+1}, w_{k+2}, \ldots$ is a basis for the null space of R we may arrange that the sum

$$\sum_{j=1}^{\infty} Rw_j \wedge w_j^* = \sum_{j=1}^{K} Rw_j \wedge w_j^*$$

is a finite sum [here $\{w_j^*\}$ is the dual basis $(w_j^*, w_k) = \delta_{jk}$.] Theorem 3.1 then implies that $\theta(\exp(\frac{1}{2}R))$ has induced rotation $(1 - RQ_-)^{-1}(1 + RQ_+)$ and it is clear that $\langle \theta(\exp(\frac{1}{2}R)) \rangle_Q = 1$. Thus

$$\theta(\exp(\tfrac{1}{2}R)) = \theta(\exp(\tfrac{1}{2}R)e),$$

where on the left θ is defined by (3.5) [$R \in \Lambda_0^2(W)$] and on the right Definition 3.2 applies.

We now consider making "finite-dimensional" approximations to elements $g \in \mathrm{Spin}_Q(W)$ with $\langle g \rangle_Q \neq 0$. The condition $\langle g \rangle_Q \neq 0$ is equivalent to $D(g)$ being invertible which in turn implies $(I - RQ_-)$ is invertible where $R = R(g)$. Thus if we write $G = T(g)$ we have $G = (I - RQ_-)^{-1} \times (I + RQ_+)$. We will now approximate G in $\mathrm{SO}_0(W)$ by making finite rank approximations to R. Let P_n denote a sequence of finite rank orthogonal projections on W converging strongly to the identity with the further property that the range of each P_n is a subspace of W invariant under both P and Q. Define $R_n = P_n R P_n$. It is shown in Ref. 2 at the end of the proof of Theorem 3.4 that $R_n \in R_Q(W)$, R_n converges to R in $R_Q(W)$, for n sufficiently large $(I - R_n Q_-)$ is invertible, and finally that

$$G_n = (I - R_n Q_-)^{-1}(1 + R_n Q_+) \in \mathrm{SO}_0(W)$$

converges to G in $\mathrm{SO}_Q(W)$. By Theorem 2.1 the element $\theta(\exp(\frac{1}{2}R_n))$ in $G_0(W)$ has induced rotation G_n. It is not in $\mathrm{Spin}_0(W)$ in general but we have

$$1 = \langle \theta(\exp(\tfrac{1}{2}R_n)) \rangle_Q^2 = \mathrm{nr}(\theta(\exp(\tfrac{1}{2}R_n)))\det D_n,$$

where $D_n = D(G_n)$ (see Theorem 2.1 in Ref. 2). Thus since $\mathrm{nr}(\cdot)$ is a "quadratic" homomorphism we can put $\theta(\exp(\frac{1}{2}R_n))$ in $\hat{\mathrm{Spin}}_0(W)$ by multiplying it by a square root of $(\det D_n)$. Let λ_n denote a square root of $(\det D_n)$. Choose the sequence λ_n so that it converges to $\langle g \rangle_Q = \sqrt{\det D}$. Then let

$$g_n \overset{\text{def}}{=} \lambda_n \theta(\exp(\tfrac{1}{2}R_n)) \in \mathrm{Spin}_0(W).$$

Since $T(g_n) = G_n$ converges to G in $\mathrm{SO}_Q(W)$ and $\langle g_n \rangle_Q = \lambda_n$ converges to $\langle g \rangle_Q \neq 0$ it follows from Proposition 3.5 in Ref. 2 that g_n converges to g in $\mathrm{Spin}_Q(W)$. Note that $\lambda_n \theta(\exp(\frac{1}{2}R_n))$ gives a "formula" for g_n that depends only on the induced rotation for g_n and a choice of square root for $\det D_n$.

We are now prepared to state the principal result of this

section. As in the proof of Theorem 3.0 it is convenient to unburden the notation writing $\hat{\Gamma}$, F, and N for $\hat{\Gamma}_Q$, F_Q, and N_Q.

Theorem 3.3: Suppose $R \in R_Q(W)$ and R is rank $K < \infty$. Let $g \in \hat{\mathrm{Spin}}_Q(W)$ with $\langle g \rangle_Q \neq 0$ and suppose $I - R'Q_-$ is invertible where $R' = R + R(g)$. Then

$$\hat{\Gamma} \circ \theta\left(\exp\left(\frac{1}{2}R\right)g\right) = \sum_{k=0}^{K} (2^k k!)^{-1} N(R^k g).$$

Proof: We first prove this result for $G \in \mathrm{Spin}_Q(W)$ and then show that the general result for $\hat{\mathrm{Spin}}_Q(W)$ follows from this special case. Let $g \in \mathrm{Spin}_Q(W)$, and let $R_n(g) = P_n R(g) P_n$ denote a sequence of finite rank approximations to $R(g)$ as above. Choose λ_n so that $g_n = \lambda_n \theta(\exp \frac{1}{2}R_n(g)) \in \mathrm{Spin}_0(W)$ converges in $\mathrm{Spin}_Q(W)$ to g. Suppose R is a rank K element of $R_Q(W)$, and let $R' = R + R(g)$. Suppose $I - R'Q_-$ is invertible. Let $R'_n = R + R_n(g)$. Then since R'_n converges to R' in $R_Q(W)$, and hence in uniform norm, it follows that $I - R'_n Q_-$ is invertible for all sufficiently large n. By Definition 3.2 the group element $\theta(\exp(\frac{1}{2}R_n)g_n) \in \hat{\mathrm{Spin}}_Q(W)$ has induced rotation $(I - R'_n Q_-)^{-1}(I + R'_n Q_+)$ and vaccum expectation $\langle g_n \rangle_Q = \lambda_n$. Since R'_n is finite rank, Theorem 3.1 applies and this same element of $G_0(W)$ may be written $\lambda_n \theta(\exp(\frac{1}{2}R'_n))$. Thus

$$\theta(\exp(\tfrac{1}{2}R)g_n) = \lambda_n \theta(\exp(\tfrac{1}{2}R'_n)). \tag{3.6}$$

But

$$\theta(\exp(\tfrac{1}{2}R'_n)) = \theta(\exp(\tfrac{1}{2})(R + R_n(g))) = \theta(\exp(\tfrac{1}{2}R)\exp(\tfrac{1}{2}R_n(g)))$$

since R and $R_n(g)$, regarded as elements of $\Lambda_0^2(W)$, commute with one another. Since R has rank K we choose a basis $\{w_j\}$ of W so that

$$\sum_{j=1}^{\infty} Rw_j \wedge w_j^* = \sum_{j=1}^{K} Rw_j \wedge w_j^*.$$

Thus

$$\exp\left(\frac{1}{2}R\right) = \sum_{k=0}^{\infty} (2^k k!)^{-1} R^k = \sum_{k=0}^{K} (2^k k!)^{-1} R^k.$$

Thus

$$\theta\left(\exp\left(\frac{1}{2}R\right)g_n\right) = \lambda_n \sum_{k=0}^{K} (2^k k!)^{-1} \theta\left(R^k \exp\left(\frac{1}{2}R_n(g)\right)\right). \tag{3.7}$$

Suppose now that $v_j \in W$, $j = 1,\ldots,m$, and consider

$$\theta(v_m \wedge \cdots \wedge v_1 \wedge \exp(\tfrac{1}{2}R_n(g))) = v_m^+ \theta(v_{m-1} \wedge \cdots \wedge v_1 \wedge \exp(\tfrac{1}{2}R_n(g))) + (-1)^{m-1}\theta(v_{m-1} \wedge \cdots \wedge v_1 \wedge \exp(\tfrac{1}{2}R_n(g)))v_m^-, \tag{3.8}$$

where $v_m^{\pm} = Q_{\pm} v_m$, and we used (3.5) and the fact that $\exp(\frac{1}{2}R_n(g))$ is an even element of the Grassmann algebra. This reduction formula is the same as that which defines the normal ordered products (3.1). Thus since $\lambda_n F\theta(\exp(\frac{1}{2}R_n(g))) = \Gamma_Q(g_n)$ it follows that

$$\lambda_n F\theta(v_m \wedge \cdots \wedge v_1 \wedge \exp(\tfrac{1}{2}R_n(g))) = N(v_m \cdots v_1 g_n).$$

Hence $\lambda_n F\theta(R^k \exp(\frac{1}{2}R_n(g))) = N(R^k g_n)$ and we have

$$F\theta(\exp(\tfrac{1}{2}R)g_n) = \sum_{k=0}^{K} (2^k k!)^{-1} N(R^k g_n). \tag{3.9}$$

To obtain the desired result we want to pass to the limit $n \to \infty$ in (3.9). To do this we regard each side of (3.9) as an element of $L(\mathscr{D})$. Since g_n converges to g in $\mathrm{Spin}_Q(W)$ it follows that $\Gamma_Q(g_n)$ converges to $\Gamma_Q(g)$ strongly on the dense domain D (Theorem 3.10 in Ref. 2). It is straightforward to use this fact and the reduction formula (3.1) to show that $N(R^k g_n)$ converges strongly on $\mathscr{D}$ to $N(R^k g)$ as $n \to \infty$. The element $\theta(\exp(\frac{1}{2}R)g_n) \in G_0(W)$ is not necessarily in $\mathrm{Spin}_0(W)$ but since $T\theta(\exp(\frac{1}{2}R)g_n)$ converges in $\mathrm{SO}_Q(W)$ as $n \to \infty$ and its vacuum expectation $\langle g_n \rangle_Q$ converges to a nonzero limit it follows from Proposition 3.5 in Ref. 2 that $\theta(\exp(\frac{1}{2}R)g_n)$ differs from a convergent sequence in $\mathrm{SO}_Q(W)$ by a scalar that also converges. It follows from Theorem 3.10 in Ref. 2 that $F\theta(\exp(\frac{1}{2}R)g_n)$ converges strongly on $\mathscr{D}$ to $\hat{\Gamma}\theta(\exp(\frac{1}{2}R)g)$ as $n \to \infty$. Thus

$$\hat{\Gamma}\theta\left(\exp\left(\frac{1}{2}R\right)g\right) = \sum_{k=0}^{K} (2^k k!)^{-1} N(R^k g).$$

We wish to extend this from $g \in \mathrm{Spin}_Q(W)$ to $g' \in \hat{\mathrm{Spin}}_Q(W)$. Suppose then that $g' = g \times a(\ker) \in \mathrm{Spin}_Q(W) \times_\alpha \mathrm{GL}(W_+)/\ker = \hat{\mathrm{Spin}}_Q(W)$ and that $\langle g' \rangle_Q = \langle g \rangle_Q \neq 0$. Let

$$T(g) = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

so that

$$T'(g) = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & a^{-\tau} \end{bmatrix}.$$

One easily calculates

$$R(g) = \begin{bmatrix} D^{-\tau} - 1 & BD^{-1} \\ D^{-1}C & 1 - D^{-1} \end{bmatrix}$$

and

$$R(g') = \begin{bmatrix} D^{-\tau}a - 1 & BD^{-1} \\ a^{\tau}D^{-1}Ca & 1 - a^{\tau}D^{-1} \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}^{\tau} R(g) \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} a - 1 & 0 \\ 0 & 1 - a^{\tau} \end{bmatrix}. \tag{3.10}$$

The element $\theta(\exp(\frac{1}{2}R)g')$ is the unique element of $\hat{\mathrm{Spin}}_Q(W)$ with R matrix $R + R(g')$ and vacuum expectation $\langle g' \rangle_Q = \langle g \rangle_Q$. Let e denote the identity in $\mathrm{Spin}_Q(W)$ and write $e \times a$ for the element $e \times a(\ker) \in \hat{\mathrm{Spin}}_Q(W)$. Let

$$R_a = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}^{-\tau} R \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}^{-1}.$$

Note that $R_a \in R_Q(W)$ and R_a is finite rank. Consider the element $\theta(\exp(\frac{1}{2}R_a)g) \cdot (e \times a) \in \hat{\mathrm{Spin}}_Q(W)$. A simple calculation using (3.10) shows that the R matrix of this element is

$$\begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix}^{\tau} (R_a + R(g)) \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} a - 1 & 0 \\ 0 & 1 - a^{\tau} \end{bmatrix} = R + R(g')$$

and since $\Gamma(e \times a)1 = 1$ the vacuum expectation is $\langle \theta(\exp(\frac{1}{2}R_a)g) \rangle_Q = \langle g \rangle_Q$. Thus

$$\theta(\exp(\tfrac{1}{2}R)g') = \theta(\exp(\tfrac{1}{2}R_a)g)\cdot(e\times a). \tag{3.11}$$

If we apply the homomorphism $\hat{\Gamma}$ to both sides of (3.11) then

$$\hat{\Gamma}\theta(\exp(\tfrac{1}{2}R)g') = \hat{\Gamma}\theta(\exp(\tfrac{1}{2}R_a)g)\hat{\Gamma}(e\times a).$$

But we know that

$$\hat{\Gamma}\theta\left(\exp\left(\frac{1}{2}R_a\right)g\right) = \sum_{k=0}^{K}(2^k k!)^{-1}N(R_a^k g).$$

Thus we will have the result we desire if we can show that

$$N(R_a^k g)\hat{\Gamma}(e\times a) = N(R^k g\times a) = N(R^k g').$$

To see this, first consider what happens when $\hat{\Gamma}(e\times a)$ is moved inside a normal ordered product $N(v_n\cdots v_1 g)$. Using the reduction formula (3.1) one finds

$$N(v_n\cdots v_1 g)\hat{\Gamma}(e\times a) = N(v_n'\cdots v_1' g\times a), \tag{3.12}$$

where

$$v_j' = \begin{bmatrix}1 & 0\\ 0 & a^\tau\end{bmatrix} v_j$$

(remember $T(e\times a) = \left[\begin{smallmatrix} a & 0\\ 0 & a^{-\tau}\end{smallmatrix}\right]$). Recall that, thought of as an element of $\Lambda_0^2(W)$, we have

$$R = \sum_{j=1}^{K} Rw_j\wedge w_j^*.$$

The basis

$$w_j(a) \overset{\text{def}}{=} \begin{bmatrix}a & 0\\ 0 & 1\end{bmatrix} w_j$$

is then the right sort of basis to use in calculating $R_a\in\Lambda_0^2(W)$. The basis

$$w_j(a)^* = \begin{bmatrix}a & 0\\ 0 & 1\end{bmatrix}^{-\tau} w_j^* = \begin{bmatrix}1 & 0\\ 0 & a^{-\tau}\end{bmatrix} w_j^*$$

is dual to $w_j(a)$. Thus

$$R_a = \sum_{j=1}^{K} R_a w_j(a)\wedge w_j(a)^*$$

$$= \sum_{j=1}^{K}\begin{bmatrix}1 & 0\\ 0 & a^{-\tau}\end{bmatrix} Rw_j\wedge\begin{bmatrix}1 & 0\\ 0 & a^{-\tau}\end{bmatrix} w_j^*.$$

Comparing this with (3.12) one sees that $N(R_a^k g)\hat{\Gamma}(e\times a) = N(R^k(g\times a))$ and we have finished the proof of Theorem 3.3. Q.E.D.

There is a slightly more symmetrical rephrasing of Theorem 3.3 that is useful.

Corollary 3.4: Suppose $g_j\in\hat{\mathrm{Spin}}_Q(W)$ and $\langle g_j\rangle_Q\neq 0$, for $j=1,2$. Suppose $T(g_2)-T(g_1)$ is finite rank and define $\Delta R = R(g_2)-R(g_1)$. Then ΔR is finite rank and one has

$$\hat{\Gamma}(g_2) = \frac{\langle g_2\rangle_Q}{\langle g_1\rangle_Q}\sum_{k=0}^{K}(2^k k!)^{-1}N((\Delta R)^k g_1),$$

where $K = \operatorname{rank}\Delta R$.

Proof: It is straightforward to show that $T(g_2)-T(g_1)$ is finite rank if and only if $R(g_2)-R(g_1)$ is finite rank [given that $D(g_j), j=1,2$, is invertible]. The rest is just a translation of Theorem 3.3.

IV. A GENERALIZED WICK THEOREM

In this section we prove a generalization of Wick's theorem. The proof is based on an idea in Sato, Miwa, and Jimbo[1] and requires the introduction of generalized Q functionals on $\mathscr{C}_0(W)$. To avoid technical complications we will introduce these Q functionals only in the case where W is finite dimensional. We will prove Wick's theorem directly in the finite-dimensional case and then extend the result to $\hat{\mathrm{Pin}}_Q(W)$ in much the same way that Theorem 3.3 was proved.

To begin, suppose that W is a finite-dimensional complex vector space with nondegenerate bilinear form $(\cdot,\cdot)$. Let Q denote an involution ($Q^2 = I$) on W that is skew symmetric ($Q^\tau = -Q$). In Ref. 2 it is shown that each such skew symmetric involution gives rise to a Fock representation F_Q of the Clifford algebra on $A(W_+)$ [where $W_\pm = Q_\pm W$ and $Q_\pm = (I\pm Q)/2$]. We have, for $x\in W$,

$$F_Q(x) = c(x_+) + c^\tau(x_-),$$

$$x = x_+ + x_-,\quad x_\pm\in W_\pm,$$

where $c(x_\pm) = x_\pm\wedge\cdot$ are creation operators on $A(W_\pm)$ and $c^\tau(x_-)$ is the transpose of $c(x_-)$, with $A(W_+)$ identified as the dual of $A(W_-)$ via the bilinear form $(\cdot,\cdot)$. The anticommutator

$$c(x)c^\tau(y) + c^\tau(y)c(x) = (x,y),\quad x\in W_+,\quad y\in W_-. \tag{4.1}$$

More details can be found in Ref. 2, Sec. 2.

The vacuum vectors $1\in A(W_+)$ and $1^*\in A(W_-)\simeq A(W_+)^*$ give rise to a linear functional $\langle\cdot\rangle_Q$ on $\mathscr{C}(W)$ defined by

$$\langle X\rangle_Q = (F_Q(X)1,1^*),\quad X\in\mathscr{C}(W),$$

where $(\cdot,\cdot)$ is the dual pairing between $A(W_+)$ and $A(W_-)$. We will refer to the linear functional $\langle\cdot\rangle_Q$ as the Q-Fock state on $\mathscr{C}(W)$. We will now extend this notion by removing the condition that Q is an involution. Let $W\oplus W$ denote the vector space direct sum of W with itself with the nondegenerate bilinear form

$$(x_1\oplus y_1, x_2\oplus y_2) = (x_1,x_2) - (y_1,y_2)$$

(note the minus sign here!).Let $W\oplus W^*$ denote the vector space direct sum of W with itself with the nondegenerate pairing

$$(x_1\oplus y_1, x_2\oplus y_2) = (x_1,y_2) + (y_1,x_2).$$

Evidently this pairing identifies the second component with the dual of the first and this is the reason for the notation $W\oplus W^*$. Now let Q be any skew symmetric map on W and write $Q_\pm = (I\pm Q)/2$ (note that the maps $Q_\pm$ are not, in general, projections). The matrix $\left[\begin{smallmatrix} I_{Q_-} & -I_{Q_+}\end{smallmatrix}\right]$ defines an orthogonal map from $W\oplus W$ to $W\oplus W^*$ and hence extends to an algebra homomorphism from $\mathscr{C}(W\oplus W)$ to $\mathscr{C}(W\oplus W^*)$ which we denote by $o(Q)$. The natural inclusion $W\ni w\to w\oplus 0\in W\oplus W$ extends to an algebra homomorphism from $\mathscr{C}(W)$ into $\mathscr{C}(W\oplus W)$. We will use this homomorphism to identify $\mathscr{C}(W)$ with a subalgebra of $\mathscr{C}(W\oplus W)$ without introducing special notation. The map $Q_0 = \left[\begin{smallmatrix} I & 0\\ 0 & -I\end{smallmatrix}\right]$ on $W\oplus W^*$ is a skew-symmetric involution. Let $\langle\cdot\rangle_0$ denote the Q_0-Fock state on $\mathscr{C}(W\oplus W^*)$. We define the Q functional $\langle\cdot\rangle_Q$ on $\mathscr{C}(W)$ by

$$\langle X\rangle_Q \overset{\text{def}}{=} \langle o(Q)X\rangle_0,\quad X\in\mathscr{C}(W)\subseteq\mathscr{C}(W\oplus W). \tag{4.2}$$

For this notation to be consistent we must check that when Q is an involution this reproduces the Fock state on $\mathscr{C}(W)$. It is enough to check this for $X = w_1 \cdots w_n$ a monomial. It is well known (and easy to prove) that

$$\langle w_1 \cdots w_n \rangle_Q = \mathrm{Pf}\, A, \tag{4.3}$$

where A is the skew symmetric matrix with ij entry $(i<j)$ given by $(Q_- w_i, w_j)$. The Pfaffian of A, $\mathrm{Pf}\, A$, is defined inductively for an $n \times n$ skew symmetric matrix $A = (\alpha_{ij})$ as follows:

$$\mathrm{Pf}[0] = 0, \quad \mathrm{Pf}\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix} = a,$$

$$\mathrm{Pf}\, A = \sum_{k=2}^{n} (-1)^{k+1} a_{1k}\, \mathrm{Pf}(A)_{1\times k},$$

where $(A)_{1\times k}$ is the $(n-2)\times(n-2)$ skew symmetric matrix obtained from A by deleting the columns 1 and k and the rows 1 and k. It is known that $[\mathrm{Pf}(A)]^2 = \det A$.

If we apply (4.3) to $\langle \mathrm{o}(Q) w_1 \cdots w_n \rangle_0$ then we find that this vacuum expectation is the Pfaffian of the skew symmetric matrix B with ij entry

$$\begin{aligned}
&\left(Q_0^- \mathrm{o}(Q)\begin{bmatrix} w_i \\ 0 \end{bmatrix}, \mathrm{o}(Q)\begin{bmatrix} w_j \\ 0 \end{bmatrix}\right)_{W\oplus W^*} \\
&\quad = \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} w_i \\ Q_- w_i \end{bmatrix}, \begin{bmatrix} w_j \\ Q_- w_j \end{bmatrix}\right)_{W\oplus W^*} \\
&\quad = (Q_- w_i, w_j) \quad (i<j).
\end{aligned}$$

Thus $B = A$ and we have $\langle \mathrm{o}(Q) w_1 \cdots w_n \rangle_0 = \langle w_1 \cdots w_n \rangle_Q$, where Q is a skew symmetric involution and on the right $\langle \cdot \rangle_Q$ is the Q-Fock state on $\mathscr{C}(W)$. It is convenient for a later application to observe that the calculation we just did shows that (4.3) remains valid for general Q functionals with the skew symmetric matrix A that has entries

$$a_{ij} = (Q_- w_i, w_j) = \langle w_i w_j \rangle_Q, \quad \text{for } i<j.$$

There is a Grassmann calculus for $\mathscr{C}(W)$ associated with these generalized Q functionals.[1a] In general, however, the subspace $W_\pm = Q_\pm W$ need not be isotropic (think of $Q=0$) and so the "normal ordering" prescription we gave for θ earlier does not apply. It is in fact simpler to give a formula for the inverse of θ (which is denoted Nr in Ref. 1a). Suppose $X \in \mathscr{C}(W)$ and Q is a skew-symmetric map on W. Define $\mathrm{Nr}_Q(X) \in A(W)$ by

$$\mathrm{Nr}_Q(X) = F_0(\mathrm{o}(Q)X)1_0, \tag{4.4}$$

where F_0 is the Q_0-Fock representation and $1_0 \in A(W)$ is the Q_0-Fock vacuum vector. It is not hard to show that Nr_Q is surjective. Since $\mathscr{C}(W)$ and $A(W)$ both have dimension $2^{\dim W}$ it follows that Nr_Q is bijective. Thus we may define $\theta_Q = (\mathrm{Nr}_Q)^{-1}$. Again, in order to see that this is consistent with previous usage we must check that it reproduces the normal ordering prescription for θ_Q when Q is an involution. We will demonstrate this inductively. Suppose Q is a skew symmetric involution. Let θ denote the normal ordering map (3.5) and Nr_Q the map (4.4). We wish to show that $\theta(\mathrm{Nr}_Q(w_1 \cdots w_n)) = w_1 \cdots w_n$, where $w_j \in W$. This is trivially true when $n=1$ and for the sake of an induction we suppose that it is true for products of length less than or equal to $(n-1)$. By (4.4) we have

$$\mathrm{Nr}_Q(w_1 \cdots w_n) = (c(w_1) + c^\tau(w_1^-))\mathrm{Nr}_Q(w_2 \cdots w_n),$$

where $w_1^- = Q_- w_1$. If we move the annihilation operator $c^\tau(w_1^-)$ past the factors $c(w_j) + c^\tau(w_j^-)$ in $\mathrm{Nr}_Q(w_2 \cdots w_n)$ until it hits the vacuum $c^\tau(w_1^-)1_0 = 0$ and make use of the anticommutators (4.1) then we find

$$\begin{aligned}
&c^\tau(w_1^-)\mathrm{Nr}_Q(w_2 \cdots w_n) \\
&\quad = \sum_{k=2}^{n} (-1)^k (w_1^-, w_k)\mathrm{Nr}_Q(w_2 \cdots \hat{w}_k \cdots w_n),
\end{aligned}$$

where the $\hat{}$ over w_k means that w_k is absent from the product. It is clear from (4.4) that $\mathrm{Nr}_Q(w_2 \cdots w_n)$ is the sum of elements in the Grassmann algebra with the same parity as $(n-1)$. Thus

$$\begin{aligned}
c(w_1^-)\mathrm{Nr}_Q(w_2 \cdots w_n) &= w_1^- \wedge \mathrm{Nr}_Q(w_2 \cdots w_n) \\
&= (-1)^{n-1}\mathrm{Nr}_Q(w_2 \cdots w_n) \wedge w_1^-.
\end{aligned}$$

Thus we find

$$\begin{aligned}
&\mathrm{Nr}_Q(w_1 \cdots w_n) \\
&\quad = w_1^+ \wedge \mathrm{Nr}_Q(w_2 \cdots w_n) \\
&\qquad + (-1)^{n-1}\mathrm{Nr}_Q(w_2 \cdots w_n) \wedge w_1^- \\
&\qquad + \sum_{k=2}^{n} (-1)^k (w_1^-, w_k)\mathrm{Nr}_Q(w_2 \cdots \hat{w}_k \cdots w_n).
\end{aligned}$$

If we apply θ to both sides of this equation and make use of (3.5) and the inductive hypothesis then we find

$$\begin{aligned}
&\theta\, \mathrm{Nr}_Q(w_1 \cdots w_n) \\
&\quad = w_1^+ w_2 \cdots w_n + (-1)^{n-1} w_2 \cdots w_n w_1^- \\
&\qquad + \sum_{k=2}^{n} (-1)^k (w_1^-, w_k)(w_2 \cdots \hat{w}_k \cdots w_n).
\end{aligned}$$

Making use of the Clifford relations one finds

$$\begin{aligned}
&(-1)^{n-1} w_2 \cdots w_n w_1^- \\
&\quad = w_1^- w_2 \cdots w_n - \sum_{k=2}^{n} (-1)^k (w_1^-, w_k) \\
&\qquad \times (w_2 \cdots \hat{w}_k \cdots w_n).
\end{aligned}$$

Thus

$$\theta(\mathrm{Nr}_Q(w_1, \ldots, w_n)) = (w_1^+ + w_1^-) w_2 \cdots w_n = w_1 w_2 \cdots w_n.$$

This completes the inductive step and shows that $\theta = \mathrm{Nr}_Q^{-1}$.

Suppose Q is map on W with $Q^\tau = -Q$. Suppose $G \in \mathrm{O}(W)$ such that $Q_- G + Q_+$ is invertible. Define

$$R_Q(G) = (G - I)(Q_- G + Q_+)^{-1}.$$

Then $R_Q(G)^\tau = -R_Q(G)$ and we may identify $R_Q(G)$ with an element of $\Lambda^2(W)$ as before. The following result will be of use to us.

Theorem 4.0: Suppose $g \in G(W)$ and $\langle g \rangle_Q \neq 0$. Let $G = T(g)$. Then $Q_- G + Q_+$ is invertible and

$$g = \langle g \rangle_Q \theta_Q \exp(\tfrac{1}{2} R_Q(G)).$$

Conversely, suppose R: $W \to W$ is skew symmetric and $I - RQ_-$ is invertible. Then $g = \theta_Q(\exp \tfrac{1}{2} R_Q(G)) \in G(W)$ and

$$T(g) = (I - RQ_-)^{-1}(I + RQ_+) \in \mathrm{SO}(W).$$

Proof: This result is a consequence of Theorem 3.1 as the

reader can find worked out in the proof of Theorem 2.0 in Ref. 3. Q.E.D.

We next introduce a slight extension of (4.4) that will be used in the proof of the generalized Wick theorem.

Theorem 4.1: Suppose $Q^\tau = -Q$ and that $g \in G(W)$ with $\langle g \rangle_Q \neq 0$. Let $w_j \in W$ $(j = 1,...,n)$. Then

$$\langle w_1 \cdots w_n g \rangle_Q = \langle g \rangle_Q \, \mathrm{Pf}(H),$$

where H is the $n \times n$ skew symmetric matrix with ij entry $\langle w_i w_j g \rangle_Q / \langle g \rangle_Q$ $(i < j)$.

Proof: This is a special case of Theorem 2.1 in Ref. 3 so we only sketch the proof. It is convenient to first prove this result in the case that Q is an involution. Let $G = T(g)$. Then $\langle g \rangle_Q \neq 0$ implies $Q_- G + Q_+$ is invertible. Since $Q_- G + Q_+$ is invertible we may factor $G = UL$, where $U \in \mathrm{SO}(W)$ is upper triangular and $L \in \mathrm{SO}(W)$ is lower triangular with respect to the decomposition $W = W_+ \oplus W_-$ (with $W_\pm = Q_\pm W$). There exist elements $u, l \in G(W)$ such that $T(u) = U$ and $T(l) = L$ (Sec. 2 in Ref. 2) and we may normalize u and l so that $l 1_Q = \langle g \rangle_Q 1_Q$ and $u^\tau 1_Q^* = 1_Q^*$. It follows that $g = ul$ and hence that

$$\langle w_1 \cdots w_n g \rangle_Q = \langle g \rangle_Q \langle w_1 \cdots w_n u \rangle_Q = \langle g \rangle_Q \langle w_1' \cdots w_n' \rangle_Q,$$

where $w_j' = U^{-1} w_j$. Thus using (4.3) we find

$$\langle w_1 \cdots w_n g \rangle_Q = \langle g \rangle_Q \, \mathrm{Pf}(H),$$

where H is the skew-symmetric matrix with ij entry

$$\langle w_i' w_j' \rangle_Q = \langle w_i w_j u \rangle_Q = \langle w_i w_j g \rangle_Q / \langle g \rangle_Q.$$

This finishes the proof when Q is an involution. If Q is not an involution then $\langle X \rangle_Q = \langle \mathrm{o}(Q) X \rangle_0$ expresses the Q functional in terms of the Q_0 functional. But the map Q_0 is an involution and it is straightforward that the result for Q_0 implies the result for Q. Q.E.D.

Before we state the principal results of this section we introduce some notation. Let M be a positive integer. For $j = 1,...,M$, let l_j denote a non-negative integer. For j such that $l_j > 0$ let u_{jk} denote an element of W for $1 \leqslant k \leqslant l_j$. Let Q denote a skew symmetric map on W and for $j = 1,...,M$ let $g_j \in G(W)$ with $\langle g_j \rangle_Q \neq 0$. For $j = 1,...,M$ define $h_j \in \mathscr{C}(W)$ by

$$\mathrm{Nr}_Q(h_j) = \begin{cases} u_{j1} \wedge u_{j2} \cdots \wedge u_{jl_j} \wedge \mathrm{Nr}_Q(g_j), & \text{if } l_j > 0 \\ \mathrm{Nr}_Q(g_j), & \text{if } l_j = 0. \end{cases}$$

Let $v_i = u_{jk}$ where $i = l_1 + l_2 + \cdots + l_{j-1} + k$ and $l_0 = 0$. The v_i are the vectors u_{jk} recorded as they appear in the product $h_1 h_2 \cdots h_M$. We will say that v_i belongs to the jth string of the set of vectors $\{u_{\alpha\beta}\}$ if $v_i = u_{jk}$ for some k. We write $s(m) = l_1 + l_2 + \cdots + l_m$. It might help the reader to recognize the familiar form of Wick's theorem as the special case of the following result obtained by setting all the group elements g_j equal to the identity I.

Theorem 4.2: (Generalized Wick Theorem): Suppose in addition to the conditions listed above that $\langle g_1 \cdots g_M \rangle_Q \neq 0$. Then

$$\langle h_1 \cdots h_M \rangle_Q = \langle g_1 \cdots g_M \rangle_Q \, \mathrm{Pf}(H),$$

where H is the $s(M) \times s(M)$ skew symmetric matrix with ij entry $(i < j)$:

$$H_{ij} = \langle g_1' \cdots g_M' \rangle / \langle g_1 \cdots g_M \rangle,$$

where $\mathrm{Nr}(g_k') = v_i \wedge v_j \wedge \mathrm{Nr}(g_k)$ if both v_i and v_j belong to the k th string in $\{u_{\alpha\beta}\}$, $\mathrm{Nr}(g_k') = v_i \wedge \mathrm{Nr}(g_k)$ if v_i is in the k th string but v_j is not, $\mathrm{Nr}(g_k') = v_j \wedge \mathrm{Nr}(g_k)$ if v_j is in the k th string but v_i is not, and finally, $\mathrm{Nr}(g_k') = \mathrm{Nr}(g_k)$ if neither v_i nor v_j is in the k th string of $\{u_{\alpha\beta}\}$.

Proof: For the proof we will construct a representation for $\langle h_1 \cdots h_M \rangle_Q$ that will permit a direct application of Theorem 4.1. For each j with $l_j > 0$ let W_j denote the vector space $\mathbf{C}^{l_j}$ with the standard basis $e_{j1} = (1,0,...,0)$, $e_{j2} = (0,1,0,...,0)$, and $e_{jl_j} = (0,0,...,0,1)$. Let

$$\hat{W} = W \oplus \sum_{j:\, l_j > 0} \oplus W_j.$$

The space $\hat{W}$ is a complex vector space with a distinguished nondegenerate bilinear form given by the direct sum of the bilinear form on W with the standard bilinear forms on $W_j \simeq \mathbf{C}^{l_j}$.

Let $\hat{Q} = Q \oplus 0$ denote the skew symmetric map on $\hat{W}$ equal to Q on W and 0 on each W_j. Let ϵ denote a positive real number, with $\mathrm{Nr} = \mathrm{Nr}_Q$ and $\mathrm{Nr} = \mathrm{Nr}_{\hat{Q}}$ and define $\hat{g}_j \in \mathscr{C}(\hat{W})$ by

$$\mathrm{Nr}(\hat{g}_j) = \begin{cases} \exp\left(\epsilon \sum_{k=1}^{l_j} e_{jk} \wedge u_{jk}\right) \wedge \mathrm{Nr}(g_j), & \text{if } l_j > 0 \\ \mathrm{Nr}(g_j), & \text{if } l_j = 0, \end{cases}$$

where here and in what follows we regard W and W_j as subspaces of $\hat{W}$. Let $f_i = 2\epsilon^{-1} e_{jk}$ with $i = l_1 + \cdots + l_{j-1} + k$ for $i = 1,...,s(M)$. The representation we are interested in for $\langle h_1 \cdots h_M \rangle_Q$ is

$$\langle h_1 \cdots h_M \rangle_Q = \langle f_1 \cdots f_{s(M)} \hat{g}_1 \cdots \hat{g}_M \rangle_{\hat{Q}}. \tag{4.5}$$

In order to prove this we first establish a formula for $\hat{g}_j$. Write $\theta = \theta_Q$ and $\hat{\theta} = \theta_{\hat{Q}}$. Then for $(l_i > 0)$,

$$\hat{g}_j = \sum_{k=0}^{\infty} \epsilon^k \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant l_j} e_{ji_k} \cdots e_{ji_1} \times \theta(u_{ji_1} \wedge \cdots \wedge u_{ji_k} \wedge \mathrm{Nr}(g_j)). \tag{4.6}$$

Setting $e_i = e_{ji}$, $u_i = u_{ji}$ and $l = l_j$ to unburden the notation we have

$$\exp\left(\epsilon \sum_{i=1}^{l} e_i \wedge u_i\right) = \sum_{k=0}^{\infty} \epsilon^k \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant l} e_{i_k} \wedge \cdots \wedge e_{i_1} \wedge u_{i_1} \wedge \cdots \wedge u_{i_k}.$$

Thus to prove (4.6) it suffices to show that for $i_1 < i_2 \cdots < i_k$ we have

$$\hat{\theta}(e_{i_k} \wedge \cdots \wedge e_{i_1} \wedge u_{i_1} \wedge \cdots u_{i_k} \wedge \mathrm{Nr}(g_j)) = e_{i_k} \cdots e_{i_1} \theta(u_{i_1} \wedge \cdots \wedge u_{i_k} \wedge \mathrm{Nr}(g_j)). \tag{4.7}$$

To calculate $\hat{\theta}$ one may recast the definition of $\hat{\theta}$ as follows:

$$\hat{\theta}(w \wedge X) = w \cdot \hat{\theta}(X) + \hat{\theta}(w \wedge X - \mathrm{Nr}(w \cdot \hat{\theta}(X))),$$

where $w \in \hat{W}$ and $x \in A(\hat{W})$. Thus since

$$\mathrm{Nr}(w \cdot \hat{\theta}(X)) = [c(w) + c^\tau(w_-)]X = w \wedge X + c^\tau(w_-)X$$

it follows that

$$\hat{\theta}(w \wedge X) = w \cdot \hat{\theta}(X) - \hat{\theta}(c_\tau(w_-)X), \tag{4.8}$$

where $w_- = \hat{Q}_- w$. When X is a monomial $w_1 \wedge \cdots \wedge w_m = c(w_1) \cdots c(w_m) 1$ of order m it is clear that $c^\tau(w_-)X$ is a

sum of monomials of order $m-2$ [see (4.1)]. Thus (4.5) gives an inductive means for calculating $\hat{\theta}$. If we apply (4.8) to the left-hand side of (4.7) with $w=e_{i_1}$ then the term $-\hat{\theta}(c^\tau(w_-)X)$ vanishes since $w_-=\hat{Q}_-e_{i_1}=\frac{1}{2}e_{i_1}$ and $c^\tau(e_{i_1})$ anticommutes with $c(e_i)$ $(i\neq i_1)$ and with $c(u)$ $(u\in W)$. Repeated application of this observation establishes (4.7) with θ on the right-hand side replaced by $\hat{\theta}$. However, $\hat{\theta}=\theta$ on $A(W)\subseteq A(\hat{W})$ and this finishes the proof of (4.6).

In order to prove (4.5) observe that if we let $e_i=\epsilon e_{jk}$ with $i=l_1+\cdots+l_{j-1}+k$ then for $i_1<\cdots i_k$ and $j_1<\cdots<j_l$ we have

$$\langle f_{i_1}\cdots f_{i_k}e_{j_l}\cdots e_{j_1}\rangle_{\hat{Q}}=\delta_{lk}\delta_{i_1j_1}\cdots\delta_{i_kj_k}. \tag{4.9}$$

It follows that the only term in (4.6) that survives after being substituted into (4.5) is

$$\epsilon^{l_j}e_{jl_j}\cdots e_{j_1}\theta(u_{j_1}\wedge\cdots\wedge u_{jl_j}\wedge \mathrm{Nr}(g_j)).$$

This and (4.9) prove (4.5) up to a possible sign difference. The sign change required to disentangle the factors $e_{jl_j}\cdots e_{j_1}$ from the factors $h_k=\theta(u_{k_1}\wedge\cdots\wedge u_{kl_k}\wedge \mathrm{Nr}(g_k))$ is

$$l_1l_2+(l_1+l_2)l_3+\cdots+(l_1+\cdots+l_{M-1})l_M=\sum_{i<j}l_il_j.$$

However, the groups of factors $e_{jl_j}\wedge\cdots\wedge e_{j_1}$ appears from left to right in increasing order with j. To put them in decreasing order requires

$$(l_1+\cdots+l_{M-1})l_M+(l_1+\cdots+l_{M-2})l_{M-1}$$
$$+\cdots+l_1l_2=\sum_{i<j}l_il_j$$

sign interchanges. These sign changes cancel out and (4.5) is proved.

Now we wish to show that Theorem 4.1 applies to the evaluation of the right-hand side of (4.5). We need to know that $\hat{g}_1\cdots\hat{g}_M\in G(\hat{W})$ and that $\langle\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}\neq 0$. We will show that both these conditions are satisfied for ϵ sufficiently small. To show that $\hat{g}_1\cdots\hat{g}_M\in G(\hat{W})$ it suffices to prove that $\hat{g}_j\in G(\hat{W})$, for $j=1,...,M$. But $\mathrm{Nr}(\hat{g}_j)$ is (a multiple of) the exponential of a quadratic element in $A(\hat{W})$. Theorem (4.0) implies that $\hat{g}_j\in G(\hat{W})$ provided the associated skew symmetric map $R(\hat{g}_j)$: $\hat{W}\to\hat{W}$ determines an invertible map $I-R(\hat{g}_j)\hat{Q}_-$. However $R(\hat{g}_j)$ differs by order ϵ from a map $R(g_j)$ for which $I-R(g_j)\hat{Q}_-$ is invertible. Thus by choosing ϵ small enough we may insure that $\hat{g}_j\in G(\hat{W})$. The vacuum expectation $\langle\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}$ is polynomial in ϵ. It is clear that

$$\lim_{\epsilon\to 0^+}\langle\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}=\langle g_1\cdots g_M\rangle_{\hat{Q}}=\langle g_1\cdots g_M\rangle_Q\neq 0.$$

Thus by choosing ϵ small enough we have $\langle\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}\neq 0$ and it follows Theorem 4.1 applies to the evaluation of the right-hand side of (4.5),

$$\langle f_1\cdots f_{s(M)}\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}=\langle\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}\,\mathrm{Pf}(H), \tag{4.10}$$

where H is the $s(M)\times s(M)$ skew symmetric matrix with ij entry $(i<j)$:

$$H_{ij}=\langle f_if_j\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}/\langle\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}.$$

But (4.5) applies to the numerator of H_{ij}. Thus

$$\langle f_if_j\hat{g}_1\cdots\hat{g}_M\rangle_{\hat{Q}}=\langle g_1'\cdots g_M'\rangle_Q. \tag{4.11}$$

Because of (4.5) the left-hand side of (4.10) does not depend on ϵ. Thus we may put $\epsilon=0$ on the right-hand side and since $\langle\hat{g}_1\cdots\hat{g}_m\rangle_{\hat{Q}}|_{\epsilon=0}=\langle g_1\cdots g_M\rangle_Q$ the theorem follows from (4.10) and (4.11). Q.E.D.

Theorem 4.2 has an immediate generalization to infinite dimensions. We turn now to this result. In the rest of this section W will denote a complex Hilbert space with a distinguished conjugation P. We will suppose that Q: $W\to W$ is an involution with $Q^*=Q$ and $Q^\tau=-Q$ as in Sec. I. We reiterate the notational conventions of Theorem 4.2.

Let M be a positive integer. For $j=1,...,M$ let l_j denote a non-negative integer. For j such that $l_j>0$ and k such that $1\leqslant k\leqslant l_j$ let u_{jk} denote an element of W. For $j=1,...,M$ let $g_j\in\hat{\mathrm{Spin}}_Q(W)$ with $\langle g_j\rangle_Q\neq 0$. For $j=1,...,M$ define $h_j\in L(\mathscr{D})$ by

$$h_j=\begin{cases}N(u_{j_1}\cdots u_{jl_j}g_j), & \text{if } l_j>0\\ \hat{\Gamma}(g_j)=N(g_j), & \text{if } l_j=0,\end{cases}$$

where $N=N_Q$ and $\hat{\Gamma}=\Gamma_Q$.

Let $v_i=u_{jk}$, where $i=l_1+\cdots+l_{j-1}+k$, and as above we will say that v_i is in the jth string of $\{u_{\alpha\beta}\}$ if $v_i=u_{jk}$ for some k. Recall that $s(m)=l_1+l_2+\cdots+l_m$, and to lighten the notation we write $\langle h_1\cdots h_M\rangle_Q$ for $\langle h_1\cdots h_M 1_Q,l_Q\rangle$.

Theorem 4.3: Suppose in addition to the conditions listed above that $\langle g_1\cdots g_M\rangle_Q\neq 0$. Then

$$\langle h_1\cdots h_M\rangle_Q=\langle g_1\cdots g_M\rangle_Q\,\mathrm{Pf}(H),$$

where H is the $s(M)\times s(M)$ skew symmetric matrix with ij entry $(i<j)$ given by

$$H_{ij}=\langle g_1'\cdots g_M'\rangle_Q/\langle g_1\cdots g_M\rangle_Q,$$

where $g_k'=N(v_iv_jg_k)$ if both v_i and v_j are in the k th string, $g_k'=N(v_ig_k)$ if v_i is in the k th string but v_j is not, $g_k'=N(v_jg_k)$ if v_j is in the k th string but v_i is not, $g_k'=N(g_k)$ if neither v_i nor v_j is in the k th string.

Proof: We first prove this result assuming $g_j\in\mathrm{Spin}_Q(W)$. As in the proof of Theorem 3.3 we can approximate $g_j\in\mathrm{Spin}_Q(W)$ by a sequence $g_{jn}\in\mathrm{Spin}_Q(W)$ converging to g_j in $\mathrm{Spin}_Q(W)$ as $n\to\infty$. Theorem 3.10 of Ref. 2 implies that $\Gamma_Q(g_{jn})$ converges strongly on the dense domain $\mathscr{D}$ as $n\to\infty$ to $\Gamma_Q(g_j)$. Let h_{jn} denote the approximation to h_j obtained by replacing g_j by g_{jn}. Then h_{jn} converges strongly to h_j on $\mathscr{D}$ as $n\to\infty$. Now approximate $\langle h_1\cdots h_M\rangle_Q$ by $\langle h_{1n_1}\cdots h_{Mn_M}\rangle_Q$. Each $h_{jn_j}\in F_Q(\mathscr{C}_0(W))$ so that Theorem 4.2 applies provided that $\langle g_{1n_1}\cdots g_{Mn_M}\rangle_Q\neq 0$. Strong convergence on $\mathscr{D}$ guarantees that we can make $\langle g_{1n_1}\cdots g_{Mn_M}\rangle_Q\neq 0$ by choosing $n_1,...,n_M$ sufficiently large since $\langle g_1\cdots g_M\rangle_Q\neq 0$. Evaluate $\langle h_{1n_1}\cdots h_{Mn_M}\rangle_Q$ using Theorem 4.2 and then in reverse order pass to the limits $n_M\to\infty$, $n_{M-1}\to\infty$, and finally $n_1\to\infty$, making use of strong convergence on $\mathscr{D}$ to evaluate the limits on both sides of the resulting equation. This establishes the desired result for $g_j\in\mathrm{Spin}_Q(W)$ $(j=1,...,M)$.

Suppose now that $g_j=g_j'\times a_j$ when $g_j'\in\mathrm{Spin}_Q(W)$ and $a_j\in\mathrm{GL}(W_+)$ and we write $g_j'\times a_j$ for the element $g_j'\times a_j(\ker)$ of $\hat{\mathrm{Spin}}_Q(W)$. Use (3.12) to pull $\hat{\Gamma}(e\times a_j)$ out of the normal ordered products defining h_j. Then move each $\hat{\Gamma}(e\times a_j)$ to the right until it hits the vacuum using

$$\hat{\Gamma}(e\times a_j)F_Q(w)\hat{\Gamma}(e\times a_j)^{-1} = F_Q(a_j \oplus a_j^{-\tau}w),$$
$$\hat{\Gamma}(e\times a_j)\Gamma_Q(g')\hat{\Gamma}(e\times a_j)^{-1} = \Gamma_Q(\alpha(a_j)g'),$$

and

$$\hat{\Gamma}(e\times a_j)1 = 1.$$

What results is a vacuum expectation of a product of normal ordered forms with elements $\tilde{g}_j$ from $\mathrm{Spin}_Q(W)$. Apply the result just established for $\mathrm{Spin}_Q(W)$ to evaluate this in terms of the Pfaffian of a matrix H. In the formulas for the matrix elements of the matrix H put the operators $\hat{\Gamma}(e\times a_j)$ back inside the normal ordered products by reversing the procedure above. In this fashion one finds the theorem is true for $g_j \in \hat{\mathrm{Spin}}_Q(W)$. Q.E.D.

V. DIFFERENCE IDENTITIES FOR THE ISING FIELD

In this section we derive local difference identities for the Ising field using Corollary 3.4. Such identities provide the foundation for the SMJ analysis of the scaled correlations in the version of this analysis presented in Ref. 4. Here we will show that the quadratic identities for the Ising correlations discovered by McCoy and Wu[7] and Perk[9] are simple consequences of these identities coupled with the generalized version of Wick's theorem (Theorem 4.3).

We review the representation established in Ref. 5 for the two-dimensional Ising correlations. The ingredients of the description are a Hilbert space W; an isotropic splitting of $W = W_+ \oplus W_-$ associated with the induced rotation T of the transfer matrix; and the "spin field" $\sigma \in \hat{\mathrm{Spin}}_Q(W)$.

Let W denote the Hilbert space $l^2(\mathbf{Z}_{1/2},\mathbf{C}^2)$, where $\mathbf{Z}_{1/2} = \mathbf{Z} + \frac{1}{2}$ and $l^2(\mathbf{Z}_{1/2},\mathbf{C}^2)$ has the inner product

$$\langle f,g\rangle = \sum_{k\in\mathbf{Z}_{1/2}} f(k)\cdot\overline{g(k)},$$

where $x\cdot y = x_1y_1 + x_2y_2$. For $f\in l^2(\mathbf{Z}_{1/2},\mathbf{C}^2)$ let $\mathrm{Pf}(k) = \overline{f(k)}$ denote the conjugation on W that acts by complex conjugation on each of the components of f. The distinguished bilinear form on W associated with P is thus

$$(f,g) = \sum_{k\in\mathbf{Z}_{1/2}} f(k)\cdot g(k).$$

We often identify W with $L^2(S^1,\mathbf{C}^2)$ via the Fourier series

$$f(z) = \sum_{k\in\mathbf{Z}_{1/2}} z^k f(k), \quad z\in S^1.$$

The multivalued functions $z^k = e^{k\log z}$, where $\log z = i\theta$ and $0\leqslant\theta<2\pi$. The branch cut is located at $z=1$.

Define the induced rotation T for the transfer matrix as the matrix multiplication operator on $L^2(S^1,\mathbf{C}^2)$ given by

$$Tf(z) = T(z)f(z),$$

where

$$T(z) = s_2^{-1}\begin{bmatrix} c_1c_2 - s_1\cos\theta & \sin\theta - i(s_1c_2 - c_1\cos\theta) \\ \sin\theta + i(s_1c_2 - c_1\cos\theta) & c_1c_2 - s_1\cos\theta \end{bmatrix}, \tag{5.1}$$

and $z = e^{i\theta}$, $c_j = \cosh(2K_j)$, $s_j = \sinh(2K_j)$, $j = 1,2$. The constants K_j are the interaction strengths for the horizontal (K_1) and vertical (K_2) bonds in the two-dimensional Ising model. The map T on W is complex orthogonal on W and when the constants K_1 and K_2 are real T is self-adjoint with respect to the Hermitian inner product on W. We will suppose from now on that K_1 and K_2 are real and positive. Define $\gamma(z) > 0$ and $a(z)\in S^1$ by

$$\cosh\gamma(z) = s_2^{-1}(c_1c_2 - s_1\cos\theta),$$
$$a(z)\sinh\gamma(z) = -s_2^{-1}(\sin\theta - i(s_1c_2 - c_1\cos\theta)).$$

These functions are of interest to us since

$$T(z) = \exp-\gamma(z)\begin{bmatrix} 0 & a(z) \\ a(z)^{-1} & 0 \end{bmatrix}.$$

Let Q denote the matrix muliplication operator $\left[\begin{smallmatrix} 0 & a(z) \\ a(z)^{-1} & 0 \end{smallmatrix}\right]$ on $L^2(S^1,\mathbf{C}^2)$. Then $Q^2 = I$, $Q^* = Q$, and $Q^\tau = -Q$. Let $Q_\pm = (I \pm Q)/2$. The subspaces $W_\pm = Q_\pm W$ give an isotropic splitting of $W = W_+ \oplus W_-$. The subspace W_+ is the spectral subspace for T associated with the interval $[0,1]$.

$$sf(k) = -\epsilon(k)f(k), \quad k\in\mathbf{Z}_{1/2},$$

where

$$\epsilon(k) = \begin{cases} 1, & k>0, \\ -1, & k<0. \end{cases}$$

Let σ denote the element in $\hat{\mathrm{Spin}}_Q(W)$ with the induced rotation $T(\sigma) = s$ and such that $\sigma^2 = I$. Such an element exists if $s_1s_2 \neq 1$ ($s_1s_2 = 1$) is the critical point for the Ising field. The conditions $T(\sigma) = s$ and $\sigma^2 = I$ determine σ up to an ambiguity in sign. The condition $s_1s_2 > 1$ corresponds to the Ising model below the critical temperature. In this case we normalize σ so that $\langle\sigma\rangle_Q > 0$. Now suppose $a\in\mathbf{Z}^2$ and let $v(a) = z^{a_1}T^{a_2}$, where $z =$ multiplication by $z = e^{i\theta}$ on $L^2(S^1,\mathbf{C}^2) \simeq W$. Write $v_+(a)$ for the restriction of $v(a)$ to W_+ [note that $v_+(a)$ is multiplication by $z^{a_1}e^{-a_2\gamma(z)}$ on W_+]. Then $v_+(a)\in\mathrm{GL}(W_+) \subseteq \hat{\mathrm{Spin}}_Q(W)$ and we define

$$\sigma(a) = v_+(a)\sigma v_+(a)^{-1} \in \hat{\mathrm{Spin}}_Q(W) \quad (s_1s_2 \neq 1)$$

[note that $T(v_+(a)) = v(a)$]. In Ref. 5 it is proved that the $+$ state infinite volume correlations at sites $a_1, a_2 \cdots a_n \in \mathbf{Z}^2$ for the two-dimensional Ising model below the critical temperature is given by $\langle\mathcal{T}\sigma(a_1)\cdots\sigma(a_n)\rangle_Q$, where the "time ordering" $\mathcal{T}$ arranges the factors $\sigma(a_i)$ in order of increasing second coordinates from left to right. The fields $\sigma(a)$ and $\sigma(b)$ commute when a and b have coincident second coordinates so there is no ambiguity in this circumstance.

The result we are interested in is a formula expressing $\sigma(a+u)$ in terms of $\sigma(a)$ where $u = \pm e_1, \pm e_2$, and $e_1 = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, $e_2 = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. We will first consider the case $T < T_c$ and then use the representation derived in Ref. 5 to obtain results for $T > T_c$. It will suffice to consider the case $a = (0,0)$. The general result will follow from an application

of the translation operator $v(a)$. Let $R = R(\sigma(0))$ and $R(u) = R(\sigma(u))$ where $u = \pm e_1, \pm e_2$. Define $\Delta R(u)$ by $R(u) - R$. Then since $\langle\sigma(u)\rangle_Q = \langle\sigma\rangle_Q \neq 0$ $(T < T_c)$ and since (as we shall see in a moment) $\Delta R(u)$ is finite rank it follows that

$$\sigma(u) = \theta_Q(\exp[\tfrac{1}{2}\Delta R(u)]\sigma(0)). \tag{5.2}$$

Much of the work in this section will be devoted to calculating $\Delta R(u)$ regarded as an element of the algebraic Grassmann algebra $A_0(W)$.

Let $[X,Y] = XY - YX$. Then using the fact that $v(u)$ commutes with $Q_\pm$ and the derivation property for $[v(u),\cdot]$ one finds that

$$\begin{aligned}
\Delta R(u) &= v(u)(s-I)(Q_-s+Q_+)^{-1}v(u)^{-1} \\
&\quad - (s-I)(Q_-s+Q_+)^{-1} \\
&= [v(u),(s-I)(Q_-s+Q_+)^{-1}]v(u)^{-1} \\
&= \{[v(u),s](Q_-s+Q_+)^{-1} \\
&\quad + (s-I)[v(u),(Q_-s+Q_+)^{-1}]\}v(u)^{-1} \\
&= \{[v(u),s] - (s-I)(Q_-s+Q_+)^{-1}Q_-[v(u),s]\} \\
&\quad \times (Q_-s+Q_+)^{-1}v(u)^{-1} \\
&= (I-(s-I)(Q_-s+Q_+)^{-1}Q_-)[v(u),s] \\
&\quad \times (Q_-s+Q_+)^{-1}v(u)^{-1}.
\end{aligned}$$

But $I-(s-I)(Q_-s+Q_+)^{-1}Q_- = (Q_+ + sQ_-)^{-1}$ so that

$$\Delta R(u) = (Q_+ + sQ_-)^{-1}[v(u),s](Q_-s+Q_+)^{-1}v(u)^{-1}. \tag{5.3}$$

The commutator $[v(u),s]$ is finite rank for each choice of u. A closer examination of $[v(u),s]$ will suggest a basis for calculating $\Delta R(u) \in A_0(W)$. Let P_k $(k \in \mathbf{Z}_{1/2})$ denote the orthogonal projection in $l^2(\mathbf{Z}_{1/2},\mathbf{C}^2)$ on the two-dimensional subspace spanned by $\delta(\cdot - k)e_j$ $(j = 1,2)$. One calculates

$$\begin{aligned}
[z,s] &= 2P_{1/2}z = 2zP_{-1/2}, \\
[z^{-1},s] &= -2P_{-1/2}z^{-1} = -2z^{-1}P_{1/2}.
\end{aligned} \tag{5.4}$$

Since $v(\pm e_1) = z^{\pm 1}$ this gives two of the desired commutators. From (5.1) and $\cos\theta = (z+z^{-1})/2$, $\sin\theta = (z-z^{-1})/2i$ one finds

$$T(z) = T_+z + T_0 + T_-z^{-1}, \tag{5.5}$$

where

$$\begin{aligned}
T_+ &= -(2s_2)^{-1}\begin{bmatrix} s_1 & -i(c_1-1) \\ i(c_1+1) & s_1 \end{bmatrix}, \\
T_0 &= \frac{c_2}{s_2}\begin{bmatrix} c_1 & -is_1 \\ is_1 & c_1 \end{bmatrix}, \\
T_- &= -(2s_2)^{-1}\begin{bmatrix} s_1 & i(c_1+1) \\ -i(c_1-1) & s_1 \end{bmatrix}.
\end{aligned}$$

Using (5.4), (5.5), and $T(z)^{-1} = T(z^{-1})^\tau$ one finds that

$$\begin{aligned}
[T(z),s] &= 2(T_+zP_{-1/2} - T_-z^{-1}P_{1/2}), \\
[T(z)^{-1},s] &= 2(T^\tau_-zP_{-1/2} - T^\tau_+z^{-1}P_{1/2}).
\end{aligned} \tag{5.6}$$

Let $e_j(k) = \delta(\cdot - k)e_j \in l^2(\mathbf{Z}_{1/2},\mathbf{C}^2)$ denote the standard basis for l^2. Recalling (5.3) one sees the basis $v(u)(Q_-s+Q_+)^{-1}e_j(k)$ is appropriate to calculate $\Delta R(u) \in A(W)$ since it extends a basis for the null space of $\Delta R(u)$. We will not use this basis but a modification of it on the complement of the null space of $\Delta R(u)$ that takes advantage of the fact that the 2×2 matrices $T_\pm$ are both singular. Observe that

$$\begin{bmatrix} \sinh K_1 \\ -i\cosh K_1 \end{bmatrix} \in \ker(T_+), \quad \begin{bmatrix} \sinh K_1 \\ i\cosh K_1 \end{bmatrix} \in \ker(T^\tau_-),$$
$$\begin{bmatrix} \cosh K_1 \\ i\sinh K_1 \end{bmatrix} \in \ker(T^\tau_+), \quad \begin{bmatrix} -\cosh K_1 \\ i\sinh K_1 \end{bmatrix} \in \ker(T_-).$$

Consulting (5.3) and (5.6) one is lead to introduce the following basis for $P_{1/2}W \oplus P_{-1/2}W$:

$$\begin{aligned}
u(-,-) &= (\sqrt{2}s_1)^{-1}\begin{bmatrix} \sinh K_1 \\ i\cosh K_1 \end{bmatrix} \otimes \delta\left(\cdot + \frac{1}{2}\right), \\
u(-,+) &= (\sqrt{2}s_1)^{-1}\begin{bmatrix} \sinh K_1 \\ -i\cosh K_1 \end{bmatrix} \otimes \delta\left(\cdot + \frac{1}{2}\right), \\
u(+,-) &= (\sqrt{2}s_1)^{-1}\begin{bmatrix} \cosh K_1 \\ i\sinh K_1 \end{bmatrix} \otimes \delta\left(\cdot - \frac{1}{2}\right), \\
u(+,+) &= (\sqrt{2}s_1)^{-1}\begin{bmatrix} -\cosh K_1 \\ i\sinh K_1 \end{bmatrix} \otimes \delta\left(\cdot - \frac{1}{2}\right).
\end{aligned} \tag{5.7}$$

The reason for notation $u(\pm,\pm)$ and the choice of normalization will be apparent only after we have done some further calculations. The basis B we use to calculate $\Delta R(u)$ is the union of

$$\{v(u)(Q_-s+Q_+)u(\epsilon_1,\epsilon_2) | \epsilon_1 = \pm, \epsilon_2 = \pm\}$$

with

$$\{v(u)(Q_-s+Q_+)e_j(k) | k \in \mathbf{Z}_{1/2,j} = 1,2 \text{ and } k \neq \pm\tfrac{1}{2}\}.$$

We also require the basis B^* dual to B with respect to the distinguished bilinear form on W. Let

$$\{u^*(\epsilon_1,\epsilon_2) | \epsilon_1 = \pm, \epsilon_2 = \pm\}$$

denote the basis of $P_{1/2}W \oplus P_{-1/2}W$ dual to

$$\{u(\epsilon_1,\epsilon_2) | \epsilon_1 = \pm, \epsilon_2 = \pm\}$$

with respect to the bilinear form on W. Then

$$\begin{aligned}
u^*(-,-) &= \sqrt{2}\begin{bmatrix} \cosh K_1 \\ -i\sinh K_1 \end{bmatrix} \otimes \delta\left(\cdot + \frac{1}{2}\right), \\
u^*(-,+) &= \sqrt{2}\begin{bmatrix} \cosh K_1 \\ i\sinh K_1 \end{bmatrix} \otimes \delta\left(\cdot + \frac{1}{2}\right), \\
u^*(+,-) &= \sqrt{2}\begin{bmatrix} \sinh K_1 \\ -i\cosh K_1 \end{bmatrix} \otimes \delta\left(\cdot - \frac{1}{2}\right), \\
u^*(+,+) &= \sqrt{2}\begin{bmatrix} -\sinh K_1 \\ -i\cosh K_1 \end{bmatrix} \otimes \delta\left(\cdot - \frac{1}{2}\right).
\end{aligned} \tag{5.8}$$

Making use of $v(u)^\tau = v(u)^{-1}$, $Q^\tau_\pm = Q_\mp$, and $s^\tau = s$ we find that B^* is the union of

$$\{v(u)(sQ_+ + Q_-)^{-1}u^*(\epsilon_1,\epsilon_2)|, \epsilon_1 = \pm, \epsilon_2 = \pm\}$$

with

$$\{v(u)(sQ_+ + Q_-)^{-1}e_j(k) | k \in \mathbf{Z}_{1/2},\ j = 1,2 \text{ and } k \neq \pm\tfrac{1}{2}\}.$$

Using (5.3)–(5.5) one finds

$$\begin{aligned}
\Delta R(u) = \sum_{\epsilon_1 = \pm, \epsilon_2 = \pm} &(Q_+ + sQ_-)^{-1}[v(u),s]u(\epsilon_1,\epsilon_2) \\
&\wedge v(u)(Q_- + sQ_+)^{-1}u^*(\epsilon_1,\epsilon_2),
\end{aligned}$$

where the basis $u(\epsilon_1,\epsilon_2)$ has been chosen so that only two terms of the sum defining $\Delta R(u)$ survive in each case ($u = \pm e_1, \pm e_2$). Without difficulty one may verify that

$$T_+zu(-,-) = (2s_2)^{-1}u^*(+,+),$$
$$T_+zu(-,+) = 0,$$
$$T_-z^{-1}u(+,-) = -(2s_2)^{-1}u^*(-,+),$$
$$T_-z^{-1}(+,+) = 0,$$

and

$$T^\tau_- zu(-,-) = 0,$$
$$T^\tau_- zu(-,+) = -(2s_2)^{-1}u^*(+,-),$$
$$T^\tau_+ z^{-1}u(+,-) = 0,$$
$$T^\tau_+ z^{-1}u(+,+) = -(2s_2)^{-1}u^*(-,-).$$

Making use of these identities and

$$(Q_- + sQ_+)^{-1}u^*(\epsilon_1,\epsilon_2)$$
$$= (sQ_- + Q_+)^{-1}su^*(\epsilon_1,\epsilon_2)$$
$$= -\epsilon_1(sQ_- + Q_+)^{-1}u^*(\epsilon_1,\epsilon_2)$$

one finds

$$\begin{aligned}
\Delta R(e_1) &= s_1^{-1}(w(+,-)\wedge zw(-,+) \\
&\quad - w(+,+)\wedge zw(-,-)), \\
\Delta R(-e_1) &= s_1^{-1}(w(-,+)\wedge z^{-1}w(+,-) \\
&\quad - w(-,-)\wedge z^{-1}w(+,+)), \\
\Delta R(e_2) &= s_2^{-1}(w(+,+)\wedge T(z)w(-,-) \\
&\quad - w(-,+)\wedge T(z)w(+,-)), \\
\Delta R(-e_2) &= s_2^{-1}(w(-,-)\wedge T(z)^{-1}w(+,+) \\
&\quad - w(+,-)\wedge T(z)^{-1}w(-,+)),
\end{aligned} \tag{5.9}$$

where

$$w(\epsilon_1,\epsilon_2) \overset{\text{def}}{=} (Q_+ + sQ_-)^{-1}u^*(\epsilon_1,\epsilon_2).$$

We will now calculate the vectors $w(\epsilon_1,\epsilon_2)$ in coordinates that are natural for the transfer matrix. This calculation will simplify (5.9) and will also show that the vectors $w(\epsilon_1,\epsilon_2)$ "cry out" to be located on the half-integer lattice $\mathbf{Z}_{1/2}\times\mathbf{Z}_{1/2}$. We begin by using the Weiner–Hopf method to calculate $(Q_+ + sQ_-)^{-1}$. Let $\epsilon_\pm = (1\pm\epsilon)/2$ so that $s = \epsilon_- - \epsilon_+$. Then $(Q_+ + sQ_-) = (\epsilon_- + \epsilon_- Q)$.

Now suppose that $Q(z) = A_-(z)A_+(z)$, where $A_-(z)$ has an analytic (invertible) continuation into the exterior of the circle $|z| = 1$ and $A_+(z)$ has an analytic (invertible) continuation into the interior of the circle. Standard Weiner–Hopf arguments show that

$$(\epsilon_- + \epsilon_+A_-A_+)^{-1} = I + A_+^{-1}\epsilon_+(A_-^{-1} - A_+). \tag{5.10}$$

For $T < T_c$ the matrix

$$Q(z) = \begin{bmatrix} 0 & a(z) \\ a(z)^{-1} & 0 \end{bmatrix}$$

does have a Weiner–Hopf factorization as we now demonstrate. Going back to the definition of $a(z)$ one finds

$$a(z)^2 = \frac{a(z)}{\overline{a(z)}} = \frac{\sin\theta - i(s_1c_2 - c_1\cos\theta)}{\sin\theta + i(s_1c_2 - c_1\cos\theta)}$$
$$= -\frac{1-\alpha_2z^{-1}}{1-\alpha_1z^{-1}}\cdot\frac{1-\alpha_1z}{1-\alpha_2z},$$

where $\alpha_1 = e^{-2K_1^*}e^{-2K_2}$ and $\alpha_2 = e^{2K_1^*}e^{-2K_2}$. Here the "dual" interaction strengths K_j^* are defined by $\sinh 2K_j^* \sinh 2K_j = 1$.

The condition $T < T_c$ is equivalent to $\alpha_1 < \alpha_2 < 1$. We define

$$\phi_-(z) = \left(\frac{1-\alpha_2z^{-1}}{1-\alpha_1z^{-1}}\right)^{1/2}, \quad \phi_+(z) = i\left(\frac{1-\alpha_1z}{1-\alpha_2z}\right)^{1/2},$$

where the square roots are normalized so that they are positive for $z = -1$. It is clear then that $a(z) = \phi_-(z)\phi_+(z)$, where $\phi_-(z)$ has an analytic continuation into the exterior of $|z| = 1$ and $\phi_+(z)$ has an analytic continuation into the interior of $|z| = 1$. We define

$$A_-(z) = \begin{pmatrix} \phi_-(z) & 0 \\ 0 & \phi_-(z)^{-1} \end{pmatrix},$$
$$A_+(z) = \begin{pmatrix} 0 & \phi_+(z) \\ \phi_+(z)^{-1} & 0 \end{pmatrix},$$

so that $Q(z) = A_-(z)A_+(z)$. We now calculate $(Q_+ + sQ_-)^{-1}e_j(\pm\frac{1}{2})$ using (5.10). Since $\epsilon_+A_+e_j(\frac{1}{2}) = A_+e_j(\frac{1}{2})$ we have

$$(I + A_+^{-1}\epsilon_+(A_-^{-1} - A_+))e_j(\tfrac{1}{2})$$
$$= A_+\epsilon_+A_-(z)^{-1}e_j(\tfrac{1}{2})$$
$$= A_+(z)A_-(\infty)^{-1}e_j(\tfrac{1}{2})$$
$$= A_+(z)e_j(\tfrac{1}{2}).$$

Since

$$\epsilon_+(A_+e_j(-\tfrac{1}{2})) = A_+e_j(-\tfrac{1}{2}) - A_+(0)e_j(-\tfrac{1}{2})$$

we have

$$(I + A_+^{-1}\epsilon_+(A_-^{-1} - A_+))e_j(-\tfrac{1}{2})$$
$$= e_j(-\tfrac{1}{2}) - A_+^{-1}\epsilon_+A_+e_j(-\tfrac{1}{2})$$
$$= A_+^{-1}(z)A_+(0)e_j(-\tfrac{1}{2}).$$

Thus

$$\begin{aligned}
(Q_+ + sQ_-)^{-1}e_1(\tfrac{1}{2}) &= \begin{bmatrix} 0 \\ \phi_+^{-1}(z)\sqrt{z} \end{bmatrix}, \\
(Q_+ + sQ_-)^{-1}e_2(\tfrac{1}{2}) &= \begin{bmatrix} \phi_+(z)\sqrt{z} \\ 0 \end{bmatrix}, \\
(Q_+ + sQ_-)^{-1}e_1(-\tfrac{1}{2}) &= \begin{bmatrix} -i\phi_+(z)\sqrt{z}^{-1} \\ 0 \end{bmatrix}, \\
(Q_+ + sQ_-)^{-1}e_2(-\tfrac{1}{2}) &= \begin{bmatrix} 0 \\ i\phi_+^{-1}(z)\sqrt{z}^{-1} \end{bmatrix}.
\end{aligned} \tag{5.11}$$

Recall that the square root $\sqrt{z}$ is calculated with $0 < \arg z < 2\pi$. With this choice note that $(z^{-1})^{1/2} = -(z^{1/2})^{-1}$. Next we calculate the vectors in (5.11) in a spectral representation for $T(z)$. A more complete rationale for the transformation to this representation is given in Ref. 11. Let $M(z) = (a(z)\sinh\gamma(z))^{1/2}$ where the square root is normalized so that $M(-1) = [(1+i)/\sqrt{2}]\times$(positive number). Since

$a(z)\sinh\gamma(z)$ does not wind around 0 for $T<T_c$ a unique continuous square root is picked out by this normalization. Now define

$$f_+(z)=[(1+i)/2](\bar{M}(z)f_1(z)+M(z)f_2(z)),$$
$$f_-(z)=[(1-i)/2](M(z^{-1})\bar{f}_1(z^{-1})-\bar{M}(z^{-1})\bar{f}_2(z^{-1})). \tag{5.12}$$

Then

$$\int_{S^1} f(z)\cdot\overline{g(z)}\,\frac{dz}{2\pi iz}$$
$$=\int_{S^1} f_+(z)\,\overline{g_+(z)}(\sinh\gamma(z))^{-1}\,\frac{dz}{2\pi iz}$$
$$+\int_{S^1}\overline{f_-(z)}g_-(z)(\sinh\gamma(z))^{-1}\,\frac{dz}{2\pi iz},$$

and the map

$$\begin{bmatrix} f_1(z)\\ f_2(z)\end{bmatrix}\to\begin{bmatrix} f_+(z)\\ f_-(z)\end{bmatrix}$$

is a spectral representation for $T(z)$ in the sense that $T(z)$ is given by the diagonal matrix multiplication operator $\left[\begin{smallmatrix} e^{-\gamma(z)} & 0\\ 0 & e^{\gamma(z)}\end{smallmatrix}\right]$ in the $\left[\begin{smallmatrix} f_+\\ f_-\end{smallmatrix}\right]$ coordinates.

It is straightforward to calculate the $\left[\begin{smallmatrix} f_+\\ f_-\end{smallmatrix}\right]$ coordinates for $(Q_++sQ_-)^{-1}e_j(\pm\frac{1}{2})$ from (5.11) and (5.12). One finds

$$(Q_++sQ_-)^{-1}e_1\left(\frac{1}{2}\right)=\frac{1+i}{\sqrt{2}}\begin{bmatrix} F(z)\sqrt{z}\\ iF(z)\sqrt{z}\end{bmatrix},$$

$$(Q_++sQ_-)^{-1}e_2\left(\frac{1}{2}\right)=\frac{1+i}{\sqrt{2}}\begin{bmatrix} G(z)\sqrt{z}\\ iG(z)\sqrt{z}\end{bmatrix},$$

$$(Q_++sQ_-)^{-1}e_1\left(-\frac{1}{2}\right)=\frac{1+i}{\sqrt{2}}\begin{bmatrix} -iG(z)\sqrt{z}^{-1}\\ -G(z)\sqrt{z}^{-1}\end{bmatrix},$$

$$(Q_++sQ_-)^{-1}e_2\left(-\frac{1}{2}\right)=\frac{1+i}{\sqrt{2}}\begin{bmatrix} iF(z)\sqrt{z}^{-1}\\ F(z)\sqrt{z}^{-1}\end{bmatrix},$$

where $F(z)=[\phi_-(z)\sinh\gamma(z)/\phi_+(z)]^{1/2}$ is the continuous square root normalized so that $F(-1)=[(1-i)/\sqrt{2}]\times$(positive number), and $G(z)=[\phi_+(z)\sinh\gamma(z)/\phi_-(z)]^{1/2}$ is the continuous square root normalized so that $G(-1)=[(1+i)/\sqrt{2}]\times$(positive number). We also used the fact that

$$i\begin{bmatrix} f_+\\ f_-\end{bmatrix}=\begin{bmatrix} if_+\\ -if_-\end{bmatrix}.$$

We can simplify $F(z)$ and $G(z)$ by making the following observations:

$$\sinh^2(\gamma/2)=(\cos\gamma-1)/2$$
$$=(s_1/4s_2\alpha_2)(1-\alpha_2z^{-1})(1-\alpha_2z),$$
$$\cosh^2(\gamma/2)=(\cosh\gamma-1)/2$$
$$=(s_1/4s_2\alpha_1)(1-\alpha_1z^{-1})(1-\alpha_1z)$$

[since $\cosh\gamma=s_2^{-1}(c_1c_2-s_1\cosh\theta)$] from which it follows that

$$\frac{\phi_-(z)\sinh\gamma(z)}{\phi_+(z)}=\frac{a(z)\sinh\gamma(z)}{\phi_+(z)^2}$$
$$=-2i\sqrt{\frac{\alpha_2}{\alpha_1}}\sinh^2(\gamma/2),$$

$$\frac{\phi_+(z)\sinh\gamma(z)}{\phi_-(z)}=\frac{a(z)\sinh\gamma(z)}{\phi_-(z)^2}$$
$$=2i\sqrt{\frac{\alpha_1}{\alpha_2}}\cosh^2(\gamma/2).$$

Thus

$$F(z)=(1-i)e^{K\dagger}\sinh(\gamma/2),$$
$$G(z)=(1+i)e^{-K\dagger}\cosh(\gamma/2),$$

so that

$$(Q_++sQ_-)^{-1}e_1\left(\frac{1}{2}\right)=\sqrt{2}e^{K\dagger}\begin{bmatrix}\sinh(\gamma(z)/2)\sqrt{z}\\ \sinh(\gamma(z)/2)\sqrt{z}\end{bmatrix},$$
$$(Q_++sQ_-)^{-1}e_2\left(\frac{1}{2}\right)=\sqrt{2}e^{-K\dagger}\begin{bmatrix} i\cosh(\gamma(z)/2)\sqrt{z}\\ i\cosh(\gamma(z)/2)\sqrt{z}\end{bmatrix},$$
$$(Q_++sQ_-)^{-1}e_1\left(-\frac{1}{2}\right)$$
$$=\sqrt{2}e^{-K\dagger}\begin{bmatrix}\cosh(\gamma(z)/2)\sqrt{z}^{-1}\\ -\cosh(\gamma(z)/2)\sqrt{z}^{-1}\end{bmatrix}, \tag{5.13}$$
$$(Q_++sQ_-)^{-1}e_2\left(-\frac{1}{2}\right)$$
$$=\sqrt{2}e^{K\dagger}\begin{bmatrix} i\sinh(\gamma(z)/2)\sqrt{z}^{-1}\\ -i\sinh(\gamma(z)/2)\sqrt{z}^{-1}\end{bmatrix},$$

where the terms on the right are the $\left[\begin{smallmatrix} f_+\\ f_-\end{smallmatrix}\right]$ coordinates of the vectors on the left. Finally we may use (5.13) to calculate the $[f_+,f_-]$ representation for the vectors

$$w(\epsilon_1,\epsilon_2)=(Q_++sQ_-)^{-1}u^*(\epsilon_1,\epsilon_2).$$

One finds

$$w(-,-)=\sqrt{s_1}\begin{bmatrix} e^{\gamma(z)/2}\sqrt{z}^{-1}\\ -e^{-\gamma(z)/2}\sqrt{z}^{-1}\end{bmatrix},$$
$$w(-,+)=\sqrt{s_1}\begin{bmatrix} e^{-\gamma(z)/2}\sqrt{z}^{-1}\\ -e^{\gamma(z)/2}\sqrt{z}^{-1}\end{bmatrix},$$
$$w(+,-)=\sqrt{s_1}\begin{bmatrix} e^{\gamma(z)/2}\sqrt{z}\\ -e^{-\gamma(z)/2}\sqrt{z}\end{bmatrix}, \tag{5.14}$$
$$w(+,+)=\sqrt{s_1}\begin{bmatrix} e^{-\gamma(z)/2}\sqrt{z}\\ -e^{\gamma(z)/2}\sqrt{z}\end{bmatrix}.$$

Since $T(z)$ is multiplication by $\left[\begin{smallmatrix} e^{-\gamma(z)} & 0\\ 0 & e^{\gamma(z)}\end{smallmatrix}\right]$ and z is multiplication by $\left[\begin{smallmatrix} z & 0\\ 0 & z\end{smallmatrix}\right]$ in the spectral representation (5.12) it follows from (5.14) that $T(z)w(\epsilon,-)=w(\epsilon,+)$ and $zw(-,\epsilon)=w(+,\epsilon)$. Making use of these results in (5.8) we find

$$\Delta R(e_1)=2s_1^{-1}w(+,-)\wedge w(+,+),$$
$$\Delta R(-e_1)=2s_1^{-1}w(-,+)\wedge w(-,-),$$
$$\Delta R(e_2)=2s_2^{-1}w(+,+)\wedge w(-,+), \tag{5.15}$$
$$\Delta R(-e_2)=2s_2^{-1}w(-,-)\wedge w(+,-).$$

We have established the following theorem.

Theorem 5.1: Let $u = \pm e_1, \pm e_2$. Then

$$\sigma(u) = \theta_Q \exp[\tfrac{1}{2}\Delta R(u)]\sigma(0), \tag{5.16}$$

where $\Delta R(u)$ is given by (5.15) and the vectors $w(\epsilon_1,\epsilon_2)$ have spectral representations given by (5.14).

Remark: Since $zw(-,\epsilon) = w(+,\epsilon)$ and $T(z)w(\epsilon,-) = w(\epsilon,+)$ it is quite natural to "locate" the vector $w(\epsilon_1,\epsilon_2)$ at the point $(\epsilon_1(\frac{1}{2}),\epsilon_2(\frac{1}{2}))$ on the half-integer lattice $\mathbf{Z}_{1/2}\times\mathbf{Z}_{1/2}$. If this is done then (5.15) may be remembered with the following graphical device:

$$\begin{array}{ccc} w(-,+) & \longleftarrow & w(+,+) \\ \downarrow & (0,0) \quad (1,0). & \uparrow \\ w(\mp,-) & \longrightarrow & w(+,-) \end{array}$$

To get from (0,0) to (1,0) one crosses the directed bond joining $w(+,-)$ to $w(+,+)$ in the picture. The factor $w(+,-)\wedge w(+,+)$ with weight s_1^* is just what is needed to obtain $\sigma(e_1)$ from $\sigma(0)$ via (5.16). The other cases are precisely analogous.

The vectors $w(\epsilon_1,\epsilon_2)$ have another significance that we now describe. It is easy to check that

$$w(-,+) = \sqrt{2}(Q_+ + sQ_-)^{-1}(\cosh K_1 e_1(-\tfrac{1}{2}) + i\sinh K_1 e_2(-\tfrac{1}{2}))$$

so that

$$N(w(-,+)\sigma) = \sqrt{2}[\cosh K_1 F(e_1(-\tfrac{1}{2})) + i\sinh K_1 F(e_2(-\tfrac{1}{2}))]\hat{\Gamma}(\sigma). \tag{5.17}$$

In Ref. 5 the operator on the right-hand side of (5.17) was denoted by $\mu_{-1/2}$. Here it is more appropriate to write $\mu(-\frac{1}{2},\frac{1}{2})$ for this operator. Let $V(m) = \Gamma(v_+(m))$ and define

$$\mu(k_1,k_2) = V(k_1+\tfrac{1}{2},k_2-\tfrac{1}{2})\mu(-\tfrac{1}{2},\tfrac{1}{2})V(k_1+\tfrac{1}{2},k_2-\tfrac{1}{2})^{-1}, \tag{5.18}$$

for $(k_1,k_2)\in\mathbf{Z}^2_{1/2}$. The significance of the disorder variables $\mu(k_1,k_2)$ was established in Ref. 5. Suppose one has interaction strengths K_j $(j=1,2)$ given for $T<T_c$ $(\sinh 2K_1 \times\sinh 2K_2 > 1)$. The dual interaction strengths K_j^* then define a model with $T>T_c$ (recall $\sinh 2K_j \sinh 2K_j^* = 1$). Let $\tau_{K^*}(k_1,\ldots,k_n)$ denote the infinite volume correlations for the Ising model with interaction strengths K_1^*, K_2^* at sites $k_j\in\mathbf{Z}^2_{1/2}$ on the half-integer lattice. Then it is proved in Ref. 5 (Sec. 3) that

$$\tau_{K^*}(k_1,\ldots,k_n) = \langle\mathcal{T}\mu(k_1)\cdots\mu(k_n)\rangle_{Q(K)}, \tag{5.19}$$

where the time ordering $\mathcal{T}$ puts the operators $\mu(k_j)$ in order of increasing second coordinates from left to right. The operators $\mu(k)$ and $\mu(l)$ commute when the second coordinates of k and l agree so the time ordering prescription does not lead to ambiguity in (5.19). The notation $Q(K)$ means that the vacuum expectation on the right of (5.19) should be calculated at interaction strengths K_1 and K_2 (below T_c). We may summarize (5.19) by saying that the correlations above the critical temperature are given by vacuum expectations of disorder variables below the critical temperature with interaction strengths related by the duality $\sinh 2K_j \sinh 2K_j^* = 1$.

We are almost ready to derive the McCoy, Wu, and Perk difference equations. Before we do this we will show that all the $w(\epsilon_1,\epsilon_2)$ occur on an equal footing in the sense that

$$N(w(\epsilon_1,\epsilon_2)\sigma) = \mu(\epsilon_1/2,\epsilon_2/2).$$

It will be enough to prove this for $w(+,+)$, the other cases are analogous. We have by definition $N(w(-,+)\sigma) = \mu(-\frac{1}{2},\frac{1}{2})$. Making a similarity transformation of both sides by $V(e_1)$ one finds

$$\mu(\tfrac{1}{2},\tfrac{1}{2}) = N(w(+,+)\sigma(1,0))$$

since $zw(-,+) = w(+,+)$. But Theorems 3.3 and 5.1 imply that

$$N(\sigma(1,0)) = N([1+s_1^{-1}w(+,-)w(+,+)]\sigma(0)).$$

Thus

$$\begin{aligned} &N(w(+,+)\sigma(1,0)) \\ &\quad = N(w(+,+) \\ &\qquad \times[1+s_1^{-1}w(+,-)w(+,+)]\sigma(0)) \\ &\quad = N(w(+,+)\sigma(0)) \end{aligned}$$

since $w(+,+)^2 = 0$ in the Grassmann algebra.

We are now prepared to use Theorem 5.1 to derive quadratic difference identities for the Ising correlations. It will be useful to simplify notational matters by concentrating on the two point functions. For $a\in\mathbf{Z}^2$ define

$$\tau(a) = \langle\mathcal{T}\sigma(0)\sigma(a)\rangle_{Q(K)},$$
$$\tau^*(a) = \langle\mathcal{T}\mu(k)\mu(k+a)\rangle_{Q(K)}, \quad k\in\mathbf{Z}^2_{1/2},$$

where $Q(K)$ is Q evaluated at K_1, K_2 with $\sinh 2K_1 \sinh 2K_2 > 1$.

Let $u = \pm e_1, \pm e_2$ and define

$$F_u(a) = \tau(a+u)/\tau(a), \quad F_u^*(a) = \tau^*(a+u)/\tau^*(a).$$

We shall derive the MWP difference identities by seeking a relation for $F_u(a+u')$, where $u' = \pm e_1, \pm e_2$. But

$$\begin{aligned} F_u(a+u') &= \tau(a+u+u')/\tau(a+u') \\ &= \tau(a+u+u')/\tau(a)F_{u'}(a)^{-1}. \end{aligned}$$

Thus

$$F_{u'}(a)F_u(a+u') = \tau(a+u+u')/\tau(a).$$

It is convenient to suppose that the second coordinate of $a(\pi_2(a))$ is sufficiently large so that $\pi_2(a)\geqslant 0$, $\pi_2(a+u)\geqslant 0$, $\pi_2(a+u')\geqslant 0$, and $\pi_2(a+u+u')\geqslant 0$. Then we have

$$\begin{aligned} \tau(a+u+u') &= \langle\sigma(0)\sigma(a+u+u')\rangle \\ &= \langle\sigma(-u)\sigma(a+u)\rangle \\ &= \langle(\sigma(-u)-\sigma(0))(\sigma(a+u')-\sigma(a))\rangle \\ &\quad + \langle\sigma(0)\sigma(a+u')\rangle \\ &\quad + \langle\sigma(-u)\sigma(a)\rangle - \langle\sigma(0)\sigma(a)\rangle, \end{aligned}$$

from which it follows that

$$\begin{aligned} &\frac{\tau(a+u+u')}{\tau(a)} \\ &\quad = F_u(a) + F_{u'}(a) - 1 \\ &\qquad + \frac{\langle(\sigma(-u)-\sigma(0))(\sigma(a+u')-\sigma(a))\rangle}{\tau(a)}. \end{aligned} \tag{5.20}$$

The rest of the calculation is the use of Theorem 3.3, Theorem 5.1, and the generalized Wick theorem to evaluate the last term in (5.20). In order to keep track of the points on the half-integer lattice that arise in the application of Theorem 5.1 it is useful to introduce complex notation for points in $\mathbf{Z}^2$ and $\mathbf{Z}^2_{1/2}$ writing $(a_1,a_2) = a_1 + ia_2$. We may summarize Theorem 5.1 as follows:

$$\hat{\Gamma}(\sigma(a+u)) - \hat{\Gamma}(\sigma(a))$$
$$= s(u)^{-1} N(w(a + [(1-i)/2]u)w \times (a + [(1+i)/2]u)\sigma(a)),$$

where $s(\pm e_j) = s_j$ $(j = 1,2)$. Now substitute this in the last term in (5.20) and use Theorem 4.3 to get

$$\langle (\sigma(-u) - \sigma(0))(\sigma(a+u') - \sigma(a))\rangle/\tau(a) = \mathrm{Pf}(G),$$

where G is a 4×4 skew symmetric matrix with entries above the diagonal:

$$G_{12} = \frac{\langle (\sigma(-u) - \sigma(0))\sigma(a)\rangle}{\tau(a)} = F_u(a) - 1,$$

$$G_{13} = \frac{\langle \mu([(i-1)/2]u)\mu(a + [(1-i)/2]u')\rangle}{\tau(a)} = \frac{\tau^*(a + (u+u')/2 - i(u+u')/2)}{\tau(a)},$$

$$G_{14} = \frac{\langle \mu([(i-1)/2]u)\mu(a + [(1-i)/2]u')\rangle}{\tau(a)} = \frac{\tau^*(a + (u+u')/2 + i(u-u')/2)}{\tau(a)},$$

$$G_{23} = \frac{\langle \mu(-[(1+i)/2]u)\mu(a + [(1-i)/2]u')\rangle}{\tau(a)} = \frac{\tau^*(a + (u+u')/2 + i[(u-u')/2])}{\tau(a)},$$

$$G_{24} = \frac{\langle \mu(-[(1+i)/2]u)\mu(a + [(1-i)/2]u')\rangle}{\tau(a)} = \frac{\tau^*(a + (u+u')/2 + i[(u-u')/2])}{\tau(a)},$$

$$G_{34} = \frac{\langle \sigma(0)(\sigma(a+u') - \sigma(a))\rangle}{\tau(a)} = F_{u'}(a) - 1.$$

Since $\mathrm{Pf}(G) = G_{12}G_{34} - G_{13}G_{24} + G_{14}G_{23}$ one finds after substitution in (5.20)

$$F_{u'}(a)(F_u(a+u') - F_u(a))$$
$$= -(s(u)s(u'))^{-1}\tau(a)^{-2}(\tau^*(a^*)\tau^*(a^* + iu + iu') - \tau^*(a^* + iu')\tau^*(a^* + iu)),$$

where

$$a^* \overset{\text{def}}{=} a + (u+u')/2 - i[(u+u')/2].$$

Clearing the denominator $\tau(a)^2$ one finds the more symmetrical

$$\tau(a)\tau(a+u+u') - \tau(a+u')\tau(a+u)$$
$$= -(s(u)s(u'))^{-1}(\tau^*(a^*)\tau^*(a^* + iu + iu') - \tau^*(a^* + iu')\tau^*(a^* + iu)),$$

which is a slight variant of the MWP difference identities.

ACKNOWLEDGMENT

Research for this paper was supported in part by National Science Foundation Grant No. DMS-8421289.

[1] M. Sato, T. Miwa, and M. Jimbo, "Holonomic quantum fields I–V," Publ. R.I.M.S. **14**, 223 (1978); **15**, 201, 577, 871 (1979); **16**, 532 (1980).
[2] A. Carey and J. Palmer, "Infinite complex spin groups," J. Func. Anal. (to appear).
[3] J. Palmer, "Products in spin representations," Adv. Appl. Math. **2**, 329 (1981).
[4] J. Palmer and C. Tracy, "Two dimensional Ising correlations: the SMJ analysis," Adv. Appl. Math. **4**, 46 (1983).
[5] J. Palmer and C. Tracy, "Two dimensional Ising correlations: convergence of the scaling limit," Adv. Appl. Math. **2**, 329 (1981).
[6] J. Palmer, "Critical scaling for monodromy fields," Commun. Math. Phys. **104**, 353 (1986).
[7] B. M. McCoy and T. T. Wu, "Nonlinear partial difference equations for the two spin correlation functions of the two dimensional Ising model," Nucl. Phys. B **180**, 89 (1981).
[8] B. M. McCoy, J. H. H. Perk, and T. T. Wu, "Ising field theory: quadratic difference equations for the n-point Green's function on the lattice," Phys. Rev. Lett. **45**, 675 (1980).
[9] J. H. H. Perk, "Quadratic identities for Ising correlations," Phys. Lett. A **1 79**, 3 (1980).
[10] G. Segal and G. Wilson, "Loop groups and equations of KdV type," I.H.E.S. Publ. Math. **61**, 5 (1985).
[11] J. Palmer, "Monodromy fields on Z^2," Commun. Math. Phys. **102**, 175 (1985).

Gell-Mann formula for simple complex Lie groups and geometric quantization

P. Šťovíček[a)]
Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, SU-141980 Dubna, USSR

(Received 22 October 1987; accepted for publication 20 January 1988)

A Lie–Poisson isomorphism for Lie algebras $\mathfrak{g}$ and $\mathfrak{q} = \mathfrak{k} \times \mathfrak{k}$ is derived in a closed form, where $\mathfrak{g}$ is complex and simple, and $\mathfrak{k}$ is the maximal compact subalgebra in $\mathfrak{g}$. This isomorphism enables one to find a generalized Gell-Mann formula relating unitary irreducible representations of the corresponding Lie groups G and Q. The case $\mathfrak{g} = \mathrm{sl}(n,\mathbb{C})$, $\mathfrak{k} = \mathrm{su}(n)$ is considered in detail. A simple relation between characters is obtained.

I. INTRODUCTION

The Gell-Mann formula sets up a correspondence between unitary representations of two Lie groups that have algebraically different structures but are related by their physical or geometrical meaning. The typical example is that of the de Sitter group and the Poincaré group. This correspondence is stated in an infinitesimal form, it represents a polynomial dependence between generators of the corresponding Lie algebras. Validity of the "standard" Gell-Mann formula in the case of a semisimple Lie group was investigated by Hermann.[1] The "standard" Gell-Mann formula means a correspondence of the form $P_j = T_j + c\,[\Delta^{(2)}, T_j]$. Weimar proved[2] that the validity in this case is restricted to the pseudo-orthogonal algebras only. The only attempt to generalize the Gell-Mann formula, to the author's knowledge, was made by Mukunda[3] for the groups $\mathrm{SL}(3,\mathbb{C})$ and $\mathrm{SU}(3) \circledS \mathrm{su}(3)$. Mackey[4] investigated the relationship between unitary representations of a semisimple Lie group G, on one hand, and the semidirect product $Q = K \circledS T$ of the maximal compact subgroup $K \subset G$ with the additive group T of the vector space $T_0(G/K)$, on the other hand. He did not take care for the Gell-Mann formula but compared the constructions of unitary representations for both groups and looked for an analogy, which in some cases is quite striking.

The main motivation for considering the Gell-Mann formula is the observation that unitary representations of semidirect product groups are comparatively simpler in their construction than those of semisimple groups. In the former case there exists a systematic approach based on a general theorem due to Mackey.[5] Another meaning of the formula arises when one is dealing with the enveloping algebras and fields.[6]

In this paper we exploit the idea suggested by one of the authors of Ref. 6 (Havlíček), who insists on comparing coadjoint orbits for both groups. This procedure is based on the method of orbits and geometric quantization.[7,8] Suppose we are given a symplectomorphism which, moreover, preserves the polarizations on the corresponding orbits. Then we can identify in a distinguished way the underlying Hilbert spaces and compare both representations. We confine ourselves to the case where G is a complex simple Lie group, connected and simply connected, $Q = K \circledS \mathfrak{k}$. In this case we are able to describe in a closed form a Poisson isomorphism $J: \mathfrak{q}^* \to \mathfrak{g}^*$ defined on an open dense subset. Further, we shall show that there exists a generalized Gell-Mann formula; a more detailed description for the group $G = \mathrm{SL}(n,\mathbb{C})$ is postponed to the Appendix. The obtained correspondence between unitary representations coincides with that proposed by Mackey. Moreover, there a simple relation between characters is found.

II. NOTATION

Let G be a complex simple Lie group, connected and simply connected, $n = \dim_{\mathbb{C}} G$, $l = \mathrm{rank}\ G$, $G = KAN$ (the Iwasawa decomposition), M the maximal torus in K, $H = MA$ the Cartan subgroup, and $B = MAN$ the Borel subgroup. Let $\mathfrak{g},\mathfrak{k},\mathfrak{a},\mathfrak{n},\mathfrak{m},\mathfrak{h},\mathfrak{b}$ be the corresponding Lie algebras, $B(\cdot,\cdot)$ the complex Killing form in $\mathfrak{g}$, $\langle\cdot,\cdot\rangle = \mathrm{Re}\,B(\cdot,\cdot)$ ($\langle\cdot,\cdot\rangle$ restricted to $\mathfrak{k}$ is the negative definite real Killing form in $\mathfrak{k}$). Let $\Delta_- = \Delta_+ \cup \Delta_-$ be the set of roots, $\delta = \frac{1}{2}\Sigma_{\alpha>0}\alpha$, $\{H_\alpha, E_\alpha\}_{\alpha\in\Delta}$ the Weyl generators normed by $[E_\alpha, E_{-\alpha}] = -H_\alpha$, $B(E_\alpha, E_{-\alpha}) = -1$. Let $X \mapsto \overline{X}$ be the complex conjugation in $\mathfrak{g}$ with respect to $\mathfrak{k}$ ($\mathfrak{g} = \mathfrak{k} + i\mathfrak{k}$ over reals), $\overline{H}_\alpha = -H_\alpha$, $\overline{E}_\alpha = E_{-\alpha}$. Let $\mathfrak{n}_- = \bar{\mathfrak{n}}$ and $\mathfrak{u}$ be the orthogonal complement of $\mathfrak{m}$ in $\mathfrak{k}$.

We identify the dual space $\mathfrak{g}^*$ with $\mathfrak{g}$ by the bilinear form $\langle\cdot,\cdot\rangle$. Consequently, the coadjoint representation is replaced by the adjoint one. Let $C_+ \subset \mathfrak{a}$ be the dominant Weyl chamber, $\mathfrak{m}_+ = iC_+$, $\mathfrak{h}_+ = \mathfrak{a} + \mathfrak{m}_+$, $\mathscr{W}$ the Weyl group identified with the normalizer factorized by the centralizer of A in K. Let $\epsilon_1,\ldots,\epsilon_l$ be the fundamental weights in C_+, $\Lambda = \mathbb{Z}\epsilon_1 + \cdots + \mathbb{Z}\epsilon_l$.

We denote by $Q = K \circledS \mathfrak{k}$ the semidirect product, where the second term $\mathfrak{k}$ means the additive group of the underlying vector space, the action of K in $\mathfrak{k}$ coincides with the adjoint representation, $\mathfrak{q} = \mathfrak{k} \circledS \mathfrak{k}$. We identify $\mathfrak{q}^*$ with $\mathfrak{q}$ by the bilinear form $(X_1,Y_1),(X_2,Y_2) \mapsto \langle X_1,X_2\rangle - \langle Y_1,Y_2\rangle$. Then

$$\mathrm{Ad}^*(k,Z)(X,Y) = (\mathrm{Ad}\,k(X - [Z,Y]), \mathrm{Ad}\,k(Y)),$$
$$k \in K, \quad X,Y,Z \in \mathfrak{k}.$$

The images of the above specified embeddings of K into G and Q are identified; the same is true for the Lie algebras. The identification of $\mathfrak{m} \times \mathfrak{m} \subset \mathfrak{q}$ with $\mathfrak{h} \subset \mathfrak{g}$ by the mapping $(X,Y) \mapsto X + iY$ will also turn out to be useful.

[a)] Permanent address: Department of Mathematics, Faculty of Nuclear Science, Czech Technical University, Prague, Czechoslovakia.

 0022-2488/88/061300-08$02.50

III. POISSON STRUCTURES

In this section we summarize some necessary facts on Poisson structures and refer to Weinstein's paper[9] where the subject is studied in detail.

Let P be a smooth manifold, and $\{\ ,\ \}$ designate a Poisson structure on P. Let ξ be the Lie algebra homomorphism from $C^\infty(P)$ into the Lie algebra of vector fields $\mathfrak{X}(P)$ defined by $\xi_f \cdot g = \{f,g\}$. Vector fields ξ_f span a smooth involutive distribution D, which is integrable.[10] The integrability of D is an immediate consequence of the generalized Frobenius theorem due to Sussmann and Stefan[11]: A smooth distribution D is integrable if and only if it is involutive and the dimension of D remains constant on all integral curves of vector fields belonging to D.

Consequently, there exists a foliation with singularities on P. Each leaf becomes a symplectic manifold, the symplectic form is defined by $\omega(\xi_f,\xi_g) = -\{f,g\}$. Weinstein's splitting theorem gives more information. To each point $p\in P$ there belongs a canonically defined transversal tangent Lie algebra.

A mapping $J: P_1,\{\ ,\ \}_1 \to P_2,\{\ ,\ \}_2$ is said to be Poisson if the pullback $J^*: C^\infty(P_2) \to C^\infty(P_1)$ is a homomorphism. In this case $J_*(p)$ maps $D_1(p)$ onto $D_2(J(p))$, $J_*(\xi_{J^*f})(p) = \xi_f(J(p))$. We shall need the following criterion and its corollary. We denote by ω_j the symplectic forms living on leaves of foliations on $P_j, j = 1,2$, and by V_p the subspace of $D_1(p)$ [the ω_1-orthogonal complement of the kernel $\ker J_*(p)$, i.e., $V_p = \ker(J_*|_D)_p^\perp$]. Then J is a Poisson mapping if and only if $J^*\omega_2|_V = \omega_1|_V$.

Corollary: Suppose, moreover, that the derivative $J_*(p)$ restricted to the vector space D_p is injective. Then the Poisson mapping J induces a local symplectomorphism at the point p of leaves passing through the points p and $J(p)$.

Let $\mathfrak{g}$ be any Lie algebra. The Lie–Poisson structure is defined on $\mathfrak{g}^*$ by $\{f,g\}(F) = \langle F,[df_F,dg_F]\rangle$. We identify $\mathfrak{g}^{**}$ with $\mathfrak{g}$. Leaves in $\mathfrak{g}^*$ are exactly the coadjoint orbits with their standard symplectic structure. Let P be another Poisson manifold, $J: P \to \mathfrak{g}^*$ a smooth mapping, $\lambda_X = J^*X$ for $X\in\mathfrak{g}^{**} = \mathfrak{g}$. Then J is a Poisson mapping if and only if $\lambda: \mathfrak{g} \to C^\infty(P)$ is a homomorphism. Then J is called a momentum mapping. The Poisson action λ determines an action of $\mathfrak{g}$ on P by the homomorphism $\eta: \mathfrak{g} \to \mathfrak{X}(P)$, $\eta_X = \xi_{\lambda(X)}$.

Remark about polarization: Let $(S,\omega,\mathcal{F})$ be a polarized symplectic manifold and we suppose the polarization $\mathcal{F}$ to be accessible. For open subsets $U\subset S$ we put ($k\in\mathbb{N}$)

$$A_{\mathcal{F}}(U) = \{f\in C^\infty(U);\ \xi\cdot f = 0 \text{ on } U \text{ if } \xi \text{ belongs to } \mathcal{F}\},$$

$$A^{(k)}(U) = \{f\in C^\infty(U);\ \{g_1,\{g_2,\ldots,\{g_{k+1},f\}\cdots\} = 0,$$

$$\text{for all } V\subset U \text{ open, } g_j\in A_{\mathcal{F}}(V)\}.$$

All these spaces can be considered over complex numbers as well as over reals. It holds that $A^{(1)}$ is a subsheaf of the sheaf of Poisson–Lie algebras on S and a real smooth function defined on U belongs to $A^{(1)}(U)$ if and only if the flow of the vector field ξ_f preserves the polarization.[12]

Let us consider, in addition, a symplectomorphism $J: S \to \mathfrak{g}^*$ of S onto a co-orbit $L = J(S)$ in $\mathfrak{g}^*$. The polarization is mapped onto a polarization $J_*(\mathcal{F})$ on L. The polarization $J_*(\mathcal{F})$ will be invariant with respect to the coadjoint action (supposing G to be connected) if and only if it is invariant with respect to flows of vector fields $\tilde\eta_X$, $\tilde\eta_X(F) = \mathrm{ad}^*X(F)$, $X\in\mathfrak{g}$. This is equivalent to the requirement $\lambda_X = J^*X\in A^{(1)}$, i.e., values of the homomorphism λ belong to the subalgebra $A^{(1)}$, $\lambda: \mathfrak{g} \to A^{(1)}$.

Summary: In the case of interest, i.e., for our Lie algebras $\mathfrak{g}$, $\mathfrak{q}$, we are given the following problem. Let $(\Omega,\mathcal{F})$ be a polarized coadjoint orbit in $\mathfrak{q}^*$. We look for a homomorphism $\lambda: \mathfrak{g} \to A^{(1)} = A^{(1)}(\Omega,\mathbb{R})$ such that the differential J_* of the corresponding momentum mapping is injective. Then J induces a local symplectomorphism between orbits in $\mathfrak{q}^*$ and $\mathfrak{g}^*$ that, moreover, maps the Q-invariant polarization onto a G-invariant polarization. Global aspects must be solved in the concrete case.

IV. THE POISSON ISOMORPHISM

Coadjoint orbits in $\mathfrak{g}^*=\mathfrak{g}$: We consider only orbits of maximal dimension containing regular elements. Any such orbit intersects $\mathfrak{h}$ exactly in $|\mathcal{W}|$ points. Let $F_1 = (1/2\pi i) \times (\nu + i\sigma)$, $\nu,\sigma\in\mathfrak{a}$. The orbit Ω_{F_1} satisfies the integrality condition if and only if $\nu\in\Lambda$. The orbits Ω_{F_1} with $\nu\in\mathfrak{a}$, $\sigma\in C_+$ fill up an open dense subset $\mathfrak{g}_0\subset\mathfrak{g}$ and are parametrized in a unique way by the set $\mathfrak{a}\times C_+$. On Ω_{F_1} there exist $|\mathcal{W}|$ real invariant polarizations and they are in one-to-one correspondence with Borel subalgebras containing $\mathfrak{h}$. All these polarizations lead to equivalent representations. In what follows we fix the subalgebra $\mathfrak{b}$. Let $\mathcal{O} = K/M, o = M\in\mathcal{O}$ be the origin. There exists a projection p_1: $\Omega_{F_1} \to \mathcal{O}$: $g\cdot F_1 \mapsto k\cdot o$, where $g = kan$ is the Iwasawa decomposition. The polarization satisfies the Pukanszky condition[18] since the mapping $N \to \mathfrak{n}$: $n \mapsto \mathrm{Ad}\, n(F_1) - F_1$ is a diffeomorphism.[13] Consequently, the fiber of the projection p_1 over the origin is $F_1 + \mathfrak{n}$.

Denote by χ_1: $B \to T^1$: $man \mapsto m^\nu a^{i\sigma}$ the character of the Borel subgroup, $\nu\in\Lambda$, $\sigma\in C_+$. Here we denote by $m^\nu = \exp(B(\nu,X))$, $m = \exp X$, $X\in\mathfrak{m}$, and $a^{i\sigma} = \exp(B(i\sigma,Y))$, $a = \exp Y$, $Y\in\mathfrak{a}$, the characters of M and A, respectively. The unitary representation corresponding to the orbit Ω_{F_1} is $\mathcal{T}_{\nu\sigma} = \mathrm{Ind}_B^G \chi_1$. The representations $\mathcal{T}_{\nu\sigma}$, $\nu\in\Lambda$, $\sigma\in C_+$, are irreducible and mutually nonequivalent, they belong to the principal series. The complement of the set $\Lambda\times C_+ \subset \hat G$ is of Plancherel measure zero. [The Plancherel measure is $\mu_1(\nu,\sigma)d\nu\,d\sigma$,

$$\mu_1(\nu,\sigma) = \prod_{\alpha>0} |\alpha(\nu + i\sigma)|^2,$$

$d\nu$ is the discrete counting measure on Λ, $d\sigma$ is the Lebesgue measure on C_+.]

Coadjoint orbits in $\mathfrak{q}^*=\mathfrak{q}$: A more general case was investigated by Lipsman.[14] Orbits Ω_{F_2} passing through points $F_2 = (1/2\pi i)(\nu,\sigma)$, $\nu\in\mathfrak{a}$, $\sigma\in C_+$, fill up an open dense subset $\mathfrak{q}_0 \subset \mathfrak{q}$ and are parametrized in a unique way by the set $\mathfrak{a}\times C_+$. The orbit Ω_{F_2} satisfies the integrality condition if and only if $\nu\in\Lambda$. The subalgebra $\mathfrak{m}\oplus\mathfrak{f}$ induces a real invariant polarization on Ω_{F_2}. Again, there exists a projection p_2: $\Omega_{F_2} \to \mathcal{O}$: $(k,X)\cdot F_2 \mapsto k\cdot o$. The polarization satisfies the Pukanszky condition since the linear operator $\mathrm{ad}((1/2\pi i)\sigma)$ on $\mathfrak{u}$ is bijective. Consequently, the fiber of the projection over the origin is $F_2 + \mathfrak{u}\times 0$.

Denote by

$$\chi_2\colon\ M\circledS\mathfrak{f}\to T^1\colon\ (m,Y)\mapsto m^\nu\exp(-B(\sigma,Y))$$

the character, $\nu\in\Lambda$, $\sigma\in C_+$. The unitary representation corresponding to the orbit Ω_{F_2} is $\mathscr{U}_{\nu\sigma}=\mathrm{Ind}_{M\circledS\mathfrak{f}}^{Q}\chi_2$. The representations $\mathscr{U}_{\nu\sigma}$, $\nu\in\Lambda$, $\sigma\in C_+$, are irreducible and mutually nonequivalent. The complement of the set $\Lambda\times C_+\subset\hat{Q}$ is of Plancherel measure zero. [The Plancherel measure is $\mu_2(\nu,\sigma)d\nu\,d\sigma$,

$$\mu_2(\nu,\sigma)=\prod_{\alpha>0}|\alpha(\sigma)|^2\,.]$$

Let $\gamma\colon\mathfrak{f}\to\mathfrak{X}(\mathscr{O})$ [resp. $\mathfrak{X}(T^*\mathscr{O})$] be the infinitesimal action of K on $\mathscr{O}$ (resp. $T^*\mathscr{O}$); $T^*\mathscr{O}$ is the cotangent bundle over $\mathscr{O}$. Using γ we identify tangent spaces $T_{k\cdot o}\mathscr{O}$ with $\mathfrak{f}/\mathrm{Ad}\,k(\mathfrak{m})$. Using the metric we identify cotangent spaces $T_{k\cdot o}{}^*\mathscr{O}$ with $\mathrm{Ad}\,k(u)$. There exists an unambiguously defined Q-equivariant diffeomorphism between the Q-homogeneous spaces Ω_{F_2} and $T^*\mathscr{O}$ such that F_2 is mapped on $0\in T_0{}^*\mathscr{O}$. Denote by $\rho_1\colon\mathscr{O}\to\mathscr{O}_{(1/2\pi i)\nu}$, $\rho_2\colon\mathscr{O}\to\mathscr{O}_{(1/2\pi i)\sigma}$, the K-equivariant mappings determined in a unique way by the condition $(\rho_1(o),\rho_2(o))=F_2$. Then K acts on $T^*\mathscr{O}$ by tangent mappings; $0\times\mathfrak{f}$ acts on the fiber $T_x{}^*\mathscr{O}$ as the group of translations by vectors $-\eta_Y(x):=\mathrm{ad}\,\rho_2(x)(Y)$, $Y\in\mathfrak{f}$.

Let $\tau\colon T^*\mathscr{O}\to\mathscr{O}$ be the projection, ϑ be the canonical form on $T^*\mathscr{O}$, $\vartheta_x(\zeta)=\langle x,\tau_*\zeta\rangle$. The K-orbit $\mathscr{O}_{(1/2\pi i)\nu}\subset\mathfrak{f}^*=\mathfrak{f}$ is integral; let ω^1 be the corresponding symplectic form. For the symplectic form ω on $T^*\mathscr{O}=\Omega_{F_2}$ the following expression is valid:

$$\omega=-d\vartheta+(\rho_1\circ\tau)^*\omega^1\,. \tag{1}$$

Both sides are Q-invariant, hence we can check the equality in one point only. Since $\mathscr{O}$ is embedded into $T^*\mathscr{O}$ as the null section and since $\tau_*\eta_Y=0$ and integral curves of the vector field γ_X do not leave the submanifold $\mathscr{O}\subset T^*\mathscr{O}$, $\vartheta|_{\mathscr{O}}=0$, we have (all expressions are evaluated at the origin, ζ is arbitrary)

$$(\rho_1\circ\tau)^*\omega^1(\zeta,\eta_Y)=d\vartheta(\gamma_X,\gamma_Y)=d\vartheta(\eta_X,\eta_Y)=0\,,$$

$$\omega(\gamma_X,\gamma_Y)=-\langle F_2,([X,Y],0)\rangle=-\langle(1/2\pi i)\nu,[X,Y]\rangle$$
$$=(\rho_1\circ\tau)^*\omega^1(\gamma_X,\gamma_Y)\,,$$

$$\omega(\gamma_X,\eta_Y)=-\langle F_2,(0,[X,Y])\rangle$$
$$=\langle(1/2\pi i)\sigma,[X,Y]\rangle$$
$$=-\langle\mathrm{ad}((1/2\pi i)\sigma)(Y),X\rangle$$
$$=\eta_Y\cdot\vartheta(\gamma_X)=-d\vartheta(\gamma_X,\eta_Y)\,.$$

The real polarization on $T^*\mathscr{O}$ is a foliation whose leaves are exactly fibers of the projection τ. The subalgebra $A^{(1)}$ of the Poisson algebra $C^\infty(T^*\mathscr{O})$ consists of those functions that, when restricted to $T_x{}^*\mathscr{O}$, are polynomials of at most first order for all $x\in\mathscr{O}$. We can identify

$$A^{(1)}=C^\infty(\mathscr{O})\oplus\mathfrak{X}(\mathscr{O}) \tag{2}$$

using the embeddings

$$C^\infty(\mathscr{O})\ni f\mapsto\tau^*f\in C^\infty(T^*\mathscr{O})\,,$$

$$\mathfrak{X}(\mathscr{O})\ni\zeta\mapsto\varphi^\zeta(x)=\langle x,\zeta_{\tau(x)}\rangle\in C^\infty(T^*\mathscr{O})\,.$$

The Poisson bracket then reads

$$\{f_1,f_2\}=0,\quad\{\zeta,f\}=\zeta\cdot f, \tag{3a}$$

$$\{\zeta_1,\zeta_2\}=(\rho_1{}^*\omega^1)(\zeta_1,\zeta_2)+[\zeta_1,\zeta_2]\,. \tag{3b}$$

Let $\omega^{(c)}=-d\vartheta$ be the standard symplectic structure on $T^*\mathscr{O}$; the corresponding Poisson algebra is well known.[15] Denote by ξ (resp. $\xi^{(c)}$): $C^\infty(T^*\mathscr{O})\to\mathfrak{X}(T^*\mathscr{O})$ the Lie algebra homomorphism corresponding to ω (resp. $\omega^{(c)}$). Then (3a) follows from the equality $\xi_f=\xi_f{}^{(c)}$ for $f\in A^{(0)}=C^\infty(\mathscr{O})$. In general, it holds that $\tau_*\xi_g=\tau_*\xi_g{}^{(c)}$ for all $g\in C^\infty(T^*\mathscr{O})$ since

$$(\xi_g{}^{(c)}-\xi_g)\,\lrcorner\,d\vartheta+\xi_g\,\lrcorner\,(\rho_1\circ\tau)^*\omega^1=dg-dg=0$$

and hence for $f\in C^\infty(\mathscr{O})$ we have $\tau_*\xi_f=0$,

$$0=-d\vartheta(\xi_f,\xi_g{}^{(c)}-\xi_g)=\langle df,\tau_*(\xi_g{}^{(c)}-\xi_g)\rangle\,.$$

Let $\zeta_j\in\mathfrak{X}(\mathscr{O})$, $\varphi_j=\varphi^{\zeta_j}$, $\xi_j=\xi_{\varphi_j}$, $j=1,2$. Since $\tau_*(\xi_{\varphi^\zeta})=\zeta$, $\vartheta(\xi_{\varphi^\zeta})=\varphi^\zeta$, we have

$$-\{\varphi_1,\varphi_2\}=-d\vartheta(\xi_1,\xi_2)+(\rho_1\circ\tau)^*\omega^1(\xi_1,\xi_2)$$
$$=-\xi_1\cdot\varphi_2+\xi_2\cdot\varphi_1+\varphi^{[\zeta_1,\zeta_2]}$$
$$+(\rho_1{}^*\omega^1)(\zeta_1,\zeta_2)\,.$$

Using $\xi_1\cdot\varphi_2=-\xi_2\cdot\varphi_1=\{\varphi_1,\varphi_2\}$ we obtain (3b).

Let $\tilde{\lambda}$: $\mathfrak{q}\to A^{(1)}$, λ^1: $\mathfrak{f}\to C^\infty(\mathscr{O}_{(1/2\pi i)\nu})$, λ^2: $\mathfrak{f}\to C^\infty(\mathscr{O}_{(1/2\pi i)\sigma})$ be the canonical Poisson actions. Then ($X\in\mathfrak{f}$)

$$\tilde{\lambda}_{(X,0)}=\rho_1{}^*\lambda^1_X+\gamma_X,\quad\tilde{\lambda}_{(0,X)}=-\rho_2{}^*\lambda^2_X\,. \tag{4}$$

We are now going to describe all homomorphisms $\lambda\colon\mathfrak{g}\to A^{(1)}$ that fulfill the *a priori* assumption (respecting the identification of $K\subset G$ and $K\subset Q$)

$$\lambda_X=\rho_1{}^*\lambda^1_X+\gamma_X,\quad X\in\mathfrak{f}\,. \tag{5}$$

We put $\lambda_{iX}=\tilde{f}_X+\zeta_X$, $X\in\mathfrak{f}$, $\tilde{f}_X\in C^\infty(\mathscr{O})$, $\zeta_X\in\mathfrak{X}(\mathscr{O})$. We require ($X,Y\in\mathfrak{f}$)

$$\{\lambda_X,\lambda_{iY}\}=\lambda_{i[X,Y]}\,, \tag{6a}$$

$$\{\lambda_{iX},\lambda_{iY}\}=-\lambda_{[X,Y]}\,. \tag{6b}$$

It follows that

$$[\gamma_X,\zeta_Y]=\zeta_{[X,Y]}\,, \tag{7a}$$

$$[\zeta_X,\zeta_Y]=-\gamma_{[X,Y]}\,. \tag{7b}$$

Equation (7a) implies $\zeta_X(k\cdot x)=k_*\zeta_{\mathrm{Ad}\,k^{-1}(X)}(x)$, $k\in K$. Since K acts transitively on $\mathscr{O}$, it is sufficient to know the value of ζ_X only in the origin. Using the decomposition we identify $T_0\mathscr{O}=\mathfrak{f}/\mathfrak{m}\equiv\mathfrak{u}$. We define a mapping $I\colon\mathfrak{u}\to\mathfrak{u}$ by

$$\zeta_X(o)=\gamma_{IX}(o)\,. \tag{8}$$

This definition is consistent if

$$(\mathrm{Ad}\,m)\circ I=I\circ(\mathrm{Ad}\,m),\quad m\in M\,, \tag{9a}$$

or equivalently

$$(\mathrm{ad}\,H)\circ I=I\circ(\mathrm{ad}\,H),\quad H\in\mathfrak{m}\,. \tag{9b}$$

Further, we define I on $\mathfrak{m}$ trivially, $I|_{\mathfrak{m}}=0$, and put, for $x=k\cdot o$, $k\in K$,

$$I(x)=(\mathrm{Ad}\,k)\circ I\circ(\mathrm{Ad}\,k^{-1})\,. \tag{10}$$

Then $\zeta_X(x)=\gamma_{I(x)X}(x)$. Since $\gamma_X.I(x)=-[\mathrm{ad}\,X,I(x)]$, (7b) means that

$$[I(x)X,I(x)Y] - [\mathrm{ad}(I(x)X),I(x)]\cdot Y + [\mathrm{ad}(I(x)Y),I(x)]\cdot X + [X,Y] \in \mathrm{Ad}\, k(\mathfrak{m}),$$
$$x = k\cdot o .$$

Regarding the K equivariance, this is equivalent to

$$[IX,IY] - [X,Y] - I[X,IY] - I[IX,Y] \in \mathfrak{m} . \quad (11)$$

If $H\in\mathfrak{m}$ then $IH = 0$ and hence $(\mathrm{ad}\, H)\circ(1 + I^2) = 0$. If H is regular then ad H restricted to $\mathfrak{u}$ is an isomorphism and we have

$$I^2|_{\mathfrak{u}} = -1 . \quad (12)$$

Linear mappings I with properties (9a) and (11) are in one-to-one correspondence with invariant complex structures $\tilde{I}$ on $\mathcal{O}$.[16] By linearity we define I on $\mathfrak{g} = \mathfrak{k}\otimes\mathbb{C}$. There exist $|\mathcal{W}|$ different K-invariant complex structures on $\mathcal{O}$ that are in one-to-one-correspondence with Borel subalgebras in $\mathfrak{g}$ containing $\mathfrak{h}$, $I \mapsto (1 + iI)\mathfrak{g}$. Moreover, it holds that the eigenspaces for the eigenvalues $+i$, $-i$ are maximal nilpotent subalgebras in $\mathfrak{g}$, the left-hand side in (11) is equal to zero for all X, $Y\in\mathfrak{g}$ and the mapping I is skew-Hermitian with respect to the bilinear form $\langle\cdot,\cdot\rangle$. All these assertions follow immediately from the fact that the root spaces $\mathfrak{g}_\alpha$, $\alpha\in\Delta$, are one dimensional and hence, according to (9b) and (12), H_α, E_α are the eigenvectors of I, where the only possible eigenvalues are $0, +i, -i$.

We conclude that $(X\in\mathfrak{k})$

$$\lambda_{iX}(x) = \lambda_{I(x)X}(x) + f_X(x), \quad f_X\in C^\infty(\mathcal{O}) . \quad (13)$$

It remains to determine f_X. Inserting into (6a) and (6b) we have $\gamma_X\cdot f_Y = f_{[X,Y]}$ and $\xi_X\cdot f_Y - \xi_Y\cdot f_X = 0$. The first equality means $f_X(k\cdot x) = f_{\mathrm{Ad}\, k^{-1}(X)}(x)$, i.e., f_X is determined by the value $f_X(o) = \langle(1/2\pi i)\epsilon, X\rangle$, where $\epsilon\in i\mathfrak{k}$ is unambiguously defined. This definition is consistent if $\mathrm{Ad}\, m(\epsilon) = \epsilon$ for all $m\in M$, i.e., $\epsilon\in\mathfrak{a}$. The second condition means $\langle i\epsilon,[IX,Y] - [IY,X]\rangle = 0$, but from the above-mentioned properties of I it follows that $[IX,Y] - [IY,X]\in\mathfrak{h}^\perp$ for all X, $Y\in\mathfrak{g}$.

Let J be the momentum mapping corresponding to λ. Then $2\pi i J(o) = \nu - i\epsilon$. By differentiating the relation $\langle J(x), X + iY\rangle = \lambda_{X+iY}$ at the origin we obtain

$$2\pi i J_*(o)(\gamma_U + \eta_V) = (\mathrm{ad}\,\nu - i\,\mathrm{ad}\,\epsilon)U - (1 + iI)\circ\mathrm{ad}\,\sigma(V) , \quad (14)$$

where U, $V\in\mathfrak{u}$. Since ad $\sigma|_{\mathfrak{u}}$ is injective, $\ker J_*(o)$ is isomorphic to $\ker(\mathrm{ad}\,\nu - I\circ\mathrm{ad}\,\epsilon)$. The mapping $(\mathrm{ad}\,\nu - I\circ\mathrm{ad}\,\epsilon)$ defined by linearity on $\mathfrak{h}^\perp = \mathfrak{u}\otimes C$ has eigenvectors E_α and eigenvalues $\alpha(\nu) \mp i\alpha(\epsilon)$, $\alpha\in\Delta$. Hence $J_*(o)$ is injective if and only if $(\nu - i\epsilon)$ is regular.

Since $J_*(o)(T_0{}^*\mathcal{O}) = (1 + iI)\mathfrak{u}$, the polarization on $T^*\mathcal{O}$ is mapped onto the polarization on $\Omega_{F_1'}$, $F_1' = (1/2\pi i)(\nu - i\epsilon)$, which corresponds to the Borel subalgebra $(1 + iI)\mathfrak{g}$. In what follows we fix the Borel subalgebra $\mathfrak{b}$. Then $\mathfrak{n}$, $\mathfrak{n}_-$ are the eigenspaces for the eigenvalues $-i$, i. If $H\in C_+$, then $I = -i\,\mathrm{sgn}(\mathrm{ad}\, H)$. In general, we put for $Y\in\mathfrak{k}$, Y regular,

$$I_Y = -i\,\mathrm{sgn}(i\,\mathrm{ad}\, Y) . \quad (15)$$

The just chosen complex structure is distinguished by the fact that it converts, together with the standard symplectic structure, the orbit $\mathcal{O} = \mathcal{O}_{(1/2\pi i)\sigma}$ into a Kählerian manifold, i.e., the positivity condition $\langle(1/2\pi i)\sigma,[IX,X]\rangle > 0$, $X\in\mathfrak{u}$ nonzero, is fulfilled. Moreover, this complex structure coincides with the complex structure gained by the Iwasawa decomposition $\mathcal{O} = K/M = G/B$.

We use again Ω_{F_1} instead of $T^*\mathcal{O}$ and get the following expression for the momentum mapping:

$$J(X,Y) = X + i(-f(Y) + I_Y X) , \quad (16)$$

where f: $\mathcal{O}_{(1/2\pi i)\sigma}\to\mathcal{O}_{(1/2\pi i)\epsilon}$ is the K-equivariant mapping determined by $f((1/2\pi i)\sigma) = (1/2\pi i)\epsilon$. Here J maps Ω_{F_2} into $\Omega_{F_1'}$, $F_1' = (1/2\pi i)(\nu - i\epsilon)$. Both orbits are fibered spaces over the same basis—the K-homogeneous space $\mathcal{O} = K/M$. This J is K-equivariant, preserves the fibers and maps them one onto another diffeomorphically (since $1 + iI$: $\mathfrak{u}\to\mathfrak{n}$ is an isomorphism), and projects onto the basis as the identity. Hence J is a symplectomorphism that, moreover, preserves the polarization.

In what follows we choose $\epsilon = -\sigma$. The reasons are the following: (1) the simplicity, (2) the character formula stated in Sec. VI, and (3) with this choice,

$$J(X,Y) = X + i(Y + I_Y X) . \quad (17)$$

Clearly, $J(X,Y) = X + iY$, $X\in\mathfrak{a}$, $Y\in C_+$; hence J is a diffeomorphism mapping $\mathfrak{q}_0$ onto $\mathfrak{g}_0$. The transversal tangent Lie algebra in the case of Lie–Poisson structures coincides with the Lie algebra of the isotropy group. We shall show that $J^* = (dJ_{F_2})^*$ induces the previously fixed isomorphism between $\mathfrak{h}\subset\mathfrak{g}^{**}$ and $\mathfrak{m}\times\mathfrak{m}\subset\mathfrak{q}^{**}$. It holds that

$$(J_*)_{(X,Y)}(U,V) = (1 + iI)U + iV - i(\mathrm{ad}\, X)\circ(\mathrm{ad}\, Y)^{-1}\circ I(V) , \quad (18)$$

where $X\in\mathfrak{m}$, $Y\in\mathfrak{m}_+$, and hence $(\mathrm{ad}\, Y)^{-1}$: $\mathfrak{u}\to\mathfrak{u}$ is well defined. This equality can be obtained from (14) using $\gamma_U(X,Y) = (\mathrm{ad}\, X(U), \mathrm{ad}\, Y(U))$, $\eta_V(X,Y) = (-\mathrm{ad}\, Y(V), 0)$. Another calculation of the differential $\partial I_{Y+V}/\partial V$ will be given in Sec. VI. From (18) it follows that $\langle U + iV, J_*(Z,W)\rangle = \langle U,Z\rangle - \langle V,W\rangle$, for $U,V\in\mathfrak{m}$, and hence $J^*(U + iV) = (U,V)$.

Theorem 4.1: The mapping

$$J: \mathfrak{q}^* = \mathfrak{q}\to\mathfrak{g}^* = \mathfrak{g}: (X,Y)\mapsto X + i(Y + I_Y X),$$
$$I_Y = -i\,\mathrm{sgn}(i\,\mathrm{ad}\, Y) ,$$

is defined and smooth on the open dense subset $\mathfrak{q}_0\subset\mathfrak{q}$ and maps $\mathfrak{q}_0$ diffeomorphically onto the open dense subset $\mathfrak{g}_0\subset\mathfrak{g}$. Then J is a Poisson mapping that, moreover, maps the real Q-invariant polarization on the orbits $\Omega_{F_2}\subset\mathfrak{q}_0$ corresponding to the subalgebra $\mathfrak{m}\oplus\mathfrak{k}$ onto the real G-invariant polarization on the orbits $\Omega_{F_1}\subset\mathfrak{g}_0$ corresponding to the Borel subalgebra $\mathfrak{b}$. Clearly, $J(F_2) = F_1$, $F_1 = (1/2\pi i)(\nu + i\sigma)$, $F_2 = (1/2\pi i)(\nu,\sigma)$, $\nu\in\mathfrak{a}$, $\sigma\in C_+$.

V. THE GELL-MANN FORMULA

Let $\nu\in\Lambda$, $\sigma\in C_+$, and $\mathcal{T}$, $\mathcal{U}$ be the differentiated representations $\mathcal{T}_{\nu\sigma}$, $\mathcal{U}_{\nu\sigma}$. Hilbert spaces for both representations can be identified with help of the momentum mapping J. Then $\mathcal{O} = K/M$ is connected and simply connected and hence there exists a unique up to isomorphism complex line

bundle with connection $(\mathscr{L}_\nu,\nabla)$ over $\mathscr{O}$ such that the curvature satisfies curv $\nabla = \rho_1^*\omega^1$. There exists a unique up to a multiplicative constant Hermitian metric in $(\mathscr{L}_\nu,\nabla)$. We choose on the manifold $\mathscr{O}$ a smooth measure μ from the Lebesgue measure class, the standard choice is the measure induced by the Haar measure on K normed to unity. Then the underlying Hilbert space for both representations is $\mathscr{H}_\nu = L^2(\mathscr{L}_\nu,d\mu)$ consisting of measurable sections in $\mathscr{L}_\nu$ with finite L^2-norms. The method of geometric quantization and the results of the preceding section enable us to write down the formulas ($X\in\mathfrak{k}$, $\tilde{I}$ is the complex structure on $\mathscr{O}$)

$$\mathscr{T}(X) = \mathscr{U}(X,0) = \nabla(\gamma_X) + \tfrac{1}{2}\mathrm{div}_\mu\,\gamma_X + 2\pi i(\rho_1^*\lambda^1_X)\,, \tag{19a}$$

$$\mathscr{T}(iX) = \nabla(\tilde{I}\gamma_X) + \tfrac{1}{2}\mathrm{div}_\mu(\tilde{I}\gamma_X) - 2\pi i(\rho_2^*\lambda^2_X)\,, \tag{19b}$$

$$\mathscr{U}(0,X) = -2\pi i(\rho_2^*\lambda^2_X)\,. \tag{19c}$$

Choose $\{L_1,...,L_n\}$ to be a basis in $\mathfrak{k}$ and put $P_j = iL_j$. Then $\{L_j,P_j\}$ is a basis in $\mathfrak{g}$. We identify $L_j = (L_j,0)$ and put $T_j = (0,L_j)$. Then $\{L_j,T_j\}$ is a basis in $\mathfrak{q}$. Let $\{L^1,...,L^n\}$ be the basis in $\mathfrak{k}$ determined by $\langle L^j,L_k\rangle = -\delta^j_k$. The operators $(1/2\pi i)\mathscr{U}(T_1),...,(1/2\pi i)\mathscr{U}(T_n)$ mutually commute and their common spectrum consists exactly of those points $t = (t_1,...,t_n)\in\mathbb{R}^n$ such that $t\cdot L = t_1L^1 + \cdots + t_nL^n \in\mathscr{O}_{(1/2\pi i)\sigma}$. Denote by $(I(t)^k_j)$ the matrix function defined by

$$I_{t\cdot L}L_j = I(t)^k_jL_k\,. \tag{20}$$

The representations $\mathscr{T},\mathscr{U}$ are related by ($j = 1,2,...,n$)

$$\mathscr{T}(L_j) = \mathscr{U}(L_j), \tag{21a}$$

$$\mathscr{T}(P_j) = \mathscr{U}(T_j) + \tfrac{1}{2}\{I((1/2\pi i)\mathscr{U}(T_1),...,(1/2\pi i)\times\mathscr{U}(T_n))^k_j,\mathscr{U}(L_k)\}_+\,. \tag{21b}$$

Equation (21b) follows from the equalities

$$\nabla(\tilde{I}\gamma_{L_j}) = I((1/2\pi i)\mathscr{U}(T))^k_j\nabla(\gamma_{L_k})$$

and

$$\nabla(\zeta) + \tfrac{1}{2}\mathrm{div}_\mu\,\zeta = \tfrac{1}{2}(\nabla(\zeta) - \nabla(\zeta)^+),\quad \zeta\in\mathfrak{X}(\mathscr{O})\,.$$

We shall prove

$$I_X := -i\,\mathrm{sgn}(i\,\mathrm{ad}\,X) = \sum_{s=1}^{r}(-1)^{s+1}a_s(\mathrm{ad}\,X)^{2s-1}\,, \tag{22}$$

where $X\in\mathfrak{k}$ is regular, $r = |\Delta_+| = \frac{1}{2}(\dim\mathfrak{k} - \mathrm{rank}\,\mathfrak{k})$ and the a_s are K-invariant smooth functions defined on the open dense subset of regular elements $\mathfrak{k}_r\subset\mathfrak{k}$. Hence for a given representation, the a_s are constants determined by the value on $\mathscr{O}_{(1/2\pi i)\sigma}$. Regarding the K invariance, it is sufficient to check the equality (22) for $X = iH$, $H\in\mathrm{C}_+$. Both sides are thought to be defined by linearity on $\mathfrak{g} = \mathfrak{k}\otimes\mathbb{C}$, both sides vanish on $\mathfrak{h}$ and, acting on the root vectors E_α, $\alpha\in\Delta$, we have

$$\mathrm{sgn}\,\alpha(H) = \sum_{s=1}^{r}a_s\alpha(H)^{2s-1}\,.$$

Clearly, it is sufficient to check only $\alpha\in\Delta_+$. Put $\beta_j = \alpha_j(H) > 0$, where $\Delta_+ = \{\alpha_1,...,\alpha_r\}$. Then the numbers $a_1,...,a_r$ are required to solve the system of linear equations with matrix $C = (c_{jk})$, $c_{jk} = \beta_j^{2(k-1)}$, and the right-hand side $(1/\beta_1,...,1/\beta_r)$. The matrix C is regular if and only if $\beta_j \neq \beta_k$ for $j\neq k$. This condition is satisfied for almost all $H\in C_+$. Since the system is invariant with respect to the permutations of $\beta_1,...,\beta_r$ and from the Cramer rule, it follows that

$$a_s(\beta_1,...,\beta_r) = \left[\prod_j\beta_j\prod_{j<k}(\beta_j+\beta_k)\right]^{-1}p_s(\beta_1,...,\beta_r)\,,$$

where p_s is a symmetric polynomial of order $\frac{1}{2}r(r+1) + 1 - 2s$. Regarding the continuity, the a_s are well defined for all $H\in C_+$, with the condition $\beta_j\neq\beta_k$, $j\neq k$, omitted.

It follows from (22) that the functions $I(t)^k_j$ coincide on $\mathscr{O}_{(1/2\pi i)\sigma}$ with some polynomials. Let $A(t)$ be the matrix of the linear operator $\mathrm{ad}(t\cdot L)$ expressed in the basis $\{L_1,...,L_n\}$. Then

$$I(t)^k_j = \sum_{s=1}^{r}(-1)^{s+1}a_sA(t)^{2s-1}\,. \tag{23}$$

Inserting (23) into (21b) we obtain a polynomial dependence of the generators $\mathscr{T}(L_j)$, $\mathscr{T}(P_j)$ on $\mathscr{U}(L_j)$, $\mathscr{U}(T_j)$, i.e., a generalized Gell-Mann formula. The Appendix is devoted to some concrete calculations for $K = \mathrm{SU}(n)$.

VI. CHARACTERS

Let $\nu\in\Lambda$, $\sigma\in C_+$, $F_1 = (1/2\pi i)(\nu + i\sigma)\in\mathfrak{g}$, $F_2 = (1/2\pi i)(\nu,\sigma)\in\mathfrak{q}$. Both representations $\mathscr{T}_{\nu\sigma}$, $\mathscr{U}_{\nu\sigma}$ fulfill the Kirillov character formula

$$\pi(\varphi) = \int_D\varphi(X)\pi(\exp X)\,j(X)dX,\quad \varphi\in\mathscr{D}(D)\,,$$

$$tr\,\pi(\varphi) = \int_\Omega\left\{\int_D j(X)^{1/2}\varphi(X)e^{2\pi i\langle F,X\rangle}dX\right\}d\beta(F)\,,$$

where $\pi = \mathscr{T}_{\nu\sigma}$ (resp. $\mathscr{U}_{\nu\sigma}$), D is the set containing those $X\in\mathfrak{g}$ [resp. those $(X,Y)\in\mathfrak{q}$] for which eigenvalues λ of the operator ad X satisfy $|\mathrm{Im}\,\lambda| < \pi$, $\mathscr{D}(D)$ is the test-function space, $j(X)$ is the Haar measure density in canonical coordinates normed by $j(0) = 1$, $j(X) = \det(\sinh(\mathrm{ad}\,X/2)/(\mathrm{ad}\,X/2))$, $\Omega = \Omega_{F_1}$ (resp. Ω_{F_2}), $d\beta = (d!)^{-1}\times|\omega\Lambda\cdots\Lambda\omega|$ is the canonical measure on Ω, $d = \frac{1}{2}\dim\Omega = n - l$. The formula was proved for semisimple Lie groups by Duflo and Gutkin,[17] its validity for Q follows from Kirillov's reduction theorem.[18]

We specify the normalization of Lebesgue measures on some subalgebras. The decomposition $\mathfrak{g} = \mathfrak{n}_- + \mathfrak{h} + \mathfrak{n}$ induces the Fubini decomposition of the measure on $\mathfrak{g}$, $dZ = dX_-dH\,dX_+$. The measures dX_-, dX_+ are normed by the conditions

$$\int_{N_-}a_u^{-4\delta}\,du = 1,\quad u = k_ua_un_u,\quad du = d(\exp X_-)\,,$$

$$\int_{\mathfrak{n}_-}\left\{\int_{\mathfrak{n}}\varphi(X)e^{2\pi i\langle Y,X\rangle}\,dX\right\}dY = \varphi(0),\quad \varphi\in\mathscr{D}(\mathfrak{n})\,.$$

The decomposition $\mathfrak{q} = (\mathfrak{m}+\mathfrak{u})\times(\mathfrak{m}+\mathfrak{u})$ induces the Fubini decomposition on $\mathfrak{q}$, $dZ = dX\,dY = dS\,d_1U\,dT\,d_2V$. We require the isomorphism $\mathfrak{m}\times\mathfrak{m}\to\mathfrak{h}$: $S,T\mapsto H = S + iT$ to preserve the measure $dH = dS\,dT$, the isomorphism $\mathfrak{u}\to\mathfrak{n}$: $V\mapsto X = (1+iI)V$ to preserve the measure $dX = d_2V$. The measure d_1U is normed by

$$\int_{\mathfrak{u}}\left\{\int_{\mathfrak{u}}\varphi(U)e^{2\pi i\langle U,V\rangle}\,d_1U\right\}d_2V = \varphi(0),\quad \varphi\in\mathscr{D}(\mathfrak{u})\,.$$

By a simple calculation it can be shown that the isomorphism $\mathfrak{u}\to\mathfrak{n}_-: U\mapsto Y=\frac{1}{2}(1-iI)U$ preserves the measure, $dY=d_1U$.

In what follows we shall need the following result.[19] The mapping $u,X\mapsto \mathrm{Ad}\,u(F_1+X)$ maps $N_-\times\mathfrak{n}$ diffeomorphically onto a subset of full measure in Ω_{F_1}, the canonical measure in these coordinates reads $d\beta=du\,dX$.

Here Ω_{F_1} is the fibered space, the fiber over F_1 is $F_1+\mathfrak{n}$. On this fiber we choose the Lebesgue measure dF induced by the measure on $\mathfrak{n}$ and we transfer this measure on the other fibers using mappings Ad k, $k\in K$. The Gauss decomposition $N_-HN\subset G$ induces the embedding of N_- onto a subset of full measure in $\mathcal{O}=G/B$. The standard measure on $\mathcal{O}$ is then $a^{-4\delta}\,du$, $u=kan\in N_-$. Denote by $\Omega_{(x)}$ the fiber over $x\in\mathcal{O}$; then

$$\int_\Omega f(F)d\beta(F)=\int_{\mathcal{O}}\left\{\int_{\Omega_{(x)}}f(F)dF\right\}a^{-4\delta}\,du\,. \tag{24}$$

In the case of the orbit Ω_{F_2} we can exploit the symplectomorphism J from Sec. IV. We obtain parametrization $N_-\times\mathfrak{u}\to\Omega_{F_2}$: $u,U\mapsto \mathrm{Ad}\,k_u(F_2+(U,0))$, $u=k_ua_un_u$. The measure on Ω_{F_2} now reads $d\beta=a_u{}^{-4\delta}\,du\,d_2U$.

The value of tr $\pi(\varphi)$ does not change if we replace φ by φ_K,

$$\varphi_K(x)=\int_K\varphi(\mathrm{Ad}\,k(x))dk; \tag{25}$$

the Haar measure on K is normed to unity. Supposing φ to be K-invariant we can further simplify the expression in Kirillov's formula since we integrate over Ω a K-invariant function in variable F. Hence the integral over Ω can be reduced to the integral over one fiber only and the inverse Fourier transform can be used. In the case of $\pi=\mathscr{T}_{\nu\sigma}$, using (24) we have

$$\mathrm{tr}\,\mathscr{T}_{\nu\sigma}(\varphi)=\int_{\mathfrak{b}+\mathfrak{n}}j_{\mathfrak{g}}(Z)^{1/2}\varphi(Z)e^{2\pi i\langle F_1,H\rangle}\,dH\,dX\,,\qquad Z=H+X\,. \tag{26}$$

A further simplification is possible according to the following lemma.

Lemma: Let $\psi\in\mathscr{D}(\mathfrak{g})$, ψ be K-invariant, $C\subset\mathfrak{a}$ be a Weyl chamber, $w\in\mathscr{W}$, $F\in\mathfrak{h}$. Put $\mathfrak{h}_C=\mathfrak{a}+iC\subset\mathfrak{h}$; then

$$\int_{\mathfrak{h}_C+\mathfrak{n}}\psi(Z)e^{i\langle F,Z\rangle}\,dZ=\int_{\mathfrak{h}_{w\cdot C}+\mathfrak{n}}\psi(Z)e^{i\langle w\cdot F,Z\rangle}dZ\,. \tag{27}$$

Proof: It is sufficient to check (27) only for reflections w_j corresponding to simple roots $\alpha_1,\dots,\alpha_l$, since $w_1,\dots,w_l$ generate $\mathscr{W}$. Let α be a simple root, $w=w_\alpha$, $\mathfrak{h}_r\subset\mathfrak{h}$ be the subset of regular elements. Put $\mathfrak{n}_\alpha=\Sigma_{\beta>0,\beta\neq\alpha}\mathbb{C}E_\beta$. Then $\mathfrak{n}=\mathbb{C}E_\alpha+\mathfrak{n}_\alpha$ and $\mathfrak{n}_\alpha$ is invariant with respect to ad E_α, ad $E_{-\alpha}$. We define a function

$$\mathfrak{h}_r\times\mathbb{C}^x\to K:\ H,a\mapsto k(H,a)=\exp(zE_\alpha+\bar{z}E_{-\alpha}),\qquad \mathbb{C}^x=\mathbb{C}\setminus\{0\}\,,$$

where

$$z=-\frac{\overline{\alpha(H)a}}{|\alpha(H)a|}\sqrt{\frac{2}{\langle\alpha,\alpha\rangle}}\times\arctan\left(\sqrt{\frac{2}{\langle\alpha,\alpha\rangle}}\left|\frac{\alpha(H)}{a}\right|\right).$$

By a straightforward calculation it can be verified that for $H\in\mathfrak{h}_r$, $a\in\mathbb{C}^x$, $X\in\mathfrak{n}_\alpha$, it holds that

$$\mathrm{Ad}\,k(H,a)(H+aE_\alpha+X)=w\cdot H+aE_\alpha+\mathscr{U}(H,a)X\,,$$

where $\mathscr{U}(H,a)$ is a nonspecified isomorphism $\mathfrak{n}_\alpha\to\mathfrak{n}_\alpha$ depending smoothly on H and a. Since the operators Ad k,$k\in K$, preserve the metric $(X,Y)=\mathrm{Re}\,B(\bar{X},Y)$ on $\mathfrak{g}$, the isomorphism $\mathscr{U}(H,a)$ preserves the Lebesgue measure on $\mathfrak{n}_\alpha$. In this way we get the diffeomorphism

$$\mathfrak{h}_C+\mathbb{C}^xE_\alpha+\mathfrak{n}_\alpha\to\mathfrak{h}_{w\cdot C}+\mathbb{C}^xE_\alpha+\mathfrak{n}_\alpha:\ H+aE_\alpha+X\mapsto \mathrm{Ad}\,k(H,a)(H+aE_\alpha+X)\,,$$

which preserves the Lebesgue measure. Now, to prove (27) it is enough to use this special substitution in the integral.

Noting that the Weyl group acts simply transitively on the Weyl chambers and using the above lemma we obtain the final expression

$$\mathrm{tr}\,\mathscr{T}_{\nu\sigma}(\varphi)=\sum_{w\in\mathscr{W}}\int_{\mathfrak{b}_+\mathfrak{n}}j_{\mathfrak{g}}(Z)^{1/2}\times\varphi(Z)e^{2\pi i\langle w\cdot F_1,H\rangle}\,dH\,dX,\quad Z=H+X\,. \tag{28}$$

A similar and even simpler procedure can be used in the case of $\pi=\mathscr{U}_{\nu\sigma}$. The integration over Ω_{F_2} is reduced to the integration over one fiber only, the fiber over F_2 is $F_2+\mathfrak{u}\times 0$. Since the action on $\mathfrak{m}$ of each $w\in\mathscr{W}$ can be realized as Ad $k|_{\mathfrak{m}}$ for some $k\in K$, we have directly

$$\mathrm{tr}\,\mathscr{U}_{\nu\sigma}(\varphi)=\sum_{w\in\mathscr{W}}\int_{\mathfrak{m}_+\times\mathfrak{k}}j_{\mathfrak{q}}(S)^{1/2}\varphi(S,T+V)\times e^{2\pi i\langle w\cdot F_2,(S,T)\rangle}dS\,dT\,d_2V\,. \tag{29}$$

We note that $j_{\mathfrak{q}}(X,Y)=j_{\mathfrak{k}}(X)^2$.

Theorem 6.1: The mapping

$$\Phi:\mathfrak{k}\times\mathfrak{k}\to\mathfrak{g}:\ X,Y\mapsto X-I_XY+iY$$

is defined and smooth on the open dense subset $\mathfrak{k}_r\times\mathfrak{k}\subset\mathfrak{q}$ ($\mathfrak{k}_r\subset\mathfrak{k}$ is the subset of regular elements) and maps it diffeomorphically onto the open dense subset $i\mathfrak{g}_0\subset\mathfrak{g}$. Let $\nu\in\Lambda$, $\sigma\in C_+$, $\varphi\in\mathscr{D}(i\mathfrak{g}_0)$; then

$$\mathrm{tr}\,\mathscr{T}_{\nu\sigma}(\varphi)=\mathrm{tr}\,\mathscr{U}_{\nu\sigma}((\Phi^*j_{\mathfrak{g}}/j_{\mathfrak{q}})^{1/2}\Phi^*\varphi)\,. \tag{30}$$

Remark: If $X,Y\in\mathfrak{k}_r\times\mathfrak{k}$, $Z=\Phi(X,Y)$, then the imaginary parts of eigenvalues of the operator ad Z in $\mathfrak{g}$ coincide with the eigenvalues of the operator ad X in $\mathfrak{k}_{\mathbb{C}}=\mathfrak{g}$. To show this, regarding the K invariance, we can confine ourselves to $X=S\in\mathfrak{m}_+$. Putting $Y=T+V\in\mathfrak{m}+\mathfrak{u}$, we have $Z=H+W\in\mathfrak{h}+\mathfrak{n}$, where $H=S+iT$, $W=i(1+iI)V$. Since H is regular there exists $u\in N$ such that Ad $u(H)=Z$. Hence ad Z has the same eigenvalues as ad$(S+iT)$, namely 0 and $(\alpha(S)+i\alpha(T))$, $\alpha\in\Delta$; $\alpha(S)$, $\alpha(T)$ are imaginary.

Proof: The first part follows from the relation $\Phi(X,Y)=-iJ(-Y,X)$ and Theorem 4.1. In the second part we can confine ourselves to the case where φ is K-invariant since Φ is K-equivariant and both sides in (30) do not

change if we replace φ by φ_K. Now, regarding the formulas (28) and (29), we only add that the diffeomorphism

$$\mathfrak{m}_+ \times (\mathfrak{m}+\mathfrak{u}) \to \mathfrak{h}_+ + \mathfrak{n}\colon\ S,T+V \mapsto H+X = \Phi(S,T+V),\quad H=S+iT,\quad X=i(1+iI)V,$$

preserves the measure, $dH\,dX = dS\,dT\,d_2V$.

The generalized function $\varphi \mapsto \operatorname{tr} \mathscr{T}_{v\sigma}(\varphi)$ is represented by a locally integrable function $t_{v\sigma}$,

$$\operatorname{tr} \mathscr{T}_{v\sigma}(\varphi) = \int_D \varphi(Z) t_{v\sigma}(Z) j_{\mathfrak{g}}(Z)\,dZ\,. \tag{31}$$

Moreover,[20] $t_{v\sigma}$ is analytic on the subset of regular elements $\mathfrak{g}_r \cap D$, it is G-invariant, and, for $H \in \mathfrak{h}_r$,

$$t_{v\sigma}(H) = \sum_{w\in\mathscr{W}} e^{2\pi i\langle w\cdot F_1, H\rangle} \times \left(\prod_{\alpha>0} |e^{\alpha(H)/2} - e^{-\alpha(H)/2}|^2\right)^{-1}. \tag{32}$$

This expression can be, in principle, derived by inserting into (28) a sequence φ_k of test functions converging to $\delta_{\mathscr{O}_H}$—the Dirac-type generalized function supported on the K orbit $\mathscr{O}_H$.

A similar assertion holds for the representation $\mathscr{U}_{v\sigma}$,

$$\operatorname{tr} \mathscr{U}_{v\sigma}(\varphi) = \int_D \varphi(X,Y) u_{v\sigma}(X,Y) j_{\mathfrak{q}}(X)\,dX\,dY\,. \tag{33}$$

We find explicitly the function $u_{v\sigma}$ using Theorem 6.1 and (32). Regarding the Q invariance we evaluate the function in points $(S_0,T_0)\in\mathfrak{m}_+\times\mathfrak{m}$. Put $H_0 = S_0 + iT_0 \in \mathfrak{h}_+$. Inserting into (31) and (32) a sequence φ_k converging to the Dirac-type generalized function $\delta_{\mathscr{O}}$, $\mathscr{O} = \mathscr{O}_{H_0}$, we have

$$j_{\mathfrak{g}}(H_0) t_{v\sigma}(H_0)\ \operatorname{vol}(\mathscr{O}) = (j_{\mathfrak{q}}(S_0)\, j_{\mathfrak{g}}(H_0))^{1/2} u_{v\sigma}(S_0,T_0) \times |dX\,dY/\Phi^*(dZ)|\operatorname{vol}(\mathscr{O})\,,$$

$$u_{v\sigma}(S_0,T_0) = (j_{\mathfrak{g}}(H_0)/j_{\mathfrak{q}}(S_0,T_0))^{1/2} \times |dX\,dY/\Phi^*(dZ)|^{-1} t_{v\sigma}(H_0)\,. \tag{34}$$

Calculation of the differential $d\Phi$ at the point (S_0,T_0) is clear except of $\partial(I_{S_0+U}T_0)/\partial U$. But this term can be calculated using the identity (22),

$$\frac{d}{ds}(I_{S_0+sU}T_0)_{s=0} = \sum_{j=1}^{r} (-1)^{j+1} a_j (\operatorname{ad} S_0)^{2(j-1)} \operatorname{ad} U(T_0) = -(\operatorname{ad} S_0)^{-1}\circ(\operatorname{ad} T_0)\circ I(U)\,,$$

where $(\operatorname{ad} S_0)^{-1}$ is well defined on $\mathfrak{u}$. Hence we obtain the linear mapping

$$d\Phi\colon (\mathfrak{m}+\mathfrak{u})\times(\mathfrak{m}+\mathfrak{u}) \to \mathfrak{n}_- + \mathfrak{h} + \mathfrak{n}\colon\ (S+U,T+V) \mapsto Y+H+X,$$

where

$$H = S+iT,$$
$$X = i(1+iI)V + \tfrac{1}{2}(1+iI) \times (1+(\operatorname{ad} T_0)(\operatorname{ad} S_0)^{-1}I)U,$$
$$Y = \tfrac{1}{2}(1-iI)(1+(\operatorname{ad} T_0)(\operatorname{ad} S_0)^{-1}I)U\,.$$

Regarding the measure normalization, we find

$$\left|\frac{dX\,dY}{\Phi^*(dZ)}\right| = |\det{}_{\mathfrak{u}}(1+(\operatorname{ad} T_0)(\operatorname{ad} S_0)^{-1}I)|^{-1} = \prod_{\alpha>0}\left|\frac{\alpha(S_0)}{\alpha(H_0)}\right|^2.$$

Inserting the last term into (34) we obtain the desired expression $[S,T\in(\mathfrak{m}_+\times\mathfrak{m})\cap D]$

$$u_{v\sigma}(S,T) = \sum_{w\in\mathscr{W}} e^{2\pi i\langle w\cdot F_2,(S,T)\rangle} \times \left(\prod_{\alpha>0} |e^{\alpha(S)/2} - e^{-\alpha(S)/2}|^2\right)^{-1}. \tag{35}$$

Clearly, the converse procedure is also possible, i.e., to derive (32) starting from (35).

VII. CONCLUSION

The case described with G being simple and complex and K being its maximal subgroup, suggests some obvious generalizations interesting from the point of view of physical interpretation. This G can be replaced by an arbitrary semisimple real Lie group; K need not be compact. Clearly, in the more general case new difficulties arise. We point out one of them in connection with the recent result due to Bohm and Moylan.[21] They realized unitary irreducible representations from the principal series of the de Sitter group in the direct sum of two Hilbert spaces, each of them with a definite unitary irreducible representation of the Poincaré group. This result can be overlooked using the Gell-Mann formula since it yields a Lie–Poisson isomorphism that maps two orbits in the Poincaré algebra onto one orbit in the de Sitter algebra (up to a subset of measure zero) and the images do not intersect. But we emphasize that a general formulation is lacking.

ACKNOWLEDGMENTS

The author is grateful to Professor M. Havlíček and J. Tolar for stimulating discussions.

APPENDIX: THE CASE $G=\mathrm{SL}(n,\mathbb{C})$

Let $G = \mathrm{SL}(n,\mathbb{C})$, $K = \mathrm{SU}(n)$ (of course, now $\dim_{\mathbb{C}} \mathfrak{g} = n^2-1$), and let E^β_α be the standard basis in $\mathrm{gl}(n,\mathbb{C})$, $L^\beta_\alpha = E^\beta_\alpha - (1/n)\delta^\beta_\alpha 1$. The matrices L^β_α generate $\mathrm{sl}(n,\mathbb{C})$ being constrained by one linear equation $L^\alpha_\alpha = 0$ and satisfy the commutation relation

$$[L^\beta_\alpha, L^\mu_\lambda] = \delta^\beta_\lambda L^\mu_\alpha - \delta^\mu_\alpha L^\beta_\lambda\,.$$

We identify $\mathrm{sl}(n,\mathbb{C})^*$ with $\mathrm{sl}(n,\mathbb{C})$ using now the bilinear form $\langle\cdot,\cdot\rangle = \operatorname{Re}\operatorname{tr}(\cdot)$. The fundamental weights are $\epsilon_j = L^1_1 + \cdots + L^j_j$, $j=1,\ldots,n-1$. Put $\nu = q_1\epsilon_1 + \cdots + q_{n-1}\epsilon_{n-1}$, $\sigma = \vartheta_1\epsilon_1 + \cdots + \vartheta_{n-1}\epsilon_{n-1}$, $q_j\in\mathbb{Z}$, $\vartheta_j > 0$. The representation belongs to the Gelfand–Naimark principal series and is induced from the subgroup B of upper triangular matrices by the character usually given in the form[22]

$$\chi(h) = \prod_2^n |h_{kk}|^{-m_k + i\tau_k} (h_{kk}{}^{m_k})\,,$$

where

$$-m_2 = q_1 + \cdots + q_{n-1}, \ldots, -m_n = q_{n-1}\,,$$
$$-\tau_2 = \vartheta_1 + \cdots + \vartheta_{n-1}, \ldots, -\tau_n = \vartheta_{n-1}\,.$$

A closed form can be derived if we extend by linearity the representations $\mathscr{T},\mathscr{U}$ on the complexified Lie algebras $\mathfrak{g}_{\mathbb{C}} = \mathrm{su}(n)_{\mathbb{C}} + i\,\mathrm{su}(n)_{\mathbb{C}}$, $\mathfrak{q}_{\mathbb{C}} = \mathrm{su}(n)_{\mathbb{C}} \oplus \mathrm{su}(n)_{\mathbb{C}}$, $\mathrm{su}(n)_{\mathbb{C}} = \mathrm{sl}(n,\mathbb{C})$, respectively. We put $P_\alpha^\beta = iL_\alpha^\beta$, $T_\alpha^\beta = (0, L_\alpha^\beta)$, and identify $L_\alpha^\beta = (L_\alpha^\beta, 0)$. We write $\hat{L}_\alpha^\beta$ instead of $\mathscr{T}(L_\alpha^\beta) = \mathscr{U}(L_\alpha^\beta)$, etc. The skew-Hermiticity of the representations $\mathscr{T},\mathscr{U}$ now reads

$$(\hat{T}_\alpha^\beta)^+ = \hat{T}_\beta^\alpha, \quad (\hat{P}_\alpha^\beta)^+ = \hat{P}_\beta^\alpha, \quad (\hat{L}_\alpha^\beta)^+ = \hat{L}_\beta^\alpha .$$

Let $X = (x_\beta^\alpha) \in \mathrm{su}(n)$, i.e., $X = x_\beta^\alpha L_\alpha^\beta$. By induction we can verify ($s \in \mathbb{N}$)

$$(\mathrm{ad}\, X)^s L_\alpha^\beta = [A^s(X)_{\alpha;\lambda}^{\beta;\mu}] L_\mu^\lambda ,$$

where

$$[A^s(X)_{\alpha;\lambda}^{\beta;\mu}] = \sum_{j=0}^{s} \binom{s}{j} (-1)^j (X^j)_\lambda^\beta (X^{s-j})_\alpha^\mu .$$

Then according to Sec. V,

$$I_X L_\alpha^\beta = [I_{X\,\alpha;\lambda}^{\beta;\mu}] L_\mu^\lambda ,$$

where

$$[I_{X\,\alpha;\lambda}^{\beta;\mu}] = \sum_{s=1}^{r} (-1)^{s+1} a_s [A^s(X)_{\alpha;\lambda}^{\beta;\mu}] ,$$

$$r = \tfrac{1}{2}n(n-1) ,$$

the a_s are constants depending on σ.

The common spectrum of the operator matrix $((1/2\pi i)\hat{T}_\alpha^\beta)$ is exactly the orbit $\mathscr{O}_{(1/2\pi i)\sigma}$. According to Sec. V and using

$$[\hat{L}_\alpha^\beta, (\hat{T}^s)_\lambda^\mu] = \delta_\lambda^\beta (\hat{T}^s)_\alpha^\mu - \delta_\alpha^\mu (\hat{T}^s)_\lambda^\beta$$

to simplify the final expression, we obtain [$r = \frac{1}{2}n(n-1)$]

$$\hat{P}_\alpha^\beta = \hat{T}_\alpha^\beta - i \sum_{s=1}^{r} \frac{a_s}{(2\pi)^{2s-1}} \sum_{j=0}^{2s-1} \binom{2s-1}{j} \times (-1)^j (\hat{T}^j \hat{L} \hat{T}^{2s-1-j})_\alpha^\beta .$$

The commutation relations

$$[\hat{L}_\alpha^\beta, \hat{P}_\lambda^\mu] = \delta_\lambda^\beta \hat{P}_\alpha^\mu - \delta_\alpha^\mu \hat{P}_\lambda^\beta ,$$

$$[\hat{P}_\alpha^\beta, \hat{P}_\lambda^\mu] = -\delta_\lambda^\beta \hat{L}_\alpha^\mu + \delta_\alpha^\mu \hat{L}_\lambda^\beta ,$$

can be also verified directly, by a tedious but straightforward calculation. The fact that the spectrum of $((1/2\pi i)\hat{T}_\alpha^\beta)$ lies in $\mathrm{su}(n)$, is utilized and at the end the identity

$$-[X,Y] = I[IX,Y] - [IX,IY] + I[X,IY], \qquad X,Y \in \mathrm{su}(n) ,$$

is again met.

Mukunda described a generalized Gell-Mann formula for SL(3,C).[3] Although his formula is similar in nature to the one above there is a substantial difference. In our formula the operators $\hat{T}_\alpha^\beta$ are allowed to appear only in odd powers, namely 1,3,5 for $n = 3$. On the contrary, the powers in the paper[3] are 1,2,3.

In this case ($n = 3$) we add the determination of the constants a_s in terms of the Casimir polynomials. Let $C_j(X) = (i)^j \operatorname{tr} X^j$, $j = 2,3$, be the independent invariant polynomials on su(3), $r = 3$, $\Delta_+ = \{\alpha_1,\alpha_2,\alpha_3\}$, $\sigma = \vartheta_1\epsilon_1 + \vartheta_2\epsilon_2$. Then

$$\beta_j = |\alpha_j((1/2\pi i)\sigma)| ,$$

$$\beta_1 = \frac{\vartheta_1}{2\pi}, \quad \beta_2 = \frac{\vartheta_2}{2\pi}, \quad \beta_3 = \frac{\vartheta_1 + \vartheta_2}{2\pi} .$$

Let s_1,s_2,s_3 be the elementary symmetric functions in the β's,

$$s_1 = \beta_1 + \beta_2 + \beta_3, \quad s_2 = \beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3,$$

$$s_3 = \beta_1\beta_2\beta_3 .$$

We have

$$a_j = p_j/(s_1s_2 - s_3)s_3, \quad j = 1,2,3,$$

where

$$p_1 = s_1{s_2}^2 - {s_1}^2 s_3 - s_2 s_3, \quad -p_2 = s_1 - 2s_1s_2 + s_3 ,$$

$$p_3 = s_1 , \quad s_1 = 2\sqrt{2C_2}\cos\phi,$$

$$s_2 = \tfrac{1}{2}C_2(1 + 4\cos 2\phi), \quad s_3 = (1/\sqrt{2})C_2^{3/2}\cos 3\phi ,$$

$$\phi = \tfrac{1}{3}\arcsin(\sqrt{2/3}\, C_3 C_2^{-3/2}) .$$

[1] R. Hermann, Commun. Math. Phys. **2**, 155 (1966).
[2] E. Weimar, Nuovo Cimento Lett. **4**, 43 (1972).
[3] N. Mukunda, J. Math. Phys. **10**, 897 (1969).
[4] G. W. Mackey, in *Lie Groups and Their Representations*, edited by I. M. Gelfand (Kaido, Budapest, 1975).
[5] A. O. Barut and R. Raczka, *The Theory of Group Representations and Applications* (PNW, Warsaw, 1977); G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[6] P. Božek, M. Havlíček, and O. Navrátil, University Carolina Pragensis preprint NC/TF/1, 1985.
[7] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, New York, 1976); B. Kostant, *Lecture Notes in Mathematics*, Vol. 170 (Springer, Berlin, 1970), p. 87; B. Kostant, *Lecture Notes in Physics*, Vol. 6 (Springer, Berlin, 1970), p. 237.
[8] N. E. Hurt, *Geometric Quantization in Action* (Reidel, Dordrecht, 1983).
[9] A. Weinstein, J. Differ. Geom. **18**, 523 (1983).
[10] A. A. Kirillov, Russ. Math. Surveys **31**, 56 (1976) [Usp. Mat. Nauk **31**, 57 (1976)].
[11] H. J. Sussmann, Bull. Am. Math. Soc. **79**, 197 (1973); P. Stefan, *ibid.* **80**, 1142 (1974).
[12] Reference 8, Chap. 6.
[13] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[14] Reference 8, §9.3.
[15] R. Abraham and J. Marsden, *Foundations of Mechanics* (Benjamin, Reading, MA, 1978).
[16] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry II* (Interscience, New York, 1969).
[17] M. Duflo, Funkt. Anal. Priloz. **4**, 38 (1970); E. A. Gutkin, *ibid.* **4**, 32 (1970).
[18] A. A. Kirillov, Funkt. Anal. Priloz. **2**, 40 (1968).
[19] M. Duflo in Ref. 18.
[20] Harish-Chandra, Trans. Am. Math. Soc. **76**, 485 (1954).
[21] A. Bohm and P. Moylan, J. Math. Phys. **26**, 365 (1985).
[22] D. P. Zhelobenko and A. I. Stern, *Representations of Lie Groups* (Nauka, Moscow, 1983).

τ functions and zero curvature equations of Toda–AKNS type

M. J. Bergvelt
Max Planck Institut für Mathematik, Gottfried Clarenstrasse 26, 5300 Bonn 3, Federal Republic of Germany

A. P. E. ten Kroode
Mathematisch Instituut, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

(Received 8 July 1987; accepted for publication 10 February 1988)

The connection between τ functions and zero curvature equations for the homogeneous construction of the basic module $L(\Lambda_0)$ over the simplest affine Kac–Moody algebra $A_1^{(1)}$ is studied.

I. INTRODUCTION AND SUMMARY

In the representation theoretic approach to soliton equations, initiated by Date *et al.*,[1] one derives so-called Hirota bilinear equations as defining equations for the orbit of an algebraic group associated with an affine Kac–Moody algebra in a representation space for this algebra; see, e.g., Kac.[2] The relation with the usual formulation of soliton equations is then given by a change of dependent variables exemplified by the famous formula

$$u = 2\frac{\partial^2}{\partial x^2}\log\tau. \tag{1.1}$$

If τ satisfies a certain set of Hirota equations, u will satisfy the KdV hierarchy.

Formula (1.1) is a quite mysterious and unmotivated substitution; what one would like is an explanation of this formula and, more generally, of the older theory of soliton equations including zero curvature conditions, conservation laws, Miura transformations, etc. For a review of these matters we refer the reader to Drinfeld and Sokolov.[3]

Wilson[4,5] has given a group theoretical explanation of the formula

$$v = \frac{\partial}{\partial x}\ln\frac{\tau^{(1)}}{\tau^{(0)}}, \tag{1.2}$$

where $\tau^{(0)}$ and $\tau^{(1)}$ are the τ functions associated with the principal realizations of the fundamental modules $L(\Lambda_0)$ and $L(\Lambda_1)$ over the simplest affine algebra $A_1^{(1)}$. This substitution leads to the modified KdV hierarchy, which is related to the KdV hierarchy by a Miura transformation. A slight extension of Wilson's method also provides an explanation for (1.1), where one may choose for τ either $\tau^{(0)}$ or $\tau^{(1)}$.

In the homogeneous realization of the modules $L(\Lambda_0)$ and $L(\Lambda_1)$ the τ functions are multicomponent objects $\tau^{(i)} = (\tau_l^{(i)})_{l\in\mathbb{Z}}$, $i = 0,1$, and it is interesting to see what the Hirota bilinear equations look like in this case. These equations were constructed in Ref. 6. There we found that another unmotivated substitution, namely,

$$q^l = -\tau_{l+1}^{(i)}/\tau_l^{(i)}, \quad r^l = \tau_{l-1}^{(i)}/\tau_l^{(i)}, \quad i = 0,1, \tag{1.3}$$

leads to the first two equations (nonlinear Schrödinger and mod-KdV) of the AKNS hierarchy on a lattice $L \simeq \mathbb{Z}$, while the quantities $u^l := \ln q^l$ satisfy the equations for the Toda lattice. It was, however, not clear to us whether the Hirota equations for $\tau^{(i)}$ would really imply all AKNS equations for q^l and r^l.

The AKNS equations are usually derived as integrability conditions for an infinite set of linear differential equations for the so-called wave function. In this context one often speaks of zero curvature conditions. In view of the results obtained from representation theory it was natural to look for an extension of this zero curvature formalism to an AKNS system on a lattice. In Ref. 7 we have shown that such an extension of the AKNS system can be constructed in a natural way and that solutions in different lattice sites are indeed related by Toda equations.

The main motivation for writing this paper was to explain the relation between the representation theoretic approach and the zero curvature construction of this Toda–AKNS hierarchy in the spirit of Wilson. The key ingredient is the Birkhoff decomposition, which, together with some background material on Kac–Moody algebras, will be discussed in Sec. II. In the next sections we will, using the Birkhoff decomposition of the "dressed vacuum," derive differential difference zero curvature equations both on the affine algebra (i.e., including the center) and on the loop algebra lying underneath it. These equations turn out to be the Toda–AKNS equations for a pair of fundamental fields q^l and r^l. We will proceed to show that these fields are related to τ functions via (1.3), thereby explaining this substitution and proving that the q^l and r^l from (1.3) do indeed satisfy all AKNS equations.

In Sec. VI we will review two well-known constructions of the AKNS conservation laws and we will show how they are related. Moreover, we will supplement these "continuous" conservation laws with conservation laws for the discrete evolution. It is interesting to remark that the explicit form (6.8) of the AKNS conserved densities led Flaschka, Newell, and Ratiu[8] to introduce the τ function as a kind of potential. Another reason for writing this paper was to derive their results in a pure Lie algebraic framework.

From the construction of the Toda–AKNS equations in this paper it will be clear that the orbit of the algebraic Kac–Moody group gives rise to a certain class of solutions to these equations. This then leads to the question of whether conversely every solution in this class comes from an element of the group orbit. In Sec. VII we will see that this is indeed true, showing that the Toda–AKNS equations describe the orbit of the algebraic group, just as the Hirota equations do. In order to give a group theoretical description of more general solutions, one will have to consider orbits of various completions of this algebraic group. We will not discuss this here.

 0022-2488/88/061308-13$02.50

II. PRELIMINARIES

A. Introduction

In this section we will, mainly in order to fix notation, briefly recall some of the basic facts about affine Kac–Moody algebras and the groups associated with them. For more details we refer to Kac.[2] Following Goodman and Wallach[9] we will also mention certain completions of these algebras and groups, which are needed to make some of the constructions of this paper rigorous. Finally we will describe the Birkhoff decomposition for affine Kac–Moody groups and their completions, which will be the main tool for deriving zero curvature equations in Sec. IV.

B. Affine Kac–Moody algebras

Let $\underline{g}$ be a finite-dimensional simple Lie algebra, $\tilde{\underline{g}} := \mathbb{C}[\lambda,\lambda^{-1}] \otimes \underline{g}$ the polynomial loop algebra associated with $\underline{g}$, and $\hat{\underline{g}} := \tilde{\underline{g}} \oplus \mathbb{C}c$ the universal central extension of $\tilde{\underline{g}}$. The degree derivation $d: \hat{\underline{g}} \to \hat{\underline{g}}$ is given by

$$d_{|\tilde{\underline{g}}} := \lambda \frac{d}{d\lambda}, \quad d(c) = 0. \tag{2.1}$$

The untwisted affine Kac–Moody algebra $\hat{\underline{g}}^e$ (e = extended) associated to $\underline{g}$ is then defined as the semidirect product algebra $\hat{\underline{g}}^e := \hat{\underline{g}} \oplus \mathbb{C}d$ and one has $\hat{\underline{g}} = [\hat{\underline{g}}^e, \hat{\underline{g}}^e]$.

Fix a Cartan subalgebra $\underline{h}$ of $\underline{g}$. Let $\Delta \subset \underline{h}^*$ be the corresponding root system $\{\alpha_1,\alpha_2,\ldots,\alpha_l\}$ a system of simple roots and $\theta = \sum_{i=1}^{l} a_i\alpha_i$ the highest root. Choose root vectors E_α, $\alpha \in \Delta$, normed in such a way that the triple E_α, $E_{-\alpha}$, and $H_\alpha := [E_\alpha, E_{-\alpha}]$ has the standard $\mathrm{sl}(2,\mathbb{C})$ commutation relations. The vectors $E_i := E_{\alpha_i}$, $F_i := E_{-\alpha_i}$, $i = 1,2,\ldots,l$, generate $\underline{g}$ and one has $\underline{h} = \oplus_{i=1}^{l} \mathbb{C}H_i$, where $H_i := H_{\alpha_i}$, the simple coroots.

The algebra $\hat{\underline{g}}^e$ has a Cartan subalgebra $\hat{\underline{h}}^e := \underline{h} \oplus \mathbb{C}c \oplus \mathbb{C}d$ and a corresponding root system $\hat{\Delta} \in (\hat{\underline{h}}^e)^*$. The so-called imaginary root $\delta \in \hat{\Delta}$ is given by

$$\delta_{|\underline{h}} = 0, \quad \langle \delta, c \rangle = 0, \quad \langle \delta, d \rangle = 1, \tag{2.2}$$

and $\hat{\Delta}$ is the disjoint union of a set of real roots $\hat{\Delta}^{\mathrm{re}} := \{j\delta + \alpha | j \in \mathbb{Z}, \alpha \in \Delta\}$ and a set of imaginary roots $\hat{\Delta}^{\mathrm{im}} = \{j\delta | j \in \mathbb{Z} \setminus \{0\}\}$. Define root vectors $e_i := 1 \otimes E_i$, $f_i := 1 \otimes F_i$, $i = 1,2,\ldots,l$, associated with the roots $\alpha_i \in \hat{\Delta}$ and $e_0 := \lambda \otimes E_{-\theta}$, $f_0 := \lambda^{-1} \otimes E_\theta$ associated with the root $\alpha_0 := \delta - \theta \in \hat{\Delta}$. The set $\{\alpha_0,\alpha_1,\ldots,\alpha_l\}$ is then a simple system for $\hat{\Delta}$ and the vectors e_i, f_i, $i = 0,1,\ldots$, generate $\hat{\underline{g}}$. The simple coroots are, as before, defined by $\alpha_i^\vee := [e_i, f_i]$, $i = 0,1,\ldots,l$.

For all real roots $\alpha \in \hat{\Delta}^{\mathrm{re}}$ one defines a reflection r_α: $(\hat{\underline{h}}^e)^* \to (\hat{\underline{h}}^e)^*$ as usual. The group $\hat{W}$ generated by the reflections $r_i := r_{\alpha_i}$, $i = 0,1,\ldots,l$, is called the affine Weyl group. It contains an Abelian normal subgroup T, called the translation group, which is defined as the group generated by

$$T_i := r_{\delta - \alpha_i} r_i, \quad i = 1,2,\ldots,l. \tag{2.3}$$

The quotient $\hat{W}/T$ is easily seen to be isomorphic to the (finite) Weyl group W of $\underline{g}$ and therefore one has

$$\hat{W} = W \times T. \tag{2.4}$$

C. Affine Kac–Moody groups

In order to describe groups associated to $\hat{\underline{g}}$ one uses the representation theory of this algebra. Let $\Lambda \in \hat{\underline{h}}^*$ be a dominant integral weight, i.e., $\langle \Lambda, \alpha_i^\vee \rangle \in \mathbb{Z}_+ \cup \{0\}$, $i = 0,1,\ldots,l$, and $(L(\Lambda), \pi_\Lambda)$ the unique irreducible highest weight module with highest weight Λ. We recall that such modules can be equipped with a Hermitian form $H_\Lambda: L(\Lambda) \times L(\Lambda) \to \mathbb{C}$, giving them the structure of a pre-Hilbert space. Such a form is—up to a constant factor—uniquely determined by the contravariance condition

$$H_\Lambda(v, \pi_\Lambda(x)(w)) = -H_\Lambda(\omega_0(\pi_\Lambda(x))(v), w), \tag{2.5}$$

where ω_0: $\hat{\underline{g}} \to \hat{\underline{g}}$ denotes the antilinear involution defined by $\omega_0(e_i) = -f_i$, $\omega_0(f_i) = -e_i$, $\omega_0(\alpha_i^\vee) = -\alpha_i^\vee$, $i = 0,1,\ldots,l$.

Let G be the connected and simply connected group associated to $\underline{g}$. By considering a faithful representation of G, it can always be realized as a subgroup of $\mathrm{SL}_n(\mathbb{C})$. This enables us to define the polynomial loop group $\tilde{G}$ by

$$\tilde{G} = \{g \in \mathrm{SL}_n(\mathbb{C}[\lambda,\lambda^{-1}]) | g(\lambda) \in G \ \forall \lambda \in S^1\}. \tag{2.6}$$

Let $\pi := \oplus_{i=0}^{l} \pi_{\Lambda_i}$, where the fundamental weights Λ_i are defined by $\langle \Lambda_i, \alpha_j^\vee \rangle = \delta_{ij}$, $i,j = 0,1,\ldots,l$, and let $\hat{G}$ be the group generated by $\exp t\pi(e_i)$, $\exp t\pi(f_i)$, $i = 0,1,\ldots,l$, $t \in \mathbb{C}$. This definition makes sense because $\pi_\Lambda(e_i)$ and $\pi_\Lambda(f_i)$ are locally nilpotent operators for all dominant integral weights Λ. One can show that $\hat{G}$ is a central extension of $\tilde{G}$ by $\mathbb{C}^*$, i.e., there is an exact sequence

$$1 \to \mathbb{C}^* \to \hat{G} \to \tilde{G} \to 1. \tag{2.7}$$

Instead of this "universal" group $\hat{G}$, one often considers the groups $\hat{G}^\Lambda$ generated by $\exp t\pi_\Lambda(e_i)$, $\exp t\pi_\Lambda(f_i)$, $i = 0,1,\ldots,l$, $t \in \mathbb{C}$. These groups are quotients of $\hat{G}$, i.e., there exist surjective homomorphisms $f_\Lambda: \hat{G} \to \hat{G}^\Lambda$. The kernel Z_Λ of f_Λ can be shown to be a finite central subgroup of $\hat{G}$ (see Ref. 9).

D. Completions

Consider the matrix

$$h = \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix} \in \mathrm{sl}_2(\mathbb{C}[\lambda,\lambda^{-1}]).$$

Exponentiating this matrix we obtain

$$\exp h = \begin{pmatrix} e^\lambda & 0 \\ 0 & e^{-\lambda} \end{pmatrix}, \tag{2.8}$$

which is clearly not in the polynomial loop group $\mathrm{Sl}_2(\mathbb{C}[\lambda,\lambda^{-1}])$. Still, we want to be able to exponentiate elements like this. The example shows that this only makes sense in suitable completions of the polynomial loop group $\tilde{G}$.

In Ref. 9 several completions of the groups $\tilde{G}$ and $\hat{G}$ are constructed. For the reader's convenience we will give a short summary of the results relevant to this paper.

A function $p: \mathbb{Z} \to (0,\infty)$ is called a weight function if it satisfies $p(k+m) \leqslant p(k) \cdot p(m)$. (N.B.: we use p instead of the more obvious w to avoid confusion with elements of the Weyl group.) We will also require that p is symmetric, i.e., $p(k) = p(-k)$, and that $p(0) = 1$. In this case one has

$$p(k) = \sqrt{p(k)\,p(-k)} \geqslant \sqrt{p(0)} = 1, \quad \forall k \in \mathbb{Z}. \tag{2.9}$$

Weight functions are used to define the so-called weighted Wiener algebra A_p as the Banach algebra of functions $f: S^1 \to \mathbb{C}$ with norm

$$\|f\|_p := \sum_{k=-\infty}^{+\infty} |a_k| p(k), \tag{2.10}$$

the a_k's being the Fourier coefficients of f. The ring $\mathbb{C}[\lambda, \lambda^{-1}]$ is obviously a dense subset of A_p.

We will also use the weight functions to construct completions of the algebras $\tilde{\underline{g}}$ and $\hat{\underline{g}}$; let x_i, $i = 1,2,...,n$, be a basis of $\underline{g}$ and define norms on $\tilde{\underline{g}}$ and $\hat{\underline{g}}$ by

$$\|x\|_p := \sum_{ij} |a_{ij}| p(i), \quad \forall x = \sum_{ij} a_{ij} \lambda^i \otimes x_j \in \underline{g},$$
$$\|\hat{x}\|_p := \sum_{ij} |a_{ij}| p(i) + |\alpha|, \quad \forall \hat{x} = \sum_{ij} a_{ij} \lambda^i \otimes x_j + \alpha c \in \hat{\underline{g}}. \tag{2.11}$$

(N.B.: another choice of basis yields an equivalent norm.) Denote by $(\tilde{\underline{g}})_p$ and $(\hat{\underline{g}})_p$ the completions of $\tilde{\underline{g}}$ and $\hat{\underline{g}}$ with respect to these norms. For arbitrary weight p the algebra $(\tilde{\underline{g}})_p$ is a Banach Lie algebra, i.e.,

$$\|[x,y]\|_p \leqslant M \|x\|_p \|y\|_p, \quad \forall x,y \in (\tilde{\underline{g}})_p. \tag{2.12}$$

The same holds for $(\hat{\underline{g}})_p$ if p is chosen such that $p(k) \geqslant C|k|^{1/2}$ for some constant C.

Define the Banach Lie group $\tilde{G}_p$ by

$$\tilde{G}_p := \{g \in \mathrm{Gl}_n(A_p) \mid g(\lambda) \in G, \quad \forall \lambda \in S^1\}. \tag{2.13}$$

The exponential mapping carries the Lie algebra $(\tilde{\underline{g}})_p$ into $\tilde{G}_p$ and one can show that it covers a neighborhood of the identity in $\tilde{G}$.

To construct a group associated with the algebra $(\hat{\underline{g}})_p$, we use again the integrable modules $L(\Lambda)$. Since we aim to exponentiate elements that are not nilpotent, we will certainly have to consider the Hilbert completion $H(\Lambda)$ of $L(\Lambda)$ with respect to the Hermitian form H_Λ. It can be shown that, if the weight p is chosen from the family

$$p_{\sigma,t}(k) := \exp t|k|^{1/\sigma}, \quad t > 0, \quad 1 < \sigma < 2, \tag{2.14}$$

there exists a dense subspace of $H(\Lambda)$ on which the operators $\exp \pi_\Lambda(x)$, $x \in (\hat{\underline{g}})_p$ are well defined. Therefore $\hat{G}_p$ can be defined as the group generated by these operators.

One now proceeds along the same lines as in the preceding section; the universal group $\hat{G}_p$ is the group associated with the representation $\pi = \oplus_{i=0}^{l} \pi_{\Lambda_i}$. It is again a central extension of $\hat{G}_p^{\Lambda}$ by the finite group Z^{Λ} and a central extension of $\tilde{G}_p$ by $\mathbb{C}^*$.

E. The Birkhoff decomposition

Let $\hat{U}_\pm$ be the subgroup of $\hat{G}$ generated by $\exp t\pi(E_\alpha)$, $\alpha \in \hat{\Delta}_\pm^{\mathrm{re}}$, $t \in \mathbb{C}$, and $\hat{H}$ the Cartan subgroup generated by $\exp t\pi(\alpha_i^\vee)$, $i = 0,1,...,l$, $t \in \mathbb{C}$. Define for $i = 0,1,...,l$,

$$r_i^\pi := \exp \pi(e_i) \exp[-\pi(f_i)] \exp \pi(e_i). \tag{2.15}$$

The group $\hat{W}^\pi$ generated by r_i^π, $i = 0,1,...,l$, is then an extension of the affine Weyl group $\hat{W}$ by an Abelian normal subgroup $D^\pi \subset \hat{H}$.[10]

Kac and Peterson[11] have proved that $\hat{G}$ admits the Birkhoff decomposition

$$\hat{G} = \bigcup_{w \in \hat{W}^\pi} \hat{U}_- w \hat{H} \hat{U}_+. \tag{2.16}$$

On the level of the completed group $\hat{G}_p$ one has an analogous formula,

$$\hat{G}_p = \bigcup_{w \in \hat{W}^\pi} (\hat{U}_-)_p w \hat{H} (\hat{U}_+)_p. \tag{2.17}$$

Here $(\hat{U}_\pm)_p$ are the completions of $\hat{U}_\pm$ in $\hat{G}_p$.

For our purposes we need a slight modification of (2.17); let U_+ (U_-) be the subgroup of G consisting of all upper (lower) triangular matrices with 1's on the diagonal in G and denote by $(\tilde{U}_\pm)_p$ the projections of $(\hat{U}_\pm)_p$ in $\tilde{G}_p$. Then for all $U \in (\tilde{U}_-)_p$ and $V \in (\tilde{U}_+)_p$ we may write

$$\begin{aligned} U &= (1 + U_1\lambda + U_2\lambda^2 + \cdots) U_0, \quad U_0 \in U_-, \\ V &= V_0(1 + V_1\lambda + V_2\lambda^2 + \cdots), \quad V_0 \in U_+. \end{aligned} \tag{2.18}$$

Using this and the projection of (2.17) on $\tilde{G}_p$ we see that any $\tilde{g} \in \tilde{G}_p$ can be factorized as

$$\tilde{g} = \tilde{g}_- \cdot \tilde{g}_0 \cdot \tilde{g}_+, \tag{2.19}$$

where

$$\tilde{g}_- = 1 + U_1\lambda + U_2\lambda^2 + \cdots, \quad \tilde{g}_0 = U_0 w H V_0,$$
$$\tilde{g}_+ = 1 + V_1\lambda + V_2\lambda^2 + \cdots,$$

and correspondingly any $\hat{g} \in \hat{G}_p$ can be written

$$\hat{g} = \hat{g}_- \hat{g}_0 \hat{g}_+. \tag{2.20}$$

III. THE τ FUNCTION

A. Introduction

In this section we briefly describe the homogeneous realization of the module $L(\Lambda_0)$. For more details we refer to Frenkel and Kac[10] and Segal.[12] Although most of the results of this paper can be derived for an arbitrary simply laced algebra $\hat{\underline{g}}$, we will restrict ourselves to the simplest affine algebra $A_1^{(1)}$. We proceed to write the components $\tau_l^{(0)}$ of the τ function in the homogeneous realization as vacuum expectation values of certain group elements ψ^l (dressed vacua), which will play an important role in the construction of zero curvature equations in Sec. IV. We will also study the relation between the Birkhoff factorization of these group elements and the zeros of the components $\tau_l^{(0)}$.

B. The homogeneous realization of $L(\Lambda_0)$

Let $\underline{g} = \mathrm{sl}(2,\mathbb{C})$ with standard basis

$$E_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad F_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{3.1}$$

In the sequel we will write e for E_1, h for H_1, and f for F_1. The homogeneous Heisenberg subalgebra $\hat{\underline{s}}$ of $\hat{\underline{g}}$ is given by

$$\hat{\underline{s}} := \bigoplus_{i>0} \mathbb{C}p_i \oplus \mathbb{C}c \oplus \bigoplus_{i>0} \mathbb{C}q_i, \tag{3.2}$$

where $p_i := \frac{1}{2}\lambda^i \otimes h$ and $q_i := (1/i)\lambda^{-i} \otimes h$ satisfy the Heisenberg commutation relations

$$[p_i, q_j] = \delta_{ij} c. \tag{3.3}$$

Let T be the operator

$$T := r_0^{\pi_{\Lambda_0}} r_1^{\pi_{\Lambda_1}} \tag{3.4}$$

and denote by $T^{\pi_{\Lambda_0}}$ the group generated by T. It is well known that the elements of $T^{\pi_{\Lambda_0}}$ centralize the action of $\pi_{\Lambda_0}(\underline{\hat{s}})$ and that the module $L(\Lambda_0)$ is irreducible under the action of the pair $(\pi_{\Lambda_0}(\underline{\hat{s}}), T^{\pi_{\Lambda_0}})$. Therefore $L(\Lambda_0)$ has the following structure:

$$L(\Lambda_0) \simeq \mathbb{C}[x_i;\ i>0] \otimes \mathbb{C}[Q]. \tag{3.5}$$

Here $Q := \mathbb{Z}\alpha$ is the root lattice of $\mathrm{sl}(2,\mathbb{C})$ and $\mathbb{C}[Q]$ is the group algebra of Q, i.e., the vector space spanned by formal exponentials $e^{k\alpha}$, $k\in\mathbb{Z}$. The action of $\pi_{\Lambda_0}(\underline{\hat{s}})$ and $T^{\pi_{\Lambda_0}}$ is given explicitly by

$$\begin{aligned}
&\pi_{\Lambda_0}(p_i)(P\otimes e^{k\alpha}) = \frac{\partial P}{\partial x_i}\otimes e^{k\alpha},\\
&\pi_{\Lambda_0}(q_i)(P\otimes e^{k\alpha}) = x_i P\otimes e^{k\alpha},\\
&T^l(P\otimes e^{k\alpha}) = P\otimes e^{(k+l)\alpha},\quad \forall P\in\mathbb{C}[x_i],\quad k,l\in\mathbb{Z},\quad i\in\mathbb{Z}_+.
\end{aligned} \tag{3.6}$$

The action of the other algebra elements is given in terms of vertex operators. We will not need these operators in this paper.

For future use we also mention the weight system $P(\Lambda_0)$ of $L(\Lambda_0)$;

$$P(\Lambda_0) = \{\Lambda_0 + m\alpha - (m^2+k)\delta \mid m\in\mathbb{Z},\quad k\in\mathbb{Z}_+\cup\{0\}\}. \tag{3.7}$$

Remark: Since we will only work with the module $L(\Lambda_0)$ we will from now on leave out the symbol π_{Λ_0} and simply write $x\cdot v$ for $\pi_{\Lambda_0}(x)(v)$ [$x\in\hat{\underline{g}}$, $v\in L(\Lambda_0)$]. We will also write H for H_{Λ_0}, the Hermitian form (2.5) on $L(\Lambda_0)$, and v_0 for the highest weight vector at $L(\Lambda_0)$.

C. The τ function as vacuum expectation value

Consider the Kac–Moody group $\hat{G}^{\Lambda_0}$ and the orbit $0_{\Lambda_0} := \{\tau^{(0)} = \hat{g}\cdot v_0 \mid \hat{g}\in\hat{G}^{\Lambda_0}\}$ passing through the highest weight vector. In the homogeneous realization (3.5) of $L(\Lambda_0)$ the elements of the group orbit $\tau^{(0)}$ are of the form

$$\tau^{(0)} = \sum_{l\in\mathbb{Z}} \tau_l^{(0)}(x)\otimes e^{l\alpha}. \tag{3.8}$$

Since we are working with the algebraic group $\hat{G}^{\Lambda_0}$, the components $\tau_l^{(0)}(x)$ are identically zero for almost all $l\in\mathbb{Z}$, while the nonzero components are polynomials in the x_i's.

We introduce an isomorphism of $L(\Lambda_0)$,

$$\hat{\psi}_{\mathrm{vac}}(t) := \exp\left(\sum_{i=1}^{\infty} t_i p_i\right),\quad t_i\in\mathbb{C}. \tag{3.9}$$

This operator does not belong to the group $\hat{G}^{\Lambda_0}$, but if we take almost all t_i's to be zero, it does belong to an arbitrary completion $\hat{G}_p^{\Lambda_0}$, where p is a weight of the form (2.14). It acts on the group orbit as a shift operator

$$\hat{\psi}_{\mathrm{vac}}(t)\cdot\tau^{(0)}(x) = \tau^{(0)}(x+t) = \sum_{l\in\mathbb{Z}} \tau_l^{(0)}(x+t)\otimes e^{l\alpha}. \tag{3.10}$$

Using the Hermitian form H on $L(\Lambda_0)$ we can project (3.10) on the vector $e^{l\alpha}$, thereby obtaining a useful expression for $\tau_l^{(0)}(t)$;

$$\begin{aligned}
\tau_l^{(0)}(t) &= H(e^{l\alpha}, \hat{\psi}_{\mathrm{vac}}\cdot\tau^{(0)})\\
&= H(T^l\cdot v_0, \hat{\psi}_{\mathrm{vac}}\cdot\hat{g}\cdot v_0)\\
&= H(v_0, T^{-l}\hat{\psi}_{\mathrm{vac}}\cdot\hat{g}\cdot v_0).
\end{aligned} \tag{3.11}$$

Note that, since $\tau_l^{(0)}$ is a polynomial, we may indeed take almost all t_i to be zero in (3.9).

We find that the components $\tau_l^{(0)}(t)$ are the vacuum expectation value of

$$\hat{\psi}^l(t) := T^{-l}\hat{\psi}_{\mathrm{vac}}(t)\hat{g}\in\hat{G}_p^{\Lambda_0}. \tag{3.12}$$

Expressions for τ functions in terms of vacuum expectation values were introduced by the Kyoto school, see, e.g., Ref. 1. The element $\hat{\psi}^l$ will provide the connection between representation theory and zero curvature equations in the following sections. This connection was first explained by Wilson[4,5] for the modified KdV equation [using the principal realization of $L(\Lambda_0)$].

D. Birkhoff decomposition for $\hat{\psi}^l$

In this section we will study the Birkhoff decomposition of $\hat{\psi}^l$. Since $\hat{\psi}^l(t)$ belongs to $\hat{G}_p^{\Lambda_0}$ for all l and t, it has a factorization (2.20). (We use here the projection $\hat{G}_p \to \hat{G}_p^{\Lambda_0}$ described in Sec. C.) Therefore we write

$$\hat{\psi}^l(t) = \hat{\psi}^l_-(t)\hat{\psi}^l_0(t)\hat{\psi}^l_+(t), \tag{3.13}$$

where

$$\begin{aligned}
\hat{\psi}^l_0(t) &= \exp\{a^l(t)f\}\cdot T^k\cdot w\cdot\exp\{\lambda_0^l(t)\alpha_0^v + \lambda_1^l(t)\alpha_1^v\}\\
&\quad\cdot\exp\{b^l(t)e\}.
\end{aligned} \tag{3.14}$$

Here $w\in\{1, r_1^{\pi_{\Lambda_0}}\}$ and the integer k depends on l and t.

Substituting (3.13) in (3.11) and using the fact that $\hat{\psi}^l_+$ and $\exp\{b^l e\}$ stabilize v_0 as well as the contravariance of the form H, we obtain for the components of the τ function

$$\begin{aligned}
\tau_l^{(0)}(t) &= H(v_0, \hat{\psi}^l_-\hat{\psi}^l_0\hat{\psi}^l_+ v_0)\\
&= H(v_0, T^k\cdot w\cdot\exp\{\lambda_0^l\alpha_0^v + \lambda_1^l\alpha_1^v\}v_0).
\end{aligned} \tag{3.15}$$

Since $\alpha_i^v\cdot v_0 = \delta_{i0}$, $w\cdot v_0 = v_0$ this becomes

$$\tau_l^{(0)}(t) = H(v_0, T^k v_0)e^{\lambda_0^l(t)} = \delta_{k0}e^{\lambda_0^l(t)}. \tag{3.16}$$

This formula shows that $\tau_l^{(0)}(t)$ vanishes if and only if $k\neq 0$. Remember that $\tau_l^{(0)}(t)$ is a polynomial, say in the variables $t_1, t_2, \ldots, t_n$. Therefore if it does not vanish identically, its zero set is a closed, nowhere dense subset of $\mathbb{C}^n$. For all t outside this set k must be zero. In this case the factorization (3.13) has very nice properties; projecting on the loop group $\tilde{G}_p$ one obtains [see (2.19)]

$$\tilde{\psi}^l(t) = \tilde{\psi}^l_-(t)\tilde{\psi}^l_0(t)\tilde{\psi}^l_+(t), \tag{3.17}$$

where

$$\begin{aligned}
&\tilde{\psi}^l_-(t) = 1 + A_1\lambda^{-1} + A_2\lambda^{-2} + \cdots,\\
&\tilde{\psi}^l_0(t) = B_0\in G = \mathrm{Sl}_2(\mathbb{C}),\\
&\tilde{\psi}^l_+(t) = 1 + C_1\lambda + C_2\lambda^2 + \cdots.
\end{aligned}$$

Thus we have a factorization in strictly negative powers of λ, powers zero, and strictly positive powers. Furthermore, the factors $\tilde{\psi}^l_-(t)$, $\tilde{\psi}^l_0(t)$, and $\tilde{\psi}^l_+(t)$ are differentiable with respect to t (see Pressley and Segal[13]).

On the zero set of $\tau_l^{(0)}(t)$ this is no longer true, giving rise to singularities in the solutions of the zero curvature equations to be derived in the following sections. For more information on the zeros of τ functions we refer to Helminck[14] and references therein.

IV. ZERO CURVATURE EQUATIONS

A. Introduction

In Sec. IV B we will derive linear differential equations for the operators $\hat{\psi}^l_-(t)$ and $\hat{\psi}^l_0(t)\hat{\psi}^l_+(t)$, which are valid for all t outside the zero set of the polynomial $\tau^{(0)}_l(t)$. Moreover, we will show that, if t is such that $\tau^{(0)}_{l+1}(t)$ is also nonzero, these differential equations may be supplemented with a set of linear difference equations. Compatibility of the resulting set of differential difference equations leads to the usual zero curvature conditions together with a set of lattice zero curvature conditions, both formulated on $(\hat{\underline{g}})_p$.

In the second section we will project out the center to obtain zero curvature equations on the algebra $(\tilde{\underline{g}})_p$ and see that no essential information is lost.

B. Zero curvature equations on $(\hat{\underline{g}})_p$

Using the definition (3.9) and (3.12) of the operator $\hat{\psi}^l$, we see that it satisfies the differential equations

$$\partial_{t_i}\hat{\psi}^l = p_i\hat{\psi}^l, \quad i = 1,2,\ldots . \tag{4.1}$$

Using the Birkhoff decomposition (3.13), this can be rewritten as

$$(\hat{\psi}^l_-)^{-1}\partial_{t_i}\hat{\psi}^l_- + (\partial_{t_i}\hat{\psi}^l_{0,+})(\hat{\psi}^l_{0,+})^{-1} = \hat{R}^l(p_i), \tag{4.2}$$

where we have introduced

$$\hat{\psi}^l_{0,+} := \hat{\psi}^l_0\hat{\psi}^l_+, \quad \hat{R}^l(p_i) := (\hat{\psi}^l_-)^{-1}p_i(\hat{\psi}^l_-). \tag{4.3}$$

Note that (4.2) is an equation on the algebra $\pi_{\Lambda_0}((\hat{\underline{g}})_p) \simeq \hat{\underline{g}}_p$.

Define subalgebras $\hat{\underline{g}}_+$ and $\hat{\underline{g}}_-$ of $\hat{\underline{g}}$ by

$$\begin{aligned} \hat{\underline{g}}_+ &:= \bigoplus_{i>0} \lambda^i \otimes \underline{g} \oplus \mathbb{C}c, \\ \hat{\underline{g}}_- &:= \bigoplus_{i<0} \lambda^i \otimes \underline{g}, \end{aligned} \tag{4.4}$$

and denote by $(\hat{\underline{g}}_+)_p$, $(\hat{\underline{g}}_-)_p$ their closures with respect to $\| \ \|_p$. One has

$$(\hat{\underline{g}})_p = (\hat{\underline{g}}_-)_p \oplus (\hat{\underline{g}}_+)_p. \tag{4.5}$$

Since $\tau^{(0)}_l(t)$ is assumed to be nonzero, the left-hand side of (4.2) is already decomposed according to (4.5); for the right-hand side we write

$$\hat{R}^l(p_i) = \hat{R}^l_-(p_i) + \hat{R}^l_+(p_i). \tag{4.6}$$

Now (4.2) is equivalent to the following linear equations:

$$[\partial_{t_i} + \hat{R}^l_-(p_i)](\hat{\psi}^l_-)^{-1} = 0, \quad i = 1,2,\ldots, \tag{4.7a}$$

$$[\partial_{t_i} - \hat{R}^l_+(p_i)]\hat{\psi}^l_{0,+} = 0. \tag{4.7b}$$

Define covariant derivatives by

$$\hat{D}^l_{t_i} = \partial_{t_i} - \hat{R}^l_+(p_i), \quad i = 1,2,\ldots . \tag{4.8}$$

The compatibility conditions for (4.7b) are the zero curvature equations

$$[\hat{D}^l_{t_i}, \hat{D}^l_{t_j}] = 0, \quad i,j = 1,2,\ldots . \tag{4.9}$$

One easily checks that

$$[\hat{D}^l_{t_i}, \hat{R}^l(p_j)] = 0, \quad i,j = 1,2,\ldots . \tag{4.10}$$

Hence the elements $\hat{R}^l(p_j)$, defined by (4.3), are (up to multiplication by a power of λ) resolvents for $\hat{D}^l_{t_i}$ in the sense of Dickey.[15]

Besides the differential equations (4.1), the operator $\hat{\psi}^l$ also satisfies the difference equation

$$\hat{\psi}^{l+1} = T^{-1}\hat{\psi}^l. \tag{4.11}$$

Using the factorization (3.13) this may be rewritten as

$$(\hat{\psi}^l_-)^{-1}\hat{\psi}^{l+1}_-\hat{\psi}^{l+1}_{0,+}(\hat{\psi}^l_{0,+})^{-1} = (\hat{\psi}^l_-)^{-1}T^{-1}\hat{\psi}^l_- . \tag{4.12}$$

If we define

$$\begin{aligned} \hat{U}^l &:= (\hat{\psi}^l_-)^{-1}T^{-1}\hat{\psi}^l_-, \\ \hat{U}^l_- &:= (\hat{\psi}^l_-)^{-1}\hat{\psi}^{l+1}_-, \\ \hat{U}^l_+ &:= (\hat{\psi}^{l+1}_-)^{-1}T^{-1}\hat{\psi}^l_- = \hat{\psi}^{l+1}_{0,+}(\hat{\psi}^l_{0,+})^{-1}, \end{aligned} \tag{4.13}$$

we have the factorization

$$\hat{U}^l = \hat{U}^l_-\hat{U}^l_+ . \tag{4.14}$$

Note that, if $\tau^{(0)}_{l+1}(t) \neq 0$, the factorization (4.14) has again nice properties, i.e., on the level of the loop group we have

$$\begin{aligned} &\tilde{U}^l = \tilde{U}^l_-\tilde{U}^l_+ , \\ &\tilde{U}^l_- = 1 + A_1\lambda^{-1} + \cdots, \quad \tilde{U}^l_+ = B_0 + B_1\lambda + \cdots . \end{aligned} \tag{4.15}$$

Moreover, $\tilde{U}^l_-$ and $\tilde{U}^l_+$ depend differentiably on t.

From the definitions (4.13) of $\tilde{U}^l_-$ and $\tilde{U}^l_+$ we read off the difference equations

$$\hat{\psi}^{l+1}_- = \hat{\psi}^l_-\tilde{U}^l_- , \tag{4.16a}$$

$$\tilde{\psi}^{l+1}_{0,+} = \tilde{U}^l_+\hat{\psi}^l_{0,+} . \tag{4.16b}$$

Equation (4.16b) suggests the introduction of a lattice covariant derivative $\hat{D}^l$, which is defined on group valued fields $\hat{\phi}^l(t)$ by

$$\hat{D}^l\hat{\phi}^l = (\hat{U}^l_+)^{-1}\hat{\phi}^{l+1} - \hat{\phi}^l. \tag{4.17}$$

From (4.16b) we see that $\hat{\psi}^l_{0,+}$ is covariantly constant with respect to this derivative. This equation may therefore be considered to be the discrete analog of (4.7b).

The compatibility of (4.7b) and (4.6b) gives the following differential difference zero curvature equations (see Ref. 7):

$$\hat{R}^{l+1}_+(p_i) = \hat{U}^l_+\hat{R}^l_+(p_i)(\hat{U}^l_+)^{-1} + (\partial_{t_i}\hat{U}^l_+)(\hat{U}^l_+)^{-1}. \tag{4.18}$$

We stress that, although (4.16a) and (4.16b) are always valid, (4.18) can only be derived if both $\tau^{(0)}_l$ and $\tau^{(0)}_{l+1}$ are nonzero.

One easily checks that

$$\hat{R}^l(p_i) = \hat{U}^l\hat{R}^l(p_i)(\hat{U}^l)^{-1}, \tag{4.19}$$

which shows that (4.19) is essentially the lattice resolvent, introduced in Ref. 7 (there we used $\tilde{U}^l$ rather than $\hat{U}^l$). In that paper we could only verify the factorization (4.15) by explicit calculation. Here it is a simple consequence of the specific form of the Birkhoff decomposition for $\hat{\psi}^l$, $\hat{\psi}^{l+1}$ following from the assumptions $\tau^{(0)}_l \neq 0$ and $\tau^{(0)}_{l+1} \neq 0$.

C. Zero curvature equations on $(\underline{g})_p$

In this section we calculate the central component of the zero curvature equations (4.9) and (4.18) and we will see that it is trivially satisfied. First we calculate the central term of $\hat{R}^l_+(p_i)$, which we denote by c^l_i;

$$c_i^l := \pi_*(\hat{R}^l{}_+(p_i)). \tag{4.20}$$

Here π_* is the projection on the center of the Kac–Moody algebra. Using (4.7b) we may write

$$\pi_*(\hat{R}^l{}_+(p_i)) = \pi_*((\partial_{t_i}\hat{\psi}^l_{0,+})(\hat{\psi}^l_{0,+})^{-1}) \tag{4.21}$$

and since the center is contained in $\underline{\hat{h}}$ we obtain

$$c_i^l = \pi_*((\partial_{t_i}\hat{\psi}^l_0)(\hat{\psi}^l_0)^{-1}). \tag{4.22}$$

For $\tau_l^{(0)} \neq 0$ (3.14) reads

$$\hat{\psi}^l_0 = \exp\{a^l f\} w \exp\{\lambda^l{}_0 c + (\lambda^l{}_1 - \lambda^l{}_0)\alpha_1^v\}\exp\{b^l e\}, \tag{4.23}$$

where we have used $c = \alpha_0^v + \alpha_1^v$. Substitution in (4.22) yields

$$c_i^l = (\partial_{t_i}\lambda^l{}_0)c = (\partial_{t_i}\ln\tau_l^{(0)})c. \tag{4.24}$$

Since the two-cocycle, defining the central extension, is zero on $(\hat{g}_+)_p$, the central component of (4.9) becomes

$$\partial_{t_i}c_j^l = \partial_{t_j}c_i^l, \tag{4.25}$$

which is trivial in view of (4.24).

Similarly we find, (see 4.15),

$$\pi_*((\partial_{t_i}\hat{U}^l{}_+)(\hat{U}^l{}_+)^{-1}) = (\partial_{t_i}\ln\tau_{l+1}^{(0)} - \partial_{t_i}\ln\tau_l^{(0)})c \tag{4.26}$$

and the central component of (4.18) reads

$$c_i^{l+1} = c_i^l + (\partial_{t_i}\ln\tau_{l+1}^{(0)} - \partial_{t_i}\ln\tau_l^{(0)})c, \tag{4.27}$$

which is again trivially satisfied.

From the discussion above it is clear that we may project the zero curvature equations (4.9) and (4.18) on the loop algebra without loss of information. More generally, all objects (and relations between them) introduced in Sec. IV B have, by projection, their counterparts on the loop group and loop algebra. We will denote them by the same symbols with the hat $\hat{}$ replaced by a tilde $\tilde{}$ and will refer to them by their formula number in Sec. IV B.

V. THE RESOLVENT

A. Introduction

In Sec. IV we saw that the central component of the zero curvature equations carries no information and that these equations may therefore just as well be formulated on the loop algebra. This does not mean that the center is unimportant; in this section we will use the complete affine algebra, or rather its basic representation $L(\Lambda_0)$, to compute the resolvent $\tilde{R}^l(p_i)$ explicitly.

In Sec. V B we will find an expression for $\tilde{R}^l(p_i)$ in terms of the components of the τ function; the final formula we obtain was found earlier by Flaschka, Newell, and Ratiu.[8] However, these authors introduce the τ function in a completely different context (see Sec. VI) and they—as they stress themselves—do not have a Lie algebraic interpretation for it.

In Sec. V C we express the resolvent in terms of differential polynomials in certain fundamental fields. In terms of these fields the zero curvature equations (4.9) and (4.18) become a system of AKNS equations on a lattice coupled by Toda equations.[7] This connection between Toda and AKNS equations was discovered before (see Refs. 16 and 17) by considering certain Bäcklund transformations for the AKNS system. Our construction is less *ad loc* and shows that it is natural to think of Toda and AKNS equations as members of one family, the Toda–AKNS hierarchy.

B. The resolvent in terms of τ functions

The simplest of the covariant derivatives (4.8) is the one in the $t_1 =: x$ direction. It has the form

$$\tilde{D}^l_x = \partial_x - \tilde{R}^l{}_+(p_1) = \partial_x - \lambda\otimes\frac{h}{2} - \begin{pmatrix} 0 & q^l \\ r^l & 0 \end{pmatrix}. \tag{5.1}$$

We will now derive expressions for q^l and r^l in terms of τ functions.

First consider q^l; it is the coefficient of $\lambda^0\otimes e$ in the root space decomposition of $\tilde{R}^l(p_1) = (\tilde{\psi}^l_-)^{-1}\lambda\otimes h/2\tilde{\psi}^l_-$ and therefore it is also the coefficient of $\lambda^{-1}\otimes e$ in the decomposition of $(\tilde{\psi}^l_-)^{-1}(h/2)\tilde{\psi}^l_-$. Since $(\tilde{\psi}^l_-)^{-1}(h/2)\tilde{\psi}^l_-$ and $(\hat{\psi}^l_-)^{-1}(h/2)\hat{\psi}^l_-$ differ only by a central term we may write

$$\begin{aligned} q^l &= H(\lambda^{-1}\otimes e\cdot v_0, (\hat{\psi}^l_-)^{-1}(h/2)\hat{\psi}^l_-\cdot v_0) \\ &\quad\times[H(\lambda^{-1}\otimes e\cdot v_0, \lambda^{-1}\otimes e\cdot v_0)]^{-1} \\ &= H(f_0\cdot v_0, (\hat{\psi}^l_-)^{-1}(h/2)\hat{\psi}^l_-\cdot v_0) \\ &= H((h/2)\cdot(\hat{\psi}^l_-)^{-1\dagger}\cdot f_0\cdot v_0, \hat{\psi}^l_-\cdot v_0). \end{aligned} \tag{5.2}$$

Here the dagger † means Hermitian conjugation as usual; if $(\hat{\psi}^l_-)^{-1}$ is an expression of the form $\exp\pi_{\Lambda_0}(x_1)\cdots\exp\pi_{\Lambda_0}(x_n)$, we have, using (2.5),

$$(\hat{\psi}^l_-)^{-1\dagger} = \exp\{-\pi_{\Lambda_0}(\omega_0(x_n))\}\cdots\exp\{-\pi_{\Lambda_0}(\omega_0(x_1))\}. \tag{5.3}$$

Since $(\hat{\psi}^l_-)^{-1}$ belongs to $(\hat{U}_-)_p$, its conjugate belongs to $(\hat{U}_+)_p$ and we may write

$$(\hat{\psi}^l_-)^{-1\dagger} = 1 + A_+, \tag{5.4}$$

where A_+ is an operator, which raises the weight of a vector it acts on. The vector $f_0\cdot v_0$ has weight $\Lambda_0 - \alpha_0$ and the only weight higher than this is Λ_0 itself. Therefore we have

$$(\hat{\psi}^l_-)^{-1\dagger}\cdot f_0\cdot v_0 = f_0\cdot v_0 + \mu v_0, \tag{5.5}$$

where μ is some complex number.

We now calculate

$$\begin{aligned} &(h/2)(\hat{\psi}^l_-)^{-1\dagger}\cdot f_0\cdot v_0 \\ &\quad= \tfrac{1}{2}\alpha_1^v\cdot f_0\cdot v_0 + \tfrac{1}{2}\mu\alpha_1^v\cdot v_0 \\ &\quad= \tfrac{1}{2}\langle\Lambda_0 - \alpha_0, \alpha_1^v\rangle f_0\cdot v_0 + \tfrac{1}{2}\mu\langle\Lambda_0, \alpha_1^v\rangle v_0 = f_0\cdot v_0. \end{aligned} \tag{5.6}$$

Using the definition of the translation operator T and the weight system $P(\Lambda_0)$ (3.7), we readily find

$$T\cdot v_0 = -f_0\cdot v_0 \tag{5.7}$$

and hence

$$\begin{aligned} q^l &= -H(T\cdot v_0, \hat{\psi}^l_-\cdot v_0) \\ &= -H(v_0, T^{-1}\hat{\psi}^l(\hat{\psi}^l_{0,+})^{-1}v_0). \end{aligned} \tag{5.8}$$

If $\tau_l^{(0)} \neq 0$ this becomes

$$\begin{aligned} q^l &= -H(v_0, T^{-1}\hat{\psi}^l v_0)\cdot e^{-\lambda^l_0} \\ &= -H(v_0, \hat{\psi}^{l+1}v_0)e^{-\lambda^l_0} = -\tau_{l+1}^{(0)}/\tau_l^{(0)}. \end{aligned} \tag{5.9}$$

An analogous calculation yields

$$r^l = \tau_{l-1}^{(0)}/\tau_l^{(0)}. \tag{5.10}$$

Next we consider the complete resolvent $\tilde{R}^l(p_i)$; we write it as

$$\tilde{R}^l(p_1) = \lambda \otimes \frac{h}{2} + \sum_{i=0}^{\infty} \lambda^{-i} \otimes R_i^l. \tag{5.11}$$

We already know

$$R_0^l = \begin{pmatrix} 0 & q^l \\ r^l & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\tau_{l+1}^{(0)}/\tau_l^{(0)} \\ \tau_{l-1}^{(0)}/\tau_l^{(0)} & 0 \end{pmatrix}.$$

For the other matrices R_i^l we write

$$\begin{aligned} R_i^l &= \operatorname{res}(\tilde{R}^l(p_i)) = \operatorname{res}(\tilde{R}^l{}_-(p_i)) \\ &= \operatorname{res}((\tilde{\psi}^l{}_-)^{-1}\partial_{t_i}\tilde{\psi}^l{}_-), \quad i = 1,2,\ldots. \end{aligned} \tag{5.12}$$

Here we have used (4.7a).

The operator $\hat{\psi}^l{}_-$ may be written as a series

$$\begin{aligned} \hat{\psi}^l{}_- &= 1 + b^l \pi_{\Lambda_0}(\lambda^{-1}\otimes e) + a^l \pi_{\Lambda_0}(\lambda^{-1}\otimes h) \\ &\quad + c^l \pi_{\Lambda_0}(\lambda^{-1}\otimes f) + \cdots. \end{aligned} \tag{5.13}$$

Hence after projection we obtain

$$\ddot{\tilde{\psi}}^l{}_- = 1 + \lambda^{-1}\begin{pmatrix} a^l & b^l \\ c^l & -a^l \end{pmatrix} + \cdots \tag{5.14}$$

and we find

$$R_i^l = \begin{pmatrix} \partial_{t_i} a^l & \partial_{t_i} b^l \\ \partial_{t_i} c^l & -\partial_{t_i} a^l \end{pmatrix}. \tag{5.15}$$

Using (5.13) we calculate the coefficients a^l, b^l, c^l;

$$b^l = H(\lambda^{-1}\otimes e\cdot v_0, \hat{\psi}^l{}_-\cdot v_0) = q^l, \tag{5.16}$$

$$c^l = H(\lambda^{-1}\otimes f\cdot v_0, \hat{\psi}^l{}_-\cdot v_0) = -r^l, \tag{5.17}$$

$$\begin{aligned} a^l &= H(\lambda^{-1}\otimes h\cdot v_0, \hat{\psi}^l{}_-\cdot v_0) \\ &\quad \times [H(\lambda^{-1}\otimes h\cdot v_0, \lambda^{-1}\otimes h\cdot v_0)]^{-1} \\ &= H(\lambda^{-1}\otimes (h/2)\cdot v_0, \hat{\psi}^l{}_-\cdot v_0) \\ &= H(v_0, p_1 \hat{\psi}^l (\hat{\psi}^l_{0,+})^{-1}\cdot v_0) \\ &= H(v_0, (\partial_x \hat{\psi}^l)(\hat{\psi}^l_{0,+})^{-1}\cdot v_0) \\ &= H(v_0, \{\partial_x \hat{\psi}^l{}_- + \hat{\psi}^l{}_- (\partial_x \hat{\psi}^l_{0,+})(\hat{\psi}^l_{0,+})^{-1}\}\cdot v_0) \\ &= H(v_0, (\partial_x \hat{\psi}^l_{0,+})(\hat{\psi}^l_{0,+})^{-1}\cdot v_0) \\ &= H(v_0, c_1^l \cdot v_0) \quad [\text{see (4.22)}] \\ &= \partial_x \ln \tau_l^{(0)}. \end{aligned} \tag{5.18}$$

So we finally obtain

$$R_i^l = \begin{pmatrix} \partial_{t_i}\partial_x \ln \tau_l^{(0)} & \partial_{t_i} q^l \\ \partial_{t_i} r^l & -\partial_{t_i}\partial_x \ln \tau_l^{(0)} \end{pmatrix}, \quad i = 1,2,\ldots. \tag{5.19}$$

Note that, since $\tau_l^{(0)}$ depends only on a finite number of times, $R_i^l = 0$, $\forall i > N$. This means that the resolvent $\tilde{R}^l(p_1)$ is an element of the polynomial loop algebra $\tilde{\underline{g}}$.

For the sake of completeness we also give the expression for the positive part of the lattice resolvent in terms of τ functions;

$$\begin{aligned} \tilde{U}^l{}_+ &= (\tilde{\psi}^{l+1}_-)^{-1} T^{-1} \tilde{\psi}^l{}_- \\ &= \begin{pmatrix} \lambda - \partial_x \ln(\tau_l^{(0)}/\tau_{l+1}^{(0)}) & -\tau_{l+1}^{(0)}/\tau_l^{(0)} \\ \tau_l^{(0)}/\tau_{l+1}^{(0)} & 0 \end{pmatrix} \\ &= \begin{pmatrix} \lambda - \partial_x \ln q^l & q^l \\ -1/q^l & 0 \end{pmatrix}. \end{aligned} \tag{5.20}$$

C. The resolvent as a differential polynomial

In the usual theory of resolvents (see, e.g., Dickey[15]) the matrix coefficients of resolvents are polynomials in some set of fundamental fields and their derivatives with respect to some variable; one may choose, e.g., q^l and r^l as fundamental fields and x as the special variable but other choices are also possible, as we will see in this section. The explicit form of the resolvent derived in Sec. V B does not look like this at all, since all t_i derivatives appear and the diagonal terms are of the form $\partial_{t_i}\partial_x \ln \tau_l^{(0)}$. One knows however, that the τ functions satisfy many equations and (implicitly) using these, one finds that resolvents may indeed be expressed in terms of differential polynomials, as stated above.

We start with a factorization for $\tilde{\psi}^l{}_- \in \tilde{G}_p$. It may be written as

$$\tilde{\psi}^l{}_- = 1 + A_1^l \lambda^{-1} + A_2^l \lambda^{-2} + \cdots = 1 - A^l, \tag{5.21}$$

where

$$A^l := -\sum_{i\geq 1} A_i^l \lambda^{-i} \in \tilde{G}_p.$$

We can define the logarithm of $\tilde{\psi}^l{}_-$ by

$$n^l := \log \tilde{\psi}^l{}_- := \sum_{k\geq 1} \frac{1}{k} (A^l)^k \tag{5.22}$$

and we have

$$\tilde{\psi}^l{}_- = \exp n^l. \tag{5.23}$$

We stress that n^l is, in general, not in any of the completions $(\tilde{\underline{n}}_-)_p$, but ought to be considered as a formal power series with traceless matrices as coefficients;

$$n^l = \sum_{i\geq 1} n_i^l \lambda^{-i}, \quad \operatorname{tr} n_i^l = 0. \tag{5.24}$$

The Lie algebra $\underline{g}$ may be decomposed as

$$\underline{g} = \underline{h} \oplus \underline{h}^{\perp}, \tag{5.25}$$

where $\underline{h}^{\perp} = \oplus_{\alpha\in\Delta} \underline{g}_{\alpha}$ is of course just a subspace, not a subalgebra. Associated with this decomposition we have a factorization of $\tilde{\psi}^l{}_-$.

Lemma 5.1: There exist unique formal power series

$$s^l = \sum_{i\geq 1} s_i^l \lambda^{-i}, \quad s_i^l \in \underline{h},$$

and

$$k^l = \sum_{i\geq 1} k_i^l \lambda^{-i}, \quad k_i^l \in \underline{h}^{\perp},$$

such that

$$\tilde{\psi}^l{}_- = \exp s^l \cdot \exp k^l. \tag{5.26}$$

Proof: We have

$$\exp s^l \cdot \exp k^l = \exp \sum_{n>1} c_n(s^l, k^l), \tag{5.27}$$

where the c_n's are the well-known Campbell–Baker–Hausdorf expressions

$$\begin{aligned} &c_1(s^l,k^l) = s^l + k^l, \quad c_2(s^l,k^l) = \tfrac{1}{2}[s^l,k^l], \\ &c_3(s^l,k^l) = \tfrac{1}{12}\{[s^l,[s^l,k^l]] + [k^l,[k^l,s^l]]\}, \text{ etc.} \end{aligned} \tag{5.28}$$

(See, e.g., Varadarajan.[18])

According to (5.23) we should try to solve s^l and k^l from

$$\sum_{n>1} c_n(s^l,k^l) = n^l. \tag{5.29}$$

The λ^{-i} coefficient of this equation reads

$$s_i^l + k_i^l + C_i(s_1^l \cdots s_{i-1}^l, k_1^l \cdots k_{i-1}^l) = n_i^l, \tag{5.30}$$

where C_i is a complicated expression in commutators of the arguments indicated. Using (5.25) we can, if $s_1^l \cdots s_{i-1}^l$, $k_1^l \cdots k_{i-1}^l$ are already determined, find $s_i^l \in \underline{h}$ and $k_i^l \in \underline{h}^\perp$ in exactly one way. □

The formal power series s^l and k^l are both important; in this subsection we will use k^l to express the resolvent in terms of differential polynomials. The significance of s^l will be discussed in Sec. VI.

The following lemma states that a gauge transformation by $\exp k^l$ "diagonalizes" the covariant derivatives $\tilde{D}_{t_i}^l$ (4.8).

Lemma 5.2:

$$e^{k^l}(\partial_{t_i} - R^l{}_+(p_i))e^{-k^l} = \partial_{t_i} - p_i + \partial_{t_i} s^l. \tag{5.31}$$

Proof: By (4.2) and (4.7b) we have

$$\begin{aligned} p_i &= (\partial_{t_i}\tilde{\psi}_-^l)(\tilde{\psi}_-^l)^{-1} \\ &\quad + \tilde{\psi}_-^l((\partial_{t_i}\tilde{\psi}_{0,+}^l)(\tilde{\psi}_{0,+}^l)^{-1})(\tilde{\psi}_-^l)^{-1} \\ &= (\partial_{t_i}\tilde{\psi}_-^l)(\tilde{\psi}_-^l)^{-1} + \tilde{\psi}_-^l \tilde{R}^l{}_+(p_i)(\tilde{\psi}_-^l)^{-1}. \end{aligned} \tag{5.32}$$

Substituting the factorization (5.26) and using the fact that $\exp s^l$ centralizes p_i one easily obtains the lemma. □

The following proposition is due to Drinfeld and Sokolov.[3]

Proposition 5.3: Let D_t be a covariant derivative of the form

$$D_t = \partial_t - B_t = \partial_t - \lambda^i \frac{h}{2} - \sum_{m=1}^{i+j} \lambda^{i-m} v_m, \tag{5.33}$$

where the v_m's are functions of t with values in $\underline{g} = \mathrm{sl}(2,\mathbb{C})$. Then there exists a unique formal series

$$k = \sum_{i>1} k_i \lambda^{-i}, \quad k_i \in \underline{h}^\perp,$$

such that

$$\exp(\mathrm{ad}\, k)(D_t) = \partial_t - \lambda^i \frac{h}{2} - \sum_{m>1} \lambda^{i-m} h_m, \tag{5.34}$$

where $h_m \in \underline{h}$. Moreover, the matrix coefficients of the k_i's and h_m's are polynomials in the matrix coefficients of the v_m's and their t derivatives.

Combining Lemma 5.2 and Proposition 5.3, we conclude that the matrix coefficients of the k_j^l's are polynomials in the matrix coefficients of $\tilde{R}^l{}_+(p_i)$ and their t_i derivatives, for arbitrary i. We will call such polynomials and, somewhat inaccurately, also matrices with these polynomials as coefficients t_i-differential polynomials.

Since k^l is a t_i-differential polynomial for all i, the same holds for the resolvents $\tilde{R}^l(p_j)$ and the lattice resolvent $\tilde{U}^l$; indeed we have

$$\tilde{R}^l(p_j) = (\tilde{\psi}_-^l)^{-1} p_j \tilde{\psi}_-^l = e^{-k^l} e^{-s^l} p_j e^{s^l} e^{k^l} = e^{-k^l} p_j e^{k^l} \tag{5.35}$$

and

$$\tilde{U}^l = (\tilde{\psi}_-^l)^{-1} T^{-1} \tilde{\psi}_-^l = e^{-k^l} e^{-s^l} T^{-1} e^{s^l} e^{k^l} = e^{-k^l} T^{-1} e^{k^l}, \tag{5.36}$$

where we have used that e^{s^l} centralizes both p_i and T.

Taking $i = 1$, e.g., we find that $\tilde{R}^l(p_j)$ and $\tilde{U}^l$ are x-differential polynomials, i.e., that their matrix coefficients are polynomial expressions in the fields q^l, r^l and their x-derivatives. In terms of these fields the zero curvature equations (4.9) are by definition the AKNS family formulated on a lattice, $L \simeq \mathbb{Z}$. The first two members of this family are

$$\begin{aligned} &[\tilde{D}_{t_2}^l, \tilde{D}_x^l] = 0 \Leftrightarrow \begin{cases} \partial_{t_2} q^l = \partial_x^2 q^l - 2(q^l)^2 r^l, \\ \partial_{t_2} r^l = -\partial_x^2 r^l + 2q^l (r^l)^2 \end{cases} \\ &\qquad \text{(nonlinear Schrödinger)}, \\ &[\tilde{D}_{t_3}^l, \tilde{D}_x^l] = 0 \Leftrightarrow \begin{cases} \partial_{t_3} q^l = \partial_x^3 q^l - 6q^l r^l \partial_x q^l, \\ \partial_{t_3} r^l = \partial_x^3 r^l - 6q^l r^l \partial_x r^l \end{cases} \\ &\qquad \text{(modified KdV)}. \end{aligned} \tag{5.37}$$

The differential difference zero curvature equations (4.18) tell us how the fields in different lattice points are related. It turns out that all information is contained in the case $i = 1$, which gives the Toda equations

$$q^{l+1} = -(q^l)^2 r^l + q^l \partial_x^2 \ln q^l, \quad r^{l+1} = -1/q^l \tag{5.38}$$

(see Ref. 7 for a more detailed discussion).

As mentioned in the Introduction, we will refer to the complete system of differential difference equations for q^l and r^l as the Toda–AKNS family.

We may summarize this section by the following simple proposition.

Proposition 5.4: Let $\hat{g} \in \hat{G}^{\Lambda_0}$ and

$$\tau^{(0)} := \hat{g} \cdot v_{\Lambda_0} = \sum_{l \in \mathbb{Z}} \tau_l^{(0)}(x) \otimes e^{l\alpha}$$

be its associated τ function. Then $q^l := -\tau_{l+1}^{(0)}/\tau_l^{(0)}$ and $r^l := \tau_{l-1}^{(0)}/\tau_l^{(0)}$ are solutions of the Toda–AKNS family.

VI. CONSERVATION LAWS

A. Introduction

In this section we discuss conservation laws for the Toda–AKNS system. For the AKNS family two different constructions of conserved densities exist. The first one is due to Drinfeld and Sokolov[3] and comes down to a diagonalization of the zero curvature equations by performing a gauge transformation by $\exp k^l$. The second construction expresses the conserved densities in terms of the resolvents. This method may be found in Dickey.[15] Flaschka, Newell, and Ratiu[8] noted the connection with τ functions. In Sec. VI B we will compare the conserved densities obtained by these different methods and we will see that they are essen-

tially identical. For a discussion of various constructions of conservation laws in the context of Lax equations we refer to Wilson.[19]

If one considers the AKNS system on a lattice, the conserved densities will of course not only depend on the times t_i, $i \geqslant 1$, but also on a discrete index $l \in \mathbb{Z}$. In Sec. VI C we will investigate the discrete evolution, which will lead to the formulation of discrete conservation laws. Finally we will see that, taking a different point of view, one may also obtain conserved quantities for the Toda lattice from these discrete conservation laws.

B. AKNS conservation laws

Let f be a t_i-differential polynomial for some fixed choice of t_i. The t_j derivative of f can be calculated from the t_j derivatives of the generators of t_i-differential polynomials, i.e., the coefficients of $\tilde{R}^l{}_+(p_i)$; these follow from the zero curvature equations $[\tilde{D}^l_{t_i}, \tilde{D}^l_{t_j}] = 0$, which can be rewritten as

$$\partial_{t_j} \tilde{R}^l{}_+(p_i) = [\tilde{D}^l_{t_i}, \tilde{R}^l{}_+(p_j)]. \tag{6.1}$$

Since $\tilde{R}^l{}_+(p_j)$ is a t_i-differential polynomial, we deduce from (6.1) that the t_j derivative of a t_i-differential polynomial is again a t_i-differential polynomial.

We will call a t_i-differential polynomial a t_i-conserved density if there exists for all $j \geqslant 1$ a t_i-differential polynomial g_j, such that

$$\partial_{t_j} f = \partial_{t_i} g_j. \tag{6.2}$$

We will refer to (6.2) as a t_i-conservation law.

It is easy to construct such t_i-conservation laws; define

$$h^l_j := \partial_{t_j} s^l, \tag{6.3}$$

where s^l is the formal power series defined in (5.26). Although s^l is not a t_i-differential polynomial, it is easily seen from (5.31) (with i replaced by j) that h^l_j is a t_i-differential polynomial. The obvious identity

$$\partial_{t_j} h^l_i = \partial_{t_i} h^l_j \tag{6.4}$$

tells us that h^l_i is a t_i-conserved density. Expanding (6.4) in powers of λ^{-1}, one finds an infinite sequence of t_i-conservation laws. Using (5.31), the reader easily checks that (6.4) is nothing else but the zero curvature condition $[\tilde{D}^l_{t_i}, \tilde{D}^l_{t_j}] = 0$ after a gauge transformation by $\exp k^l$. This is the construction of Drinfeld and Sokolov.[3]

Now we will discuss the second derivation of conservation laws. Remember from Sec. IV that the central component of the resolvent $\hat{R}^l(p_i)$ is given by

$$c^l_i = \pi_*(\mathrm{Ad}(\hat{\psi}^l_-)^{-1}(p_i)) = (\partial_{t_i} \ln \tau^{(0)}_l) \cdot c. \tag{6.5}$$

For the adjoint action of the Kac–Moody group $\hat{G}_p$ on the Kac–Moody algebra $(\hat{\mathfrak{g}})_p$ one has the following well-known formula:

$$\mathrm{Ad}\,\hat{g}(\tilde{x} + \alpha c) = \mathrm{Ad}\,\tilde{g}(\tilde{x}) + \{\mathrm{tr\,res}(\tilde{g}^{-1}(\partial_\lambda \tilde{g})\tilde{x}) + \alpha\}c,$$
$$\forall \hat{g} \in \hat{G}_p, \quad \tilde{x} \in (\hat{\mathfrak{g}})_p, \quad \alpha \in \mathbb{C} \tag{6.6}$$

(see, e.g., Frenkel[20]).

Here $\tilde{g}$ denotes, of course, the image of $\hat{g}$ under the projection $\hat{G}_p \to \tilde{G}_p$.

Using this formula we can express the quantity

$$F_{ij} := \partial_{t_i} \partial_{t_j} \ln \tau^{(0)}_l \tag{6.7}$$

in terms of resolvents:

$$\begin{aligned}
F_{ij} &= \partial_{t_j} \mathrm{tr\,res}(\tilde{\psi}^l_- (\partial_\lambda (\tilde{\psi}^l_-)^{-1}) p_i) \\
&= \mathrm{tr\,res}(\{(\partial_{t_j} \tilde{\psi}^l_-)(\partial_\lambda (\tilde{\psi}^l_-)^{-1}) \\
&\quad + \tilde{\psi}^l_- (\partial_\lambda \partial_{t_j} (\tilde{\psi}^l_-)^{-1})\} p_i) \\
&= \mathrm{tr\,res}(\{\tilde{\psi}^l_- \tilde{R}^l{}_-(p_j)(\partial_\lambda (\tilde{\psi}^l_-)^{-1}) \\
&\quad - \tilde{\psi}^l_- \partial_\lambda (-\tilde{R}^l{}_-(p_j)(\tilde{\psi}^l_-)^{-1})\} p_i) \\
&= -\mathrm{tr\,res}(\tilde{\psi}^l_- \partial_\lambda (\tilde{R}^l{}_-(p_j))(\tilde{\psi}^l_-)^{-1} p_i) \\
&= -\mathrm{tr\,res}(\partial_\lambda (\tilde{R}^l{}_-(p_j)) \tilde{R}^l(p_i)) \\
&= \mathrm{tr\,res}(\tilde{R}^l{}_-(p_j) \partial_\lambda (\tilde{R}^l(p_i))) \\
&= \mathrm{tr\,res}(\tilde{R}^l(p_j) \partial_\lambda (\tilde{R}^l{}_+(p_i))).
\end{aligned} \tag{6.8}$$

Here we have used (4.7a) and some obvious properties of the trace and the residue.

From the final expression it is clear that F^l_{ij} is a t_i-differential polynomial. Then the identity

$$\partial_{t_k} F^l_{ij} = \partial_{t_i} F^l_{jk} \tag{6.9}$$

shows that F^l_{ij} is for all j a t_i-conserved density.

Remark: By studying the structure of the AKNS equations, Flaschka, Newell, and Ratiu are led to introduce the quantities $F_{ij} := \mathrm{tr\,res}(\tilde{R}(p_j) \partial_\lambda \tilde{R}_+(p_i))$. They proceed to prove that the F_{ij}'s are symmetric in i and j and that $\partial_{t_k} F_{ij}$ is totally symmetric in i, j, and k. This enables them to define the τ function as a potential; $F_{ij} =: \partial_{t_i} \partial_{t_j} \ln \tau$. From this definition, however, it is not at all clear that τ has anything to do with the representation theory of the group $\hat{G}$.

The relation of F^l_{ij} with the conserved densities obtained from expanding h^l_i is as follows; substituting the factorization $\tilde{\psi}^l_- = e^{s^l} \cdot e^{k^l}$ in (6.5) we find

$$\partial_{t_j} \ln \tau^{(0)}_l = \mathrm{tr\,res}(\{\exp k^l \partial_\lambda (\exp(-k^l)) - \partial_\lambda s^l\} p_j). \tag{6.10}$$

This gives, for F^l_{ij},

$$F^l_{ij} = \partial_{t_i} K^l_j - \mathrm{tr\,res}((\partial_\lambda h^l_i) p_j), \tag{6.11}$$

where

$$K^l_j = \mathrm{tr\,res}(\exp k^l \partial_\lambda (\exp(-k^l)) p_j) \tag{6.12}$$

is a t_i-differential polynomial. If we substitute

$$h^l_i = \sum_{j>0} h^l_{ij} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \lambda^{-j} \tag{6.13}$$

in (6.11) we find

$$F^l_{ij} = j \cdot h^l_{ij} + \partial_{t_i} K^l_j. \tag{6.14}$$

So the two constructions of t_i-conserved densities differ by a total t_i derivative, which is not considered to be essential.

C. Discrete conservation laws

In the previous section we found that h^l_i is constant in all times modulo total t_i-derivatives of t_i-differential polynomials. A natural question to ask is how h^l_i changes under the

discrete evolution $l \to l+1$. For this purpose we introduce a lattice derivative Δ by

$$\Delta a^l := a^{l+1} - a^l. \tag{6.15}$$

We then have

$$\Delta h_i^l = \Delta(\partial_{t_i} s^l) = \partial_{t_i}(\Delta s^l). \tag{6.16}$$

If we want to interpret (6.16) as a discrete conservation law, we have to investigate whether Δs^l is a t_i-differential polynomial. From the definition (4.15) of $\tilde{U}^l{}_+$ and the factorization (5.26) of $\tilde{\psi}^l{}_-$ we obtain

$$\exp(-\Delta s^l) = \exp k^{l+1} \tilde{U}^l{}_+ \exp k^l T =: 1 - \sum_{i>1} A_i \lambda^{-i}. \tag{6.17}$$

Observing that K^{l+1} is a polynomial in q^{l+1}, r^{l+1}, and their x derivatives and using the expressions (5.38) for q^{l+1}, r^{l+1}, and the explicit form (5.20) of $\tilde{U}^l{}_+$ we find that the matrix coefficients of $A := \Sigma_{i>1} A_i \lambda^{-i}$ are polynomials in q^l, r^l, their x derivatives, and $1/q^l$. Since

$$\Delta s^l = \sum_{k>0} \frac{A^k}{k}, \tag{6.18}$$

the matrix coefficients of Δs^l are of the same form. Using the zero curvature equations, one can express the x derivatives of q^l and r^l in the generators of t_i-differential polynomials and their t_i derivatives, so that the matrix coefficients of Δs^l may also be written as polynomials in these generators, their t_i derivatives, and $1/q^l$. We conclude that Δs^l is not a t_i-differential polynomial and that if we want to read (6.16) as a discrete conservation law we will have to adjust the definition of a conservation law somewhat to allow powers of $1/q^l$ in the right-hand side.

One may, however, also take a different point of view towards (6.16); reading it from the right to the left, it states that Δs^l changes with a total lattice derivative of a t_i-differential polynomial under the t_i evolution. This opens the possibility of constructing conserved quantities for the Toda–AKNS system by summation over the index l.

To perform this summation we have to study in somewhat more detail the solutions q^l and r^l and the τ function associated to an element $\hat{g} \cdot v_0$ of the orbit of $\hat{G}^{\Lambda_0}$ passing through v_0. Since $\hat{G}^{\Lambda_0}$ is an algebraic group, q^l, r^l, and $\tau_l^{(0)}$ depend only on a finite number of times t_i and $\tau_l^{(0)}$ will be identically zero for almost all l. The next lemma gives some more precise information about this situation.

Lemma 6.1: (a) If q^l, r^l do not depend on t_{n+1}, they are also independent of t_{n+i}, $\forall i > 1$, i.e., the set of i's such that q^l and r^l depend on t_i is an interval $1 \leqslant i \leqslant n$.

(b) The set of l's such that $\tau_l^{(0)}$ is not identically zero is an interval $N \leqslant l \leqslant M$.

(c) At the boundary points N and M determined by (b) the conserved densities h_i^N and h_i^M are identically zero.

Proof: For (a) let q^l and r^l be determined by (5.1) and let them be independent of t_{n+1}. We then have

$$[\tilde{D}_x^l, \tilde{D}_{t_{n+1}}^l] = [\tilde{D}_x^l, -\tilde{R}^l{}_+(p_{n+1})] = 0. \tag{6.19}$$

In other words $\tilde{R}^l{}_+(p_{n+1})$ is a resolvent for $\tilde{D}_x^l$. It is easy to see that all resolvents of $\tilde{D}_x^l$ must be of the form $\exp(-\operatorname{ad} k^l)(p^l)$, where p^l is a constant element of the Heisenberg subalgebra $\underline{\tilde{s}}$, and k^l as in Proposition 5.3. Using this we write

$$\begin{aligned}\tilde{R}^l{}_+(p_{n+1}) &= \exp(-\operatorname{ad} k^l)(p_{n+1}) \\ &\quad - \exp(-\operatorname{ad} k^l)\left(\sum_{i>0} \alpha_i^l \lambda^{-i} h\right) \\ &= \tilde{R}^l(p_{n+1}) \\ &\quad - \exp(-\operatorname{ad} k^l)\left(\sum_{i>0} \alpha_i^l \lambda^{-i} h\right).\end{aligned} \tag{6.20}$$

Here the α_i^l's are constants. Expanding the resolvent $\tilde{R}^l(p_1)$ as in (5.11) and taking the residue of (6.20) we obtain

$$0 = \tilde{R}_{n+1}^l - \alpha_1^l h. \tag{6.21}$$

Using this, the zeroth-order term of the zero curvature equation $[\tilde{D}_x^l, \tilde{D}_{t_{n+2}}^l] = 0$ becomes

$$\partial_{t_{n+2}} \begin{pmatrix} 0 & q^l \\ r^l & 0 \end{pmatrix} = \begin{pmatrix} 0 & 2\alpha_1^l q^l \\ -2\alpha_1^l r^l & 0 \end{pmatrix}, \tag{6.22}$$

solving this we find

$$\begin{aligned} q^l &= e^{2\alpha_1^l t_{n+2}} \cdot q^l(t_{n+2} = 0), \\ r^l &= e^{-2\alpha_1^l t_{n+2}} \cdot r^l(t_{n+2} = 0). \end{aligned} \tag{6.23}$$

Now formulas (5.9) and (5.10) show that q^l and r^l are rational functions of the times t_i. So—provided that q^l and r^l are not identically zero—we must have $\alpha_1^l = 0$ and $\partial_{t_{n+2}} q^l = 0 = \partial_{t_{n+2}} r^l$. As a side result the proof shows that $\tilde{R}^l{}_+(p_{n+1}) = \tilde{R}^l(p_{n+1})$, i.e., that the resolvent is algebraic.

For (b) and (c) we consider a right boundary point of the lattice, i.e., a point M such that $\tau_M^{(0)} \not\equiv 0$, $\tau_{M+1}^{(0)} \equiv 0$. Formula (5.9) shows that $q^M \equiv 0$ and therefore the covariant x derivative in $l = M$ becomes

$$\tilde{D}_x^M = \partial_x - \lambda \frac{h}{2} - \begin{pmatrix} 0 & 0 \\ r^M & 0 \end{pmatrix}. \tag{6.24}$$

This enables us to calculate k^M explicitly.

Remember from Proposition 5.3 that k^M is determined by the condition

$$e^{k^M}(\tilde{D}_x^M) e^{-k^M} \in \underline{\tilde{s}}. \tag{6.25}$$

It is natural to look for a solution of (6.25) of the form

$$k^M = \sum_{i>0} \lambda^{-i} \begin{pmatrix} 0 & 0 \\ f_i^M & 0 \end{pmatrix}. \tag{6.26}$$

If such a solution exists, it is unique by Proposition 5.3. Using this ansatz, we calculate

$$\begin{aligned}&\exp(\operatorname{ad} k^M)(\tilde{D}_x^M) \\ &\quad = \partial_x - \lambda \frac{h}{2} - \lambda k^M - \partial_x k^M - \begin{pmatrix} 0 & 0 \\ r^M & 0 \end{pmatrix}.\end{aligned} \tag{6.27}$$

Using (6.25) we find

$$\lambda k^M = -\partial_x k^M - \begin{pmatrix} 0 & 0 \\ r^M & 0 \end{pmatrix}. \tag{6.28}$$

Equating the coefficients of λ^{-i}, $i \geqslant 0$, we find

$$f_i^M = (-)^i (\partial_x)^{i-1}(r^M), \quad \forall i \geqslant 1. \tag{6.29}$$

We can now compute the resolvent $\tilde{R}^M(p_i)$:

$$\tilde{R}^M(p_i) = e^{-k^M} p_i e^{k^M} = p_i - \lambda^i k^M. \tag{6.30}$$

It has zeroth-order term

$$\tilde{R}_0^M(p_i) = -\begin{pmatrix} 0 & 0 \\ f_i^M & 0 \end{pmatrix} = (-\partial_x)^{i-1}(r^M)\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{6.31}$$

The AKNS equations $[\tilde{D}_{t_i}^M, \tilde{D}_x^M]$ now become very simple;

$$\partial_{t_i} r^M = (-)^{i+1}(\partial_x)^i(r^M), \quad \forall i \geqslant 1. \tag{6.32}$$

Since r^M depends only on the times $t_1, t_2, \ldots, t_n$, we must have

$$(\partial_x)^i(r^M) = 0, \quad \forall i \geqslant n+1. \tag{6.33}$$

Using (6.26), (6.29), and (6.33) we finally find

$$k^M = \sum_{i=1}^{n+1} \lambda^{-i} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} (-)^i (\partial_x)^{i-1}(r^M). \tag{6.34}$$

Remark that k^M is algebraic and nilpotent, which implies that $\exp k^M = 1 + k^M$ belongs to the algebraic group $\mathrm{Sl}_2(\mathbb{C}[\lambda, \lambda^{-1}])$.

It is now easy to calculate $\partial_{t_i} s^M = h_i^M$; with (5.33), (6.26), (6.29), (6.30), and (6.32) we compute

$$\begin{aligned} \partial_{t_i} s^M &= p_i + e^{k^M}(\partial_{t_i} - \tilde{R}_+^M(p_i))e^{-k^M} \\ &= -\partial_{t_i} k^M - (\lambda^i k_M)_- \\ &= -\sum_{j \geqslant 1} \lambda^{-j} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \{\partial_{t_i} f_j + f_{i+j}\} = 0. \end{aligned} \tag{6.35}$$

Since we are working with an algebraic group this result can be sharpened to $s^M \equiv 0$; indeed since $\tilde{\psi}^M(0)$ is algebraic, its factors $\tilde{\psi}_{0,+}^M(0)$ and $\tilde{\psi}_-^M(0) = \exp s^M(0) \exp k^M(0)$ are also algebraic. Above we saw that $\exp k^M$ is algebraic so the same must hold for $\exp s^M(0)$, i.e., it is of the form

$$\exp s^M(0) = \begin{pmatrix} p^M(\lambda^{-1}) & 0 \\ 0 & q^M(\lambda^{-1}) \end{pmatrix}, \tag{6.36}$$

where both $p^M(\lambda^{-1})$ and $q^M(\lambda^{-1}) = 1/p^M(\lambda^{-1})$ are polynomials in λ^{-1} starting with 1. This can only happen if $p^M \equiv 1 \equiv q^M$ and hence $s^M(0) = 0$. Using (6.35) we find $s^M \equiv 0$ as desired. (If we leave the algebraic group we may still derive $\partial_{t_i} s^M = 0$ at a right boundary point M but not $s^M \equiv 0$.)

Finally we consider the Birkhoff decomposition of $\hat{\psi}^{M+i}$ for $i > 0$; we have

$$\begin{aligned} \hat{\psi}^{M+i} = T^{-i}\hat{\psi}^M &= T^{-i}\hat{\psi}_-^M \hat{\psi}_{0,+}^M \\ &= T^{-i}\hat{\psi}_-^M T^i T^{-i}\hat{\psi}_{0,+}^M \\ &= (1 + T^{-i}k^M T^i) T^{-i}\hat{\psi}_{0,+}^M. \end{aligned} \tag{6.37}$$

Now

$$T^{-i}(\lambda^j \otimes f) T^i = \lambda^{j-2i} \otimes f \tag{6.38}$$

and hence $(T^{-i}\hat{\psi}_-^M T^i)^\dagger$ stabilizes v_0. Using this we obtain for $\tau_{M+i}^{(0)}$

$$\tau_{M+i}^{(0)} = H(v_0, T^{-i}\hat{\psi}_{0,+}^M v_0) = H(v_0, T^{-i}v_0)\tau_M^{(0)} = 0. \tag{6.39}$$

The situation at a left boundary point N is similar and will be left to the reader. This completes the proof of the lemma. □

Now return to Eq. (6.16). Summation over $N \leqslant l \leqslant M-1$ yields, using the lemma,

$$\partial_{t_i} S = h_i^M - h_i^N = 0, \tag{6.40}$$

where we have defined

$$S := \sum_{l=N}^{M-1} \Delta s^l. \tag{6.41}$$

So S is a polynomial expression in the fields q^l, r^l, their x derivatives, and $1/q^l$, $N \leqslant l \leqslant M-1$, which is conserved with respect to the t_i evolutions. Note that, in the case of the algebraic group we are working with, the fields q^l and r^l are such that S is identically zero;

$$S = s^M - s^N = 0 \tag{6.42}$$

(see the proof of Lemma 6.1).

Let us now calculate the explicit form of the first two conservation laws contained in (6.40) for $t_i = t_1 = x$. Using (5.31) one finds

$$h_1^l = -q^l r^l \cdot \lambda^{-1} h + \tfrac{1}{2}(q^l \partial_x r^l - r^l \partial_x q^l) \cdot \lambda^{-2} h + \cdots \tag{6.43}$$

and with (5.38) one determines the first two terms in the expansion of (6.16),

$$\Delta(-q^l r^l) = \partial_x^2 \ln q^l = \partial_x(\partial_x \ln q^l), \tag{6.44}$$

$$\begin{aligned} &\Delta \tfrac{1}{2}(q^l \partial_x r^l - r^l \partial_x q^l) \\ &\quad = \partial_x(\tfrac{1}{2}(\partial_x \ln q^l)^2 - q^l r^l) + \tfrac{1}{2}\partial_x(\partial_x^2 \ln q^l). \end{aligned} \tag{6.45}$$

Summation over $N \leqslant l \leqslant M-1$ yields

$$\partial_x \left(\sum_{l=N}^{M-1} \partial_x \ln q^l \right) = 0, \tag{6.46}$$

$$\partial_x \left(\sum_{l=N}^{M-1} \frac{1}{2}(\partial_x \ln q^l)^2 - \sum_{l=N}^{M-1} q^l r^l \right) = 0. \tag{6.47}$$

If we substitute $u^l := \ln q^l$, $p^l := \partial_x u^l$, (6.46) and (6.47) are just the conservation of the total linear momentum and the total energy of the Toda lattice;

$$\partial_x \left(\sum_{l=N}^{M-1} p^l \right) = 0, \tag{6.48}$$

$$\partial_x \left(\sum_{l=N}^{M-1} \frac{1}{2}(p^l)^2 + \sum_{l=N}^{M-1} e^{u^l - u^{l-1}} \right) = 0. \tag{6.49}$$

VII. THE TODA–AKNS SYSTEM AS DEFINING EQUATIONS FOR $P(0_{\Lambda_0})$

A. Introduction

In the preceding sections we have seen how an element $\hat{g} \in \hat{G}^{\Lambda_0}$ gives rise to a solution $q^l = -\tau_{l+1}^{(0)}/\tau_l^{(0)}$, $r^l = \tau_{l-1}^{(0)}/\tau_l^{(0)}$ of the Toda–AKNS system. Since we are working with the algebraic group $\hat{G}^{\Lambda_0}$, the solution satisfies the following:

there exist integers N, M such that

$$\begin{aligned} &q^l, r^l \neq 0, \quad \text{for } N \leqslant l \leqslant M, \\ &q^N \neq 0, \quad r^N \equiv 0, \quad q^M \equiv 0, \quad r^M \neq 0. \end{aligned} \tag{7.1a}$$

for $N \leqslant l \leqslant M$ q^l and r^l are rational functions of a finite number of variables, say $t_1, t_2, \ldots, t_n$. (7.1b)

In this section we want to follow the reverse way; starting from a solution q^l, r^l of the Toda–AKNS equations, which is of the form (7.1a) and (7.1b), we will construct a group

element $\hat{g}\in\hat{G}^{\Lambda_0}$, which gives rise to this solution.

Such a group element is obviously not unique, since multiplication on the right with an arbitrary element of the stabilizer P_0 of the highest weight vector v_0 will not change the τ function. Moreover, since q^l and r^l are quotients of components of τ functions, $\hat{g}$ and $\lambda\hat{g}$, $\lambda\in\mathbb{C}^*$ will yield the same solution.

The explicit construction of $\hat{g}$ in this section shows that the freedom mentioned above is the only one. This means that any solution of the Toda–AKNS equations of type (7.1a) and (7.1b) corresponds to precisely one element of the projectivized group orbit $P(\hat{G}^{\Lambda_0}/P_0)=P(0_{\Lambda_0})$ and vice versa. We may also say that the Toda–AKNS equations are defining equations for $P(0_{\Lambda_0})$ as a subset of $P(L(\Lambda_0))$.

B. Outline of the construction

The strategy to construct a group element $\hat{g}$ starting from a solution q^l, r^l of the form (7.1a) and (7.1b) is as follows: suppose for a moment that we have found an element $\hat{g}$ giving rise to this solution. The factors $\hat{\psi}^l_-$, $\hat{\psi}^l_{0,+}$ from the Birkhoff decomposition (3.8) of

$$\hat{\psi}^l = T^{-l}\exp\Big(\sum_{i=1}^n p_i t_i\Big)\hat{g}$$

then satisfy the linear equations (4.7a) and (4.7b). Projecting these equations on the loop group and rewriting them slightly, we find

$$\partial_{t_i}\tilde{\psi}^l_{0,+} = \tilde{R}^l{}_+(p_i)\tilde{\psi}^l_{0,+}, \quad 1\leqslant i\leqslant n, \tag{7.2a}$$

$$\partial_{t_i}\tilde{\psi}^l_- = p_i\tilde{\psi}^l_- - \tilde{\psi}^l_-\tilde{R}^l{}_+(p_i), \quad 1\leqslant i\leqslant n. \tag{7.2b}$$

Remember that $\tilde{R}^l{}_+(p_i)$ is an x-differential polynomial, so that it is completely determined by the solution q^l, r^l. This suggests that, given such a solution, we should try to solve the linear equations. Having done this, we define

$$\tilde{\psi}^l := \tilde{\psi}^l_-\tilde{\psi}^l_{0,+} \tag{7.3}$$

and we derive from (7.2a) and (7.2b),

$$\partial_{t_i}\tilde{\psi}^l = p_i\tilde{\psi}^l, \quad 1\leqslant i\leqslant n. \tag{7.4}$$

The solution of (7.4) is simply

$$\tilde{\psi}^l = \exp\Big(\sum_{i=1}^n p_i t_i\Big)\tilde{\psi}^l(0) = \exp\Big(\sum_{i=1}^n p_i t_i\Big)\tilde{\psi}^l_-(0)\tilde{\psi}^l_{0,+}(0). \tag{7.5}$$

It is now evident that we should take for $\hat{g}$ an arbitrary lift of

$$\tilde{g} := T^l\tilde{\psi}^l_-(0)\tilde{\psi}^l_{0,+}(0) \tag{7.6}$$

to $\hat{G}^{\Lambda_0}$. This determines $\hat{g}$ up to a nonzero complex constant $\lambda\in\mathbb{C}^*$. The remaining freedom in $\hat{g}$ consists of the possible choices of the initial conditions $\tilde{\psi}^l_{0,+}(0)$ and $\tilde{\psi}^l_-(0)$ for the linear equations (7.2a) and (7.2b).

In Sec. VII C we will see that (7.2a) has a local solution $\tilde{\psi}^l_{0,+}(t)\in\mathrm{Sl}_2(A_p^+)$ for any initial condition $\tilde{\psi}^l_{0,+}(0)\in\mathrm{Sl}_2(\mathbb{C}[\lambda])\simeq P_0$, so that $\hat{g}$ can indeed be multiplied on the right by an arbitrary element of P_0. The linear equations (7.2b) will be studied in Sec. VII D. It will turn out that (7.2b) is only solvable in $\mathrm{Sl}_2(A_p^-)$ for a special class of initial conditions $\tilde{\psi}^l_-(0)$ and that this class contains only one element of the algebraic loop group $\tilde{G}$.

C. The linear system (7.2a)

We start to remark that the integrability conditions of (7.2a) are, of course, satisfied, because q^l, r^l is a solution of the AKNS equations;

$$\partial_{t_j}\partial_{t_i}\tilde{\psi}^l_{0,+} - \partial_{t_i}\partial_{t_j}\tilde{\psi}^l_{0,+} = [\tilde{D}^l_{t_i},\tilde{D}^l_{t_j}]\tilde{\psi}^l_{0,+} = 0, \quad 1\leqslant i,j\leqslant n, \tag{7.7}$$

therefore any initial condition

$$\tilde{\psi}^l_{0,+}(0) = B_0 + B_1\lambda + \cdots + B_k\lambda^k\in\mathrm{gl}_2(\mathbb{C}[\lambda]) \tag{7.8}$$

determines a local solution $\tilde{\psi}^l_{0,+}(t)$, $t=(t_1,\ldots,t_n)$, $t_i<\epsilon_i$. [Since q^l, r^l have singularities, $\tilde{\psi}^l_{0,+}(t)$ will in general not be globally defined.] Because $\tilde{R}^l{}_+(p_i)$ is traceless, $\det\tilde{\psi}^l_{0,+}(t)$ is constant in t. Hence choosing $\tilde{\psi}^l_{0,+}(0)\in\mathrm{Sl}_2(\mathbb{C}[\lambda])$, we will have $\det\tilde{\psi}^l_{0,+}(t)=1$.

We now investigate the λ dependence of the solution $\tilde{\psi}_{0,+}(t)$. Because of the integrability of (7.2a) we may perform the integration of this system in n successive steps, where in the ith step all times except t_i remain constant. The resulting ordinary differential equations (ODE's) can be solved iteratively and it is not too difficult to derive the following lemma.

Lemma 7.1: Let p be an arbitrary symmetric weight, $\tilde{\psi}^l_{0,+}(t)$ a solution of (7.2a) with initial condition $\tilde{\psi}^l_{0,+}(0)\in\mathrm{Sl}_2(\mathbb{C}[\lambda])$, then $\tilde{\psi}_{0,+}(t)\in\mathrm{Sl}_2(A_p^+)$.

D. The linear system (7.2b)

The linear system (7.2b) is fundamentally different from (7.2a). One may of course, just as in Sec. VII C, check that the integrability conditions are satisfied and conclude that any initial condition determines a local solution of (7.2b). However, because of the presence of positive powers of λ in the right-hand side of (7.2b), this solution will, even if its initial value is prescribed to be in $\mathrm{Sl}_2(\mathbb{C}[\lambda^{-1}])$, in general not remain in this group [or in one of its completions $\mathrm{Sl}_2(A_p^-)$]. This can only be achieved for a special class of initial conditions.

In order to find these initial conditions, let us try to construct a formal solution $\tilde{\psi}^l_-$ of the form

$$\tilde{\psi}^l_- = \sum_{i\geqslant 0} A_i\lambda^{-i}, \quad A_0 = I. \tag{7.9}$$

Using Lemma (5.1) we can write

$$\tilde{\psi}^l_- = e^{s^l}\cdot e^{k^l}, \tag{7.10}$$

where

$$s^l = \sum_{i\geqslant 1}\lambda^{-i}s^l_i, \quad s^l_i\in\underline{h}$$

and

$$k^l = \sum_{i\geqslant 1}\lambda^{-i}k^l_i, \quad k^l_i\in\underline{h}^\perp$$

are formal power series. Substitution of (7.10) in (7.2b) yields, of course (see 5.31),

$$e^{k^l}(\partial_{t_i} - \tilde{R}^l{}_+(p_i))e^{-k^l} = \partial_{t_i}s^l - p_i. \tag{7.11}$$

According to Proposition 5.3 such a k^l is uniquely deter-

mined in terms of the solution q^l, r^l. Furthermore, s^l should satisfy

$$\partial_{t_i} s^l = p_i + e^{k^l} \tilde{D}^l_{t_i} e^{-k^l}. \tag{7.12}$$

These equations are again integrable because

$$\partial_{t_j} \partial_{t_i} s^l - \partial_{t_i} \partial_{t_j} s^l = e^{k^l} [\tilde{D}^l_{t_i}, \tilde{D}^l_{t_j}] e^{-k^l} = 0. \tag{7.13}$$

Note that the integrability conditions (7.13) are just the conservation laws (6.4). So any initial condition $s^l(0) \in \underline{\tilde{s}}_-$ determines a local solution $s^l(t)$, $t_i < \epsilon_i$ of (7.12).

Thus we have determined all solutions of (7.2b), which contain only negative powers of λ. The possible initial conditions are

$$e^{s^l}(0) \cdot e^{k^l}(0), \tag{7.14}$$

where $s^l(0) \underline{s}_-$ is arbitrary and $k^l(0)$ is determined by q^l and r^l. The problem with (7.10) and (7.14) is of course that they are just formal power series and that we do not know at all if they belong to the algebraic group $\mathrm{Sl}_2(\mathbb{C}[\lambda^{-1}])$ or one of its completions.

To solve this problem, we use condition (7.1a). Let M be the right boundary point determined by this condition. Using the proof of Lemma 6.1(b) and 6.1(c) one shows that $\exp k^M(0) \in \mathrm{Sl}_2(\mathbb{C}[\lambda^{-1}])$. Therefore the set of possible initial conditions (7.14) contains for $l = M$ only one algebraic initial condition, namely $s^M(0) = 0$ and $\tilde{\psi}^M_-(0) = \exp k^M(0)$. Using again the proof of Lemma 6.1(b) and 6.1(c) one shows that $\partial_{t_i} s^M = 0$, $\forall i$. Hence the unique algebraic solution of the linear system (7.2b) for $l = M$ is given by

$$\tilde{\psi}^M_-(t) = e^{k^M(t)}. \tag{7.15}$$

E. Conclusion

We now combine the results of the previous sections to define, as was outlined in Sec. VII B, an element of the algebraic loop group by

$$\tilde{g} := T^M \tilde{\psi}^M_-(0) \tilde{\psi}^M_{0,+}(0) \tag{7.16}$$

and take for $\hat{g}$ an arbitrary lift of $\hat{g}$ to $\hat{G}^{\Lambda_0}$. It remains to check that the solution $\bar{q}^l, \bar{r}^l$ of the Toda–AKNS system associated to this element $\hat{g}$ coincides with our original solution q^l, r^l for all $N \leqslant l \leqslant M$. For this we remark that

$$\lambda \frac{h}{2} + \begin{pmatrix} 0 & \bar{q}^M \\ \bar{r}^M & 0 \end{pmatrix} = \left[(\tilde{\psi}^M_-)^{-1} \lambda \frac{h}{2} \tilde{\psi}^M_- \right]_+ = [e^{-k^M} \lambda (h/2) e^{k^M}] = \lambda \frac{h}{2} + \begin{pmatrix} 0 & q^M \\ r^M & 0 \end{pmatrix} \tag{7.17}$$

by the results of Secs. IV and V. So we have

$$\bar{q}^M = q^M = 0, \quad \bar{r}^M = r^M = \tau^{(0)}_{M-1} / \tau^{(0)}_M \neq 0. \tag{7.18}$$

Moreover, both $\bar{q}^l, \bar{r}^l$ and q^l, r^l satisfy the differential difference equations (5.38), which may be rewritten as

$$\begin{aligned} q^{l-1} &= -1/r^l, \quad N+1 \leqslant l \leqslant M, \\ r^{l-1} &= -q^l (r^l)^2 + r^l \partial_x^2 \ln r^l. \end{aligned} \tag{7.19}$$

From this we conclude that $\bar{q}^l, \bar{r}^l$ and q^l, r^l are indeed identical.

We may summarize this section with the following proposition.

Proposition 7.2: Let q^l, r^l be a solution of the Toda–AKNS system that satisfies (7.1a) and (7.1b). Then there exists an element $\tau = \hat{g} \cdot v_0$ of the group orbit of $\hat{G}^{\Lambda_0}$ through v_0, which has, in the homogeneous realization of $L(\Lambda_0)$, components $\tau^{(0)}_l$, such that

$$q^l = -\tau^{(0)}_{l+1} / \tau^{(0)}_l, \quad r^l = \tau^{(0)}_{l-1} / \tau^{(0)}_l, \quad N \leqslant l \leqslant M. \tag{7.20}$$

The τ function, and hence its components $\tau^{(0)}_l$, is unique up to a multiplicative constant.

Note added in proof: After completing the manuscript we received a preprint from Imbens,[21] which also discusses the connection between representation theory and zero curvature constructions. In particular, formula (1.3) is derived for $l = 0$. We also mention Ref. 22, a related paper.

[1] E. Date, M. Kashiwara, M. Jimbo, and T. Miwa, Publ. R.I.M.S. **18**, 1077 (1982).
[2] V. G. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U.P., Cambridge, 1985), 2nd ed.
[3] V. G. Drinfeld and V. V. Sokolov, J. Sov. Math. **30**, 1975 (1985).
[4] G. Wilson, C. R. Acad. Sci. Paris **299**, 587 (1984).
[5] G. Wilson, Philos. Trans. R. Soc. London Ser. A **315**, 393 (1985).
[6] A. P. E. ten Kroode and M. J. Bergvelt, Lett. Math. Phys. **12**, 139 (1986).
[7] M. J. Bergvelt and A. P. E. ten Kroode, J. Math. Phys. **28**, 302 (1987).
[8] H. Flaschka, A. C. Newell, and T. Ratiu, Physica D **9**, 300 (1983).
[9] R. Goodman and N. R. Wallach, J. R. Angew. Math. **347**, 69 (1984).
[10] I. B. Frenkel, V. G. Kac, Invent. Math. **62**, 23 (1981).
[11] V. G. Kac, D. H. Peterson, Proc. Natl. Acad. Sci. USA **80**, 1778 (1983).
[12] G. B. Segal, Commun. Math. Phys. **80**, 301 (1981).
[13] A. Pressley and G. Segal, *Loopgroups* (Clarendon, Oxford, 1986).
[14] G. F. Helminck, "Deformations of connections, the Riemann–Hilbert problem and τ-functions," Report PM-R8602, Centre for Mathematics and Computer-Science, Amsterdam, 1986.
[15] L. A. Dickey, Commun. Math. Phys. **82**, 345 (1981).
[16] H. Flaschka, in *Non-linear Integrable Systems-Classical Theory and Quantum Theory*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).
[17] A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).
[18] V. S. Varadarajan, *Lie Groups, Lie Algebras, and their Representations* (Springer, New York, 1984).
[19] G. Wilson, Quart. J. Math. Oxford **32**, 491 (1981).
[20] I. B. Frenkel, Invent. Math. **77**, 301 (1984).
[21] H. Imbens, "Generalized KdV-equations and τ-functions," Preprint No. 468, Department of Mathematics, University of Utrecht, 1986.
[22] H. Imbens, C. R. Acad. Sci. Paris **305**, 229 (1987).

Integral equations with symmetrical kernel applied to a system with a Dirac-type spectrum

Zhong-Qi Ma
Institute of High Energy Physics, P.O. Box 918, Beijing, People's Republic of China and Jiaotong University of Shanghai, Shanghai, People's Republic of China

An-Ying Dai
Beijing Institute of Technology, Beijing, People's Republic of China

(Received 30 September 1987; accepted for publication 10 February 1988)

The generalized Sturm–Liouville problem for a system with a Dirac-type spectrum was converted to that of a linear integral operator with a symmetrical kernel by Yang [Commun. Math. Phys. **112**, 205 (1987)]. A supplementary case is discussed in this paper.

I. INTRODUCTION

Recently, the Sturm–Liouville theorem was generalized[1] to $2N$ coupled first-order linear ordinary differential equations with a Dirac-type spectrum. The elegant Sturm–Liouville theorem is a powerful tool for obtaining information on the number of eigenvalues, on the nodes of the wave functions, and on the meaning of Levinson's theorem.[2–6] However, the relativistic equation of motion, the Dirac equation, is a first-order differential one, and there are different characters between the first-order and the second-order differential equations. For instance, the energy spectrum for the Dirac equation is unbounded from below.

Consider the radial Dirac equation

$$\begin{aligned} -g'(r)+(\kappa/r)g(r)&=[E-V(r)-M]f(r),\\ f'(r)+(\kappa/r)f(r)&=[E-V(r)+M]g(r), \end{aligned} \tag{1.1}$$

where the potential $V(r)$ is a real continuous function of r satisfying

$$V(r)=\begin{cases}0, & \text{at } r\geqslant a,\\ \text{const}, & \text{at } r\sim 0,\end{cases} \tag{1.2}$$

and $\kappa=\mp(j+\frac{1}{2})$. For definiteness, we discuss the case with $\kappa>0$. Changing Eq. (1.1) into the matrix form, we have

$$H(r)\psi(r)=E\psi(r),\quad \psi(r)=\begin{pmatrix}f(r)\\ g(r)\end{pmatrix}, \tag{1.3}$$

$$\begin{aligned} H(r)&=-i\sigma_2\frac{\partial}{\partial r}+U(r),\\ U(r)&=(\kappa/r)\sigma_1+M\sigma_3+V(r)\mathbb{1}_2, \end{aligned} \tag{1.4}$$

where $U(r)$ is a 2×2 symmetric matrix and σ_j is the Pauli matrix. Near the origin, one solution is divergent and the other is vanishing. From the indicial equation of (1.1), two solutions satisfy the different boundary conditions at the origin. The latter solution satisfies

$$f/g|_{r=0}=K_0=0. \tag{1.5}$$

In the region $[a,\infty)$ the solutions of Eq. (1.1) can be obtained exactly. For $|E|<M$, there is only one solution vanishing in infinity, the ratio f/g of which at $r=a+$ can be calculated as

$$f/g|_{a+}=K_a. \tag{1.6}$$

If in the region $[0,a]$ there is a solution satisfying Eq. (1.3) and both boundary conditions (1.5) and (1.6), this solution is the physically admissible one, called the bound state.

In some physical problems, for example, fermions moving in the background monopole field[2,4] or in the background Skyrmion field,[6] and the fully relativistical treatment for two particles,[7] the radial equations appear to be four coupled first-order linear ordinary differential equations which are the general forms of Eq. (1.3) with the general forms of boundary conditions (1.5) and (1.6). Yang's generalized form of the Sturm–Liouville theorem provides a powerful method to deal with those physical problems.

Furthermore, by converting the Sturm–Liouville problem to that of an integral operator with a symmetrical kernel,[8,9] one has powerful control over the properties of the eigenfunctions and eigenvalues. The generalization has been made by Yang to the first-order differential equation

$$[\epsilon-H(x)]\Phi(x)=f(x),\quad 0\leqslant x\leqslant a,$$

where ϵ is not equal to the eigenvalue of the Hamiltonian. In this paper we will generalize this method to the case where ϵ is an eigenvalue of the Hamiltonian. Some properties of the solutions will also be discussed in this paper.

We use the same notations as used in Ref. 1.

II. PROBLEM

We consider the first-order ordinary differential equations in the matrix form

$$[\epsilon-H(x)]\Phi(x)=f(x),\quad 0\leqslant x\leqslant a, \tag{2.1}$$

where $\Phi(x)$ and $f(x)$ are $2N\times 1$ column matrices, $f(x)$ is piecewise continuous in the region $[0,a]$, and

$$H(x)=\omega\,\partial_x+V(x), \tag{2.2}$$

$$\omega=\begin{pmatrix}\mathbf{0} & -\mathbb{1}_N\\ \mathbb{1}_N & \mathbf{0}\end{pmatrix}, \tag{2.3}$$

where $V(x)$ is a real symmetrical $2N\times 2N$ matrix and continuous in the region $[0,a]$, $\mathbb{1}_N$ is an $N\times N$ unit matrix, and all quantities are real. When $f(x)=0$, we have the eigenequation of $H(x)$,

$$H(x)\psi(x)=E\psi(x),\quad \psi=\begin{pmatrix}\xi\\ \eta\end{pmatrix}, \tag{2.4}$$

where ξ and η are $N\times 1$ column matrices. We shall discuss

 0022-2488/88/061321-04$02.50

the solutions ψ satisfying the boundary conditions at $x=0$ and $x=a$:

$$\xi = K_0\eta, \quad \text{at } x=0, \tag{2.5}$$

$$\xi = K_a\eta, \quad \text{at } x=a, \tag{2.6}$$

where K_0 and K_a are real symmetrical $N\times N$ matrices.

There are N linearly independent solutions of Eq. (2.4) satisfying the boundary condition (2.5). Arranging them in columns we obtain the matrix solution ψ:

$$H(x)\psi(x,E) = E\psi(x,E), \tag{2.7}$$

$$\psi = \begin{pmatrix}\xi\\ \eta\end{pmatrix}, \tag{2.8}$$

where ξ and η are $N\times N$ matrices, and, at $x=0$,

$$\eta(0,E) = 1_N, \quad \xi(0,E) = K_0. \tag{2.9}$$

Similarly, arranging N linearly independent solutions of Eq. (2.4) satisfying boundary condition (2.6), we obtain the matrix solution ψ_1:

$$\psi_1 = \begin{pmatrix}\xi_1\\ \eta_1\end{pmatrix}, \tag{2.10}$$

and, at $x=a$,

$$\eta_1(a,E) = 1_N, \quad \xi_1(a,E) = K_a. \tag{2.11}$$

Define

$$W(x,E) = \tilde{\psi}_1\omega\psi = (\xi - K_a\eta)|_{x=a}, \tag{2.12}$$

then

$$\frac{d}{dx}W(x,E) = 0. \tag{2.13}$$

As shown in Ref. 1, the necessary and sufficient condition that there is a nonvanishing solution of Eq. (2.4) satisfying both boundary conditions (2.5) and (2.6) is

$$\det W(x,E) = 0. \tag{2.14}$$

Such a solution is called an eigenfunction and corresponding E is an eigenvalue:

$$H(x)\phi_l(x) = E_l\phi_l(x), \quad \phi_l(x) = \begin{pmatrix}\mu_l(x)\\ \nu_l(x)\end{pmatrix},$$
$$\mu_l(0) = K_0\nu_l(0), \quad \text{at } x=0, \tag{2.15}$$
$$\mu_l(a) = K_a\nu_l(a), \quad \text{at } x=a.$$

For any real number ϵ that is not an eigenvalue of Eqs. (2.4)–(2.6), the only solution of Eq. (2.1) satisfying both boundary conditions (2.5) and (2.6) for ψ at $x=0$ and $x=a$ can be expressed as[1]

$$\Phi(x) = \int_0^a \langle x|\Omega|y\rangle f(y)dy$$
$$= \psi_1(x,\epsilon)\tilde{W}^{-1}(x,\epsilon)\int_0^x \tilde{\psi}(y,\epsilon)f(y)dy$$
$$+ \psi(x,\epsilon)W^{-1}(x,\epsilon)\int_x^a \tilde{\psi}_1(y,\epsilon)f(y)dy. \tag{2.16}$$

If ϵ is an eigenvalue of $H(x)$, $\epsilon = E_s$, $W(x,E_s)$ is singular, and $W^{-1}(x,E_s)$ cannot be defined. In the following, we assume, first, that E_s is nondegenerate, and the only eigenfunction is

$$\phi_s(x) = \psi(x,E_s)\zeta = \begin{pmatrix}\mu_s(x)\\ \nu_s(x)\end{pmatrix}, \tag{2.17}$$

where ζ is a real normalized $N\times 1$ matrix and

$$\nu_s(0) = \zeta, \quad \mu_s(0) = K_0\zeta,$$
$$\mu_s(a) = \xi(a,E_s)\zeta = K_a\eta(a,E_s)\zeta = K_a\nu_s(a). \tag{2.18}$$

III. NECESSARY CONDITION

If the piecewise differentiable solution $\Phi(x)$ of Eq. (2.1) satisfying both boundary conditions (2.5) and (2.6) exists when ϵ is an eigenvaue E_s of $H(x)$, we multiply the eigenfunction $\phi_s(x)$ on the both sides of Eq. (2.1) and integrate it in $[0,a]$:

$$\int_0^a \tilde{\phi}_s(x)f(x)dx = \int_0^a \tilde{\phi}_s(x)[E_s - H(x)]\Phi(x)dx$$
$$= \int_0^a \{[E_s - H(x)]\phi_s(x)\}^T\Phi(x)dx$$
$$- \tilde{\phi}_s(x)\omega\Phi(x)\Big|_{x=0}^{x=a} = 0,$$

where the superscript T denotes the transpose of the matrix. Therefore, the necessary condition for the existence of such a solution $\Phi(x)$ is

$$\int_0^a [\psi(x,E_s)\zeta]^T f(x)dx = 0. \tag{3.1}$$

IV. GENERAL SOLUTIONS

If $\Phi_1(x)$ and $\Phi_2(x)$ are the solutions of Eq. (2.1) satisfying the boundary conditions (2.5) and (2.6), the difference $\Phi_1(x) - \Phi_2(x)$ satisfies Eq. (2.4) with the eigenvalues $E = E_s$. Since the eigenvalue E_s is nondegenerate,

$$\Phi_1(x) - \Phi_2(x) = c\psi(x,E_s)\zeta = c\phi_s(x). \tag{4.1}$$

The general solutions of Eqs. (2.1), (2.5), and (2.6) can be expressed as

$$\Phi(x) + c\phi_x(x), \tag{4.2}$$

where $\Phi(x)$ is a particular solution which we will discuss.

V. PARTICULAR SOLUTION

Define X as an $N\times N$ real orthogonal matrix with ζ as its first column and Y as such a matrix whose first column is proportional to $\eta(a,E_s)\zeta = \nu_s(a)$, which is not vanishing because the eigenfunction Φ_s is not a trivial one. Now, the necessary condition (3.1) becomes

$$\int_0^a [\psi(x,E_s)X]^T f(x)dx = \begin{pmatrix}0\\ D\end{pmatrix}, \tag{5.1}$$

where D is an $(N-1)\times 1$ matrix.

It is easy to see that the first column and row of the following matrix vanish:

$$\tilde{Y}W(x,E_s)X = \tilde{Y}(\xi\eta^{-1} - K_a)\eta X|_{x=a} = \begin{pmatrix}0 & 0\\ 0 & W_1\end{pmatrix}, \tag{5.2}$$

where $W_1(x,E_s)$ is an $(N-1)\times(N-1)$ real nonsingular constant matrix because the eigenvalue E_s is nondegenerate. The first column of the matrix in Eq. (5.2) vanishes because

the first column of ηX is the lower part $\nu_s(a)$ of the eigenfunction ϕ_s; and the first column of the transposed matrix of (5.2) vanishes because the first columns of ηX and Y are proportional to each other.

Define an $N\times N$ real symmetrical matrix K':

$$K' = K_a - Y\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\tilde{Y}. \tag{5.3}$$

We get the solution $\psi'(x,E_s)$ of Eq. (2.7) with $E=E_s$, satisfying the boundary condition at $x=a$:

$$\psi' = \begin{pmatrix}\xi'\\ \eta'\end{pmatrix},$$
$$\eta'(a,E_s) = 1\!\!1_N, \quad \xi'(a,E_s) = K', \quad \text{at } x=a. \tag{5.4}$$

Instead of $W(x,E_s)$ we introduce $W'(x,E_s)$:

$$W'(x,E_s) = \tilde{\psi}'(x,E_s)\omega\psi(x,E_s)$$
$$= (\xi\eta^{-1} - K')\eta|_{x=a}. \tag{5.5}$$

In fact, $W'(x,E_s)$ is independent of x. Now,

$$\tilde{Y}W'X = \tilde{Y}\left[\xi\eta^{-1} - K_a + Y\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\tilde{Y}\right]\eta X\Big|_{x=a}$$
$$= \begin{pmatrix}0 & 0\\ 0 & W_1\end{pmatrix} + \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\tilde{Y}\eta X\Big|_{x=a}$$
$$= \begin{pmatrix}b & bS\\ 0 & W_1\end{pmatrix}, \tag{5.6}$$

where S is a $1\times(N-1)$ matrix, and b is a nonvanishing constant. Therefore, W' is nonsingular,

$$\det W' \neq 0, \tag{5.7}$$

and W'^{-1} can be defined.

Defining $\Phi(x)$ as in Eq. (2.16):

$$\Phi(x) = \int_0^a \langle x|\Omega'|y\rangle f(y)dy$$
$$= \psi'(x,E_s)\tilde{W}'^{-1}(x,E_s)\int_0^x \tilde{\psi}(y,E_s)f(y)dy$$
$$+ \psi(x,E_s)W'^{-1}(x,E_s)\int_x^a \tilde{\psi}'(y,E_s)f(y)dy, \tag{5.8}$$

and following the proof of Lemma 9 in Ref. 1, we find that $\Phi(x)$ is a piecewise differentiable solution of Eq. (2.1) with $\epsilon = E_s$, and satisfies the boundary conditions

$$\Phi(x) = \begin{pmatrix}\mu(x)\\ \nu(x)\end{pmatrix},$$
$$\mu(0) = K_0\nu(0), \quad \mu(a) = K'\nu(a). \tag{5.9}$$

However, at $x=a$ we have from Eq. (5.8)

$$\Phi(a) = \psi'(a,E_s)Y\tilde{Y}\tilde{W}'^{-1}(a,E_s)X$$
$$\times\int_0^a [\psi(y,E_s)X]^T f(y)dy$$
$$= \psi'(a,E_s)Y\begin{pmatrix}b^{-1} & 0\\ -\tilde{W}_1^{-1}\tilde{S} & \tilde{W}_1^{-1}\end{pmatrix}\begin{pmatrix}0\\ D\end{pmatrix}$$
$$= \psi'(a,E_s)Y\begin{pmatrix}0\\ \tilde{W}_1^{-1}D\end{pmatrix},$$

$$\mu(a) = K'\nu(a) = K_a\nu(a) - Y\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}0\\ \tilde{W}_1^{-1}D\end{pmatrix}$$
$$= K_a\nu(a). \tag{5.10}$$

Therefore, $\Phi(x)$ in Eq. (5.8) is the particular solution of Eqs. (2.1), (2.5), and (2.6).

VI. DEGENERATE CASE

If E_s is n-fold degenerate, there are n linearly independent eigenfunctions $\phi_{si} = \psi\zeta_i$, $i=1,2,...,n$, where ζ_i can be chosen as orthogonal and normalized. The necessary condition for the existence of the solution of Eqs. (2.1), (2.5), and (2.6) is that $f(x)$ satisfies Eq. (3.1) for each of $\psi\zeta_i$. The general solutions now can be expressed as

$$\Phi(x) + \sum_{i=1}^{n} c_i\psi(x,E_s)\zeta_i, \tag{6.1}$$

where $\Phi(x)$ is the particular solution. In the degenerate case, X is defined as an $N\times N$ real orthogonal matrix whose first n columns are ζ_i. The $\eta(a,E_s)\zeta_i$ are linearly independent because there are n linearly independent eigenfunctions with the degenerate eigenvalues E_s. Orthogonalizing and normalizing them, we define an $N\times N$ real orthogonal matrix y whose first n columns are

$$Y_i = \sum_{j=1}^{n}[\eta(a,E_s)\zeta_j](B^{-1})_{ji},$$
$$\tilde{\zeta}_j\tilde{\eta}(a,E_s)\eta(a,E_s)\zeta_i = B_{ji}^2, \tag{6.2}$$
$$i,j = 1,2,...,n, \quad \det B \neq 0.$$

Now, $W_1(x,E_s)$ in Eq. (5.2) becomes an $(N-n)\times(N-n)$ real constant matrix with a nonvanishing determinant. Defining an $N\times N$ real symmetrical matrix K' as

$$K' = K_a - Y\begin{pmatrix}1_n & 0\\ 0 & 0\end{pmatrix}\tilde{Y}, \tag{6.3}$$

and substituting it into Eqs. (5.4) and (5.5), we get

$$\tilde{Y}W'X = \begin{pmatrix}0 & 0\\ 0 & W_1\end{pmatrix} + \begin{pmatrix}1_n & 0\\ 0 & 0\end{pmatrix}\tilde{Y}\eta X\Big|_{x=a} = \begin{pmatrix}B & BS\\ 0 & W_1\end{pmatrix}, \tag{6.4}$$

where S is an $n\times(N-n)$ matrix. Thus

$$\det W' \neq 0, \tag{6.5}$$

and W'^{-1} is defined. Finally, substituting $\psi'(x,E_s)$ and $W'(x,E_s)$ into Eq. (5.8), we can also prove Eq. (5.10), namely, $\Phi(x)$ is the particular solution of Eqs. (2.1), (2.5), and (2.6), where $\epsilon = E_s$ is an n-fold degenerate eigenvalue of $H(x)$.

VII. INTEGRAL TRANSFORMATION ONTO THE EIGENFUNCTION OF $H(x)$

Choose the source function $f(x)$ as

$$f(x) = (E_s - H)\phi_l = (E_s - E_l)\phi_l, \tag{7.1}$$

where ϕ_l is an eigenfunction of $H(x)$ with the eigenvalue $E_l \neq E_s$, and $f(x)$ satisfies the necessary condition (3.1). The solution $\Phi(x)$ in Eq. (5.8) now satisfies the same equation and boundary conditions as ϕ_l, so

$$\Phi(x) - \phi_l(x) = \sum_i c_i \phi_{si}(x). \tag{7.2}$$

From Eq. (5.10), the lower component $\nu(x)$ of $\Phi(x)$ satisfies

$$K_a \nu(a) = K' \nu(a), \tag{7.3}$$

i.e.,

$$[\eta(a,E_s)\zeta_i]^T \nu(a) = 0. \tag{7.4}$$

Since the lower component of $\phi_{si}(a)$ is just $\eta(a,E_s)\zeta_i$, the coefficients c_i in Eq. (7.2) can be determined uniquely from Eqs. (7.2) and (7.4). Therefore,

$$\int_0^a \langle x|\Omega'|y\rangle \phi_l(y)dy = \frac{\Phi(x)}{E_s - E_l} = \left[\phi_l(x) + \sum_i c_i \phi_{si}(x)\right](E_s - E_l)^{-1}, \tag{7.5}$$

where the c_i's have been determined. Equation (7.5) is the generalized form of Lemma 10 in Ref. 1.

VIII. APPLICATION

Courant and Hilbert[8] pointed out that every continuous function $\Phi(x)$ that, as in Eq. (2.16), is an integral transform with a symmetric kernel $\langle x|\Omega|y\rangle$ of a piecewise continuous function $f(y)$ can be expanded in a series in the eigenfunctions $g_l(y)$ of $\langle x|\Omega|y\rangle$:

$$\int_0^a \langle x|\Omega|y\rangle g_l(y)dy = \mu_l g_l(y), \tag{8.1}$$

$$\Phi(x) = \sum_l c_l g_l(x), \tag{8.2}$$

$$c_l = \int \tilde{g}_l(x)\Phi(x)dx = \int\int \tilde{g}_l(x)\langle x|\Omega|y\rangle f(y)dx\,dy = \mu_l \int \tilde{g}_l(y) f(y)dy. \tag{8.3}$$

Now, when $\epsilon \neq E_s$, from Lemma 11 in Ref. 1,

$$H(x)g_l(x) = E_l g_l(x), \quad E_l = \epsilon - \mu_l^{-1}. \tag{8.4}$$

Therefore, $\Phi(x)$ in Eq. (2.16), which is the solution of Eqs. (2.1), (2.5), and (2.6), can be expanded in a series in the eigenfunctions $\phi_l(x)$ of $H(x)$. When $\epsilon = E_s$ and $f(x) = \Sigma_{l\neq s} d_l \phi_l(x)$, the solution $\Phi(x)$ can be expressed as

$$\Phi(x) = \int_0^a \langle x|\Omega|y\rangle f(y)dy = \sum_{l\neq s} d_l \int_0^a \langle x|\Omega|y\rangle \phi_l(y)dy = \sum_{l\neq s} \frac{d_l}{E_s - E_l}\left[\phi_l(x) + \sum_k c_k \phi_{sk}(x)\right],$$

where Eq. (7.5) has been used and c_k can be determined uniquely by Eqs. (7.2) and (7.4).

ACKNOWLEDGMENTS

The authors would like to thank Professor Chen Ning Yang for sending his new preprint[1] before publication and for his encouragement.

One of the authors (Z-Q. M.) was supported by the Natural Science Foundation of China under Grant No. 1860171.

[1] C. N. Yang, Commun. Math. Phys. **112**, 205 (1987).
[2] C. N. Yang, in *Monopole in Quantum Field Theory*, edited by N. S. Craigie, P. Goddard, and W. Nahm (World Scientific, Singapore, 1982), p. 237; Y. Kazama and C. N. Yang, Phys. Rev. D **15**, 2300 (1977).
[3] Z. R. Iwinski, L. Rosenberg, and L. Spruch, Phys. Rev. Lett. **54**, 1602 (1985); Phys. Rev. A **31**, 1229 (1985); **33**, 946 (1986).
[4] Z. Q. Ma, J. Math. Phys. **26**, 1995 (1985); Phys. Rev. D **32**, 2203, 2213 (1985); **33**, 1745 (1986).
[5] Z. Q. Ma, H. T. Nieh, and R. K. Su, Phys. Rev. D **32**, 3268 (1985).
[6] L. N. Chang and Z. Q. Ma, "Levinson's theorem for fermions moving in the background Skyrmion field," submitted to Commun. Math. Phys.
[7] A. O. Barut and N. Ünal, Physica A **142**, 467 (1987).
[8] See, e.g., E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford U.P., Oxford, 1958), Part I.
[9] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. I.

A geometric setting for the quantum planar n-body system, and a U($n-1$) basis for the internal states

Toshihiro Iwai
Department of Applied Mathematics and Physics, Kyoto University, Kyoto 606, Japan

(Received 8 July 1987; accepted for publication 10 February 1988)

In a previous paper [J. Math. Phys. **28**, 964 (1987)], the author showed that the internal motion of a molecule, a many-body system in the Born–Oppenheimer approximation, can be well described in terms of the gauge theory or of the connection theory in differential geometry. However, the scope of that paper centers on the planar triatomic molecule in order to put forward the gauge theory in an explicit manner. This paper is a continuation of the previous one and gives the generalization to the planar multiatomic molecule. The internal space of the n-atomic molecule proves to be diffeomorphic to $\mathbb{R}^+ \times \mathbb{C}P^{n-2}$, the product of the positive real numbers and the complex projective space. The internal states of the molecule are described as cross sections in complex line bundles over the internal space. Introduction of the complex line bundles is a geometric consequence of the angular momentum conservation law, because cross sections in each complex line bundle are in one-to-one correspondence with eigenstates that have a fixed total angular momentum eigenvalue. The internal Hamiltonian operator is obtained, which acts on the cross sections in the complex line bundle. Further, boson calculus is performed to obtain a complete basis of internal states of the molecule, using the harmonic oscillator annihilation and creation operators. As a result, carrier spaces of unitary irreducible representations of the unitary group $U(n-1)$, which are characterized by two integers, are realized as finite-dimensional subspaces of the space of the square integrable cross sections in the complex line bundle.

I. INTRODUCTION

A "molecule" in this and previous papers of the author means a system of several particles or atomic nuclei in the Born–Oppenheimer approximation. In quantum chemistry, theoretical treatment of nonrigid molecules has been related more or less with the Eckart frame.[1] However, the Eckart Hamiltonian is interpreted as valid in the vicinity of the equilibrium nuclear position. If one wishes to study motions of nonrigid molecules far from the equilibrium position, one must become involved with the difficulty in separating rotation and vibration. It is Guichardet[2] who showed that the vibration motion cannot, in general, be separated from the rotation motion.

On the basis of Guichardet's work, the author demonstrated in previous papers[3,4] that the internal motion of the nonrigid molecule can be well described in terms of the gauge theory or the connection theory in differential geometry: The center-of-mass system is made into a principal fiber bundle with the rotation group as the structure group. The base manifold of this fiber bundle is called the internal space, which is thought of as the set of all the molecule forms independent of the position. With this principal fiber bundle are associated the complex vector bundles assigned by the total angular momentum eigenvalues, the cross sections of which are understood as internal states of the molecule. The association of the complex vector bundle is a geometric consequence of the conservation of the total angular momentum. Hence the internal Hamiltonian operator is derived from the standard one on the center-of-mass system by using the conservation law of the total angular momentum.

Introductory remarks on the present geometric setting for the n-body system are here worth making to show that this setting could provide a profound view of n-body systems. If the molecule is regarded as a system of nucleons, i.e., as a nucleus, the geometric setting in this paper becomes related to the collective models of nuclei. In the microscopic collective model, one discusses the problem of effecting a change of coordinates on the n-body center-of-mass system $(\mathbb{R}^d)^{n-1}$ from Cartesian to collective plus intrinsic coordinates. The essential idea made clear by Rowe and Rosensteel[5] is this; consider a Lie group G acting on $(\mathbb{R}^d)^{n-1}$ and decompose $(\mathbb{R}^d)^{n-1}$ into orbits of G. The collective coordinates are taken to be a chart for the generic orbits, while the intrinsic variables are a set of coordinates for the transversal to those orbits. This idea has the same origin as the present geometric setting with $G = \mathrm{SO}(d)$, and may be developed further to fit into the principal fiber bundle theory; the center-of-mass system is indeed made into a principal fiber bundle. The collective coordinates are represented as the Euler angles for $\mathrm{SO}(d)$, and the intrinsic coordinates should be a chart for the internal space, the base manifold of the principal fiber bundle. The transversality is then taken to be the distinction between rotational and vibrational vectors, and eventually leads up to the connection on the principal fiber bundle.

In the collective models, one is interested in $G = \mathrm{SO}(d)$, $\mathrm{SL}(d,\mathbb{R})$, or $\mathrm{GL}^+(d,\mathbb{R})$, and in the collective part of the kinetic energy operator. However, in the present setting for $G = \mathrm{SO}(d)$, the whole (i.e., collective plus intrinsic) kinetic energy operator is discussed. The intrinsic (or "vibrational" in this paper's nomenclature) part is of interest from the chemical and geometric points of view, because this part is characteristic of nonrigid molecules and associated with the

 0022-2488/88/061325-13$02.50

linear connection induced on the complex vector bundles.

Though the previous papers[3,4] presented the general method for treating the internal motion of the molecule, as stated above, applications were restricted to the triatomic molecule; in Refs. 3 and 4, triatomic molecules were dealt with in two and three dimensions, respectively. This paper is a continuation of Ref. 3, and gives an application of that method to the planar multiatomic molecule.

The organization of this article is as follows: Section II is concerned with the center-of-mass system for the planar n-atomic molecule. It is shown that the center-of-mass system is made into a principal fiber bundle $\dot{\mathbb{C}}^{n-1} \to \mathbb{R}^+ \times \mathbb{C}P^{n-2}$ with the structure group $U(1) \cong SO(2)$, where the overdot means that the origin is deleted, and $\mathbb{R}^+$ and $\mathbb{C}P^{n-2}$ denote the positive real numbers and the complex projective space,[6] respectively. The connection defined by the Eckart condition is constructed and its curvature is shown to be the Kähler form associated with the Fubini–Study metric on $\mathbb{C}P^{n-2}$ (Ref. 6). The induced metric on the internal space $\mathbb{R}^+ \times \mathbb{C}P^{n-2}$ is also obtained, which turns out to be a warped product of the standard metric on $\mathbb{R}^+$ and the Fubini–Study metric on $\mathbb{C}P^{n-2}$.

Section III contains the associated complex line bundles over the internal space $\mathbb{R}^+ \times \mathbb{C}P^{n-2}$. These complex line bundles L_m are assigned by the eigenvalues m of the total angular momentum operator, i.e., all the integers. Thus cross sections in L_m are understood as internal states of the molecule with the total angular momentum eigenvalue m. The linear connection in L_m, which are induced from the connection defined on the center-of-mass system, is discussed together with its curvature.

Since $\mathbb{R}^+$ is contractible, complex line bundles over $\mathbb{R}^+ \times \mathbb{C}P^{n-2}$ are in one-to-one correspondence with complex line bundles over $\mathbb{C}P^{n-2}$, so that the contents of this section run in parallel with the theory of complex line bundles over $\mathbb{C}P^{n-2}$ (see Ref. 7, for example). The curvature two-form of the induced linear connection in L_m defines, independently of $\mathbb{R}^+$, the first Chern class on $\mathbb{C}P^{n-2}$,[7] and hence the first Chern class numbers prove to be nothing but the eigenvalues m of the total angular momentum operator.

The curvature form is interpreted as a "magnetic" field or a gauge field caused by the rotation of the molecule. Written in local coordinates, the linear connection in L_m is thought of as a minimal coupling of the internal motion with the gauge field.

Section IV deals with the internal Hamiltonian operator that acts on cross sections in the complex line bundle L_m. This operator turns out to be quadratic in the covariant derivation operator and to contain a centrifugal potential. This implies that the internal Hamiltonian operator satisfies the principle of minimal coupling with the gauge field.

Section V is devoted to boson calculus using the harmonic oscillator annihilation and creation operators in order to obtain a complete basis of internal states with the assigned angular momentum eigenvalue. In conclusion, unitary group actions on the cross sections in L_m are considered. In the space of cross sections in L_m, a series of carrier spaces of unitary irreducible representations of $U(n-1) \times U(n-1)$ and of $U(n-1)$ are found, which are characterized by two integers. Thus the internal states of the n-atomic molecule are classified in terms of a $U(n-1)$ basis.

Though this paper concentrates on the planar multiatomic molecules, the same idea of the geometric setting can run for d-dimensional multiatomic molecules. In fact, for $d = 3$, triatomic molecules were discussed in the same geometric setting, using explicit coordinates such as Euler angles, principal moments of inertia, etc. For three-dimensional multiatomic molecules, complex vector bundles V_l, $l = 0,1,...$, are used, instead of the complex line bundles L_m for the planar molecules, to describe the internal states of the molecule. The standard fiber of V_l is $\mathbb{C}^{2l+1}$, the carrier space for the unitary irreducible representation of $SO(3)$. The internal Hamiltonian operator can be obtained in the same manner as developed in Ref. 4.

While the two-dimensional molecule is of mathematical interest, its simplicity admits a complete analysis, so that the geometric setting becomes tractable. Indeed, the internal space, diffeomorphic with $\mathbb{R}^+ \times \mathbb{C}P^{n-1}$, and the complex line bundles L_m over it are topologically easy to understand. In this respect, together with the introductory remarks already made, Sec. V will give new insight into the microscopic collective models, but in two dimensions.

II. THE PRINCIPAL FIBER BUNDLE

In this section we make the center-of-mass system into a principal fiber bundle with the structure group $U(1) \cong SO(2)$, and discuss the connection and curvature due to Guichardet.[2] To carry out this program in an explicit manner, it is of great use to introduce an orthogonal system in the center-of-mass system, which is closely related with the so-called Jacobi vectors.

A. Settings on the center-of-mass system

Let y_j, $j = 1,...,n$, be position vectors of n particles in $\mathbb{R}^2$ with masses m_j, $j = 1,...,n$, respectively. The configuration space Q_0 of the planar n-atomic molecule is then the linear space of all the n-tuples $(y_1,...,y_n)$;

$$Q_0 := \{ y = (y_1,...,y_n);\ y_j \in \mathbb{R}^2 \} . \tag{2.1}$$

This is clearly isomorphic with the vector space $\mathbb{R}^{2n}$. The Q_0 is endowed with the inner product by

$$K(x,y) = \sum_{j=1}^{n} m_j (x_j | y_j) , \tag{2.2}$$

where the parentheses denote the standard inner product in $\mathbb{R}^2$.

The center-of-mass system Q is defined as the linear subspace of Q_0 by

$$Q := \left\{ x \in Q_0;\ \sum_{j=1}^{n} m_j x_j = 0 \right\} . \tag{2.3}$$

The inner product induced on Q will be also denoted by K.

The rotation group $SO(2)$ acting on $\mathbb{R}^2$ acts on Q_0 in a natural manner; for $g \in SO(2)$ one has

$$y = (y_1,...,y_n) \to gy = (gy_1,...,gy_n) . \tag{2.4}$$

It is clear that the center-of-mass system Q also admits the $SO(2)$ action.

The following proposition, easy to prove, is basic for the discussion below.

Proposition 2.1: Let e_1 and e_2 be the standard basis in $\mathbb{R}^2$. Then the following n-tuples constitute an orthonormal system in Q_0 with respect to the inner product K:

$$c_1 = N_0(e_1,...,e_1)\,,$$
$$c_2 = N_0(e_2,...,e_2)\,, \tag{2.5}$$

$$f_{2j-1} = N_j\Big(-m_{j+1}e_1,...,-m_{j+1}e_1,\Big(\sum_{k=1}^{j} m_k\Big)e_1,0,...,0\Big)\,,$$

$$f_{2j} = N_j\Big(\underbrace{-m_{j+1}e_2,...,-m_{j+1}e_2}_{j\text{ terms}},\Big(\sum_{k=1}^{j} m_k\Big)e_2,0,...,0\Big)\,,$$
$$j = 1,2,...,n-1, \tag{2.6}$$

where N_j, $j = 0,1,...,n$, are normalization constants given by

$$N_0 = \Big(\sum_{k=1}^{n} m_k\Big)^{-1/2}\,, \tag{2.7}$$

$$N_j = \Big[m_{j+1}\Big(\sum_{k=1}^{j} m_k\Big)\Big(\sum_{k=1}^{j+1} m_k\Big)\Big]^{-1/2}\,,\quad j = 1,...,n-1\,, \tag{2.8}$$

respectively.

The vectors f_k, $k = 1,...,2(n-1)$, and c_h, $h = 1,2$, form an orthonormal basis in Q and $Q^{\perp}$, the orthogonal complement of Q, respectively. Thus Q becomes isomorphic to the vector space $(\mathbb{R}^2)^{n-1}$.

The decomposition $Q_0 = Q \oplus Q^{\perp}$ is a restatement of an elementary fact of mechanics. Let B denote the center-of-mass vector:

$$B = \sum_{j=1}^{n} m_j y_j\Big(\sum_{j=1}^{n} m_j\Big)^{-1} = \sum_{h=1}^{2} B^h e_h\,. \tag{2.9}$$

Then any n-tuple $y = (y_1,...,y_n)$ in Q_0 is, as usual, broken up into

$$(y_1,...,y_n) = (x_1,...,x_n) + (B,...,B)\,, \tag{2.10}$$

where $x = (x_1,...,x_n)$ is in Q. The right-hand side of (2.10) is expressed in terms of f_k and c_h as

$$y = \sum_{k=1}^{2(n-1)} q^k f_k + \sum_{h=1}^{2} \frac{B^h c_h}{N_0}\,,\quad q^k = K(x,f_k)\,. \tag{2.11}$$

Thus one has the following.

Corollary 2.2: The $2n$ variables $(B^h/N_0, q^k)$, $h = 1,2$, $k = 1,...,2(n-1)$, serve as the Cartesian coordinates in Q_0. In particular (q^k) are the Cartesian coordinates in Q.

It is of practical importance to note that the (q^k) has a realization in $\mathbb{R}^2$ as a system of Jacobi vectors. In fact, after a calculation with (2.6), we can get

$$q^{2j-1}e_1 + q^{2j}e_2$$
$$= \Big(m_{j+1}\sum_{k=1}^{j} m_k\Big)^{1/2}\Big(\sum_{k=1}^{j+1} m_k\Big)^{-1/2}$$
$$\times\Big[x_{j+1} - \sum_{k=1}^{j} m_k x_k\Big(\sum_{k=1}^{j} m_k\Big)^{-1}\Big]\,. \tag{2.12}$$

The vectors in the right-hand side are known as Jacobi vectors and used frequently in several-particle systems. The vectors in the left-hand side lead us to the introduction of the complex vector space structure in $Q \cong (\mathbb{R}^2)^{n-1}$. We set

$$z^j = q^{2j-1} + iq^{2j}\,,\quad i = \sqrt{-1}\,,\quad j = 1,...,n-1\,. \tag{2.13}$$

In what follows, the complex coordinates (z^j) will be extensively used.

We return to the SO(2) action defined in (2.4).

Proposition 2.3: Let $g(t) \in SO(2)$;

$$g(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}. \tag{2.14}$$

Then the action of SO(2) on Q is expressed with respect to the basis $\{f_k\}$ in the block diagonal form

$$\begin{pmatrix} g(t) & & \\ & \ddots & \\ & & g(t) \end{pmatrix}, \tag{2.15}$$

where missing matrix entries are all zero. The action on $Q^{\perp}$ is expressed in the same form as (2.14) with respect to the basis $\{c_h\}$.

Proof: Computing $K(g(t)f_j, f_k)$ to get the coefficients of $g(t)f_j = \Sigma a_{kj} f_k$ results in (2.15). This ends the proof.

Corollary 2.4: The $SO(2) \cong U(1)$ action on $Q \cong \mathbb{C}^{n-1}$ is also expressed with respect to the complex variables (2.13) as

$$(z^1,...,z^{n-1}) \to (e^{it}z^1,...,e^{it}z^{n-1})\,. \tag{2.16}$$

B. The principal fiber bundle

From (2.15) or (2.16) it follows that the SO(2) action [or U(1) action] on Q is free if the origin of Q is removed. Further, we see from (2.12) that the origin corresponds to the collision of all the particles at the center of mass. By $\dot{Q}$ we mean the Q whose origin is removed. The $\dot{Q}$ becomes diffeomorphic to $\dot{\mathbb{C}}^{n-1} := \mathbb{C}^{n-1} - \{0\}$. We now show the following theorem.

Theorem 2.5: For the planar n-atomic molecule, the center-of-mass system $\dot{Q}$ without the origin is made into a principal fiber bundle with structure group U(1). The base manifold $M := \dot{Q}/U(1)$, called the internal space, is diffeomorphic to $\mathbb{R}^+ \times \mathbb{C}P^{n-2}$, where $\mathbb{R}^+$ and $\mathbb{C}P^{n-2}$ denote the positive real numbers and the complex projective space, respectively. Let the natural projection be denoted by π. Then one has

$$\pi: \dot{Q} \cong \dot{\mathbb{C}}^{n-1} \to M = \dot{Q}/U(1) \cong \mathbb{R}^+ \times \mathbb{C}P^{n-2}\,. \tag{2.17}$$

Proof: Note that $\dot{\mathbb{C}}^{n-1} \cong \mathbb{R}^+ \times S^{2n-3}$. It is proved in Refs. 6 and 8 that S^{2m+1} is a principal fiber bundle over $\mathbb{C}P^m$ with structure group U(1); $S^{2m+1} \to \mathbb{C}P^m$, called the Hopf fibering. The action of the U(1) is expressed in the form (2.16) with $m = n-2$ and $\Sigma|z^i|^2 = 1$. Thus we have (2.17). This ends the proof.

Remark: For $n = 3$, we recover a result in Ref. 3 (Theorem 4). Note that the base manifold, diffeomorphic with $\mathbb{R}^+ \times \mathbb{C}P^1$, is then diffeomorphic with $\mathbb{R}^+ \times S^2 \cong \mathbb{R}^3$.

We now define a local coordinate system that gives a local picture of the principal fiber bundle (2.17). Let U_{n-1} denote the open subset of $\mathbb{C}^{n-1}$ such that $z^{n-1} \neq 0$. Then we can introduce the local coordinates in U_{n-1} by

$$z^a = re^{i\theta}\rho w^a\,,\quad a = 1,...,n-2\,,\quad z^{n-1} = re^{i\theta}\rho\,,$$

with

$$r^2=\sum_{j=1}^{n-1}|z^j|^2,\quad \rho^{-2}=1+\sum_{a=1}^{n-2}|w^a|^2. \tag{2.18}$$

The θ and r are the coordinates of S^1, the fiber, and $\mathbb{R}^+$, respectively, and (w^a), $a=1,\ldots,n-2$, serve as local coordinates in $\mathbb{C}P^{n-2}$. The internal space, which is, so to speak, the space of all molecule forms independent of the position in $\mathbb{R}^2$, has then the local coordinate system (r,w^a). Note also that r^2 is the moment of inertia of the molecule;

$$r^2=\sum_{j=1}^{n-1}|z^j|^2=\sum_{k=1}^{2(n-1)}(q^k)^2$$
$$=K(x,x)=\sum_{j=1}^{n}m_j(x_j|x_j),\quad x\in Q. \tag{2.19}$$

When restricted to $r=1$, the coordinate system (2.18) becomes the one used by Trautman[9] for discussing the Yang–Mills equation associated with the Hopf fibering $S^{2m+1}\to\mathbb{C}P^m$. Another choice of local coordinate system is, of course, possible for the open subset U_j of $\mathbb{C}^{n-1}$ such that $z^j\neq 0$, $1\leqslant j\leqslant n-1$.

C. The connection and curvature

In this section we apply the connection theory due to Guichardet[2] to our principal fiber bundle $\dot{\mathbb{C}}^{n-1}\to\mathbb{R}^+\times\mathbb{C}P^{n-2}$. Rotational vector fields are defined as the infinitesimal generators of the rotation group. In our case, since that group action is given by (2.16), the rotational vector field is found to be

$$F=\sum_{j=1}^{n-1}\left(-q^{2j}\frac{\partial}{\partial q^{2j-1}}+q^{2j-1}\frac{\partial}{\partial q^{2j}}\right)$$
$$=i\sum_{j=1}^{n-1}\left(z^j\frac{\partial}{\partial z^j}-\bar{z}^j\frac{\partial}{\partial \bar{z}^j}\right),\quad i=\sqrt{-1}, \tag{2.20}$$

where

$$\frac{\partial}{\partial z^j}=\frac{1}{2}\left(\frac{\partial}{\partial q^{2j-1}}-i\frac{\partial}{\partial q^{2j}}\right),$$
$$\frac{\partial}{\partial \bar{z}^j}=\frac{1}{2}\left(\frac{\partial}{\partial q^{2j-1}}+i\frac{\partial}{\partial q^{2j}}\right).$$

In Ref. 6, the generators of the structure group action are called fundamental or vertical as long as there is a connection. The vector field F is interpreted also as the total angular momentum.

The vector fields Y orthogonal to F are called vibrational[2] (or horizontal[6]);

$$K_x(Y_x,F_x)=0,\quad x\in\dot{Q}, \tag{2.21}$$

where K_x is the inner product naturally induced in the tangent space $T_x(\dot{Q})$. For convenience, we extend the definition of K_x to complex vectors so that K_x may be a symmetric bilinear form on $T_x(\dot{Q})^{\mathbb{C}}$, the complexified tangent space. Then, for a vector field

$$Y=\sum\left(\xi^j\frac{\partial}{\partial z^j}+\eta^j\frac{\partial}{\partial \bar{z}^j}\right)$$

to be vibrational, it is necessary and sufficient that

$$\sum(z^j\eta^j-\bar{z}^j\xi^j)=0. \tag{2.22}$$

To see this, we have only to note that from Corollary 2.2 K_x can be expressed as the standard flat metric on $\mathbb{C}^{n-1}$;

$$K_x=\sum_{j=1}^{n-1}d\bar{z}^j\,dz^j. \tag{2.23}$$

In the following, differential forms on $\dot{Q}$ will be tacitly assumed to be extended in order to be defined for complex vector fields on $\dot{Q}$.

Let $W_{x,\mathrm{vib}}$ denote the linear subspace of $T_x(\dot{Q})$ spanned by all the vibrational tangent vectors at x. The connection due to Guichardet[2] is the assignment: $x\to W_{x,\mathrm{vib}}$. The connection is given, in a dual manner, in terms of differential forms as follows.

Theorem 2.6: The connection form ω defined on the center-of-mass system is expressed as

$$\omega=\left(\sum_{k=1}^{2(n-1)}(q^k)^2\right)^{-1}\sum_{j=1}^{n-1}(q^{2j-1}\,dq^{2j}-q^{2j}\,dq^{2j-1})$$
$$=i\left(2\sum_{j=1}^{n-1}|z^j|^2\right)^{-1}\sum_{j=1}^{n-1}(z^j d\bar{z}^j-\bar{z}^j\,dz^j). \tag{2.24}$$

Proof: We identify u(1), the Lie algebra of U(1), with $\mathbb{R}$. For the rotational vector field F and vibrational vector fields Y, we obtain, from (2.20), (2.22), and (2.24),

$$\omega(F)=1,\quad \omega(Y)=0. \tag{2.25}$$

Thus, by definition,[6] ω proves to be the connection form.

Note that the connection form ω is associated only with the inertia moment and the angular momentum. This is observed from (2.19) and (2.20).

The curvature form Ω is given by the structure equation[6] $\Omega=d\omega-\omega\wedge\omega$. Applied to (2.24), this formula gives the following.

Theorem 2.7: The curvature form Ω is expressed as

$$\Omega=i\left(\sum_{j=1}^{n-1}|z^j|^2\right)^{-2}\left(\sum_{j=1}^{n-1}|z^j|^2\sum_{k=1}^{n-1}dz^k\wedge d\bar{z}^k\right.$$
$$\left.-\sum_{j,k=1}^{n-1}\bar{z}^k z^j\,dz^k\wedge d\bar{z}^j\right), \tag{2.26}$$

and defines a two-form on $\mathbb{C}P^{n-2}$.

Proof: When applied to (2.24), the formula $\Omega=d\omega$ in our case results in (2.26) after a calculation. The curvature form is known to vanish for any vertical vector fields. For (2.26), we can verify this fact by a simple calculation with (2.20). Further, as is easily seen, Ω is invariant under the U(1) action, so that Ω defines a two-form on the internal space $\mathbb{R}^+\times\mathbb{C}P^{n-2}$. Moreover, since Ω vanishes for the radial vector field $\Sigma(z^k\,\partial/\partial z^k+\bar{z}^k\,\partial/\partial\bar{z}^k)$, it comes to define a two-form on $\mathbb{C}P^{n-2}$. This ends the proof.

We proceed to express the connection and curvature forms in the local coordinates introduced by (2.18). A straightforward calculation with (2.18) yields

$$\omega=d\theta+\frac{i}{2}\frac{\Sigma(w^a d\bar{w}^a-\bar{w}^a\,dw^a)}{1+\Sigma|w^a|^2}, \tag{2.27}$$

$$\Omega=i\frac{(1+\Sigma|w^a|^2)\Sigma\,dw^a\wedge d\bar{w}^a-\Sigma\,\bar{w}^a w^b\,dw^a\wedge d\bar{w}^b}{(1+\Sigma|w^a|^2)^2}, \tag{2.28}$$

where a and b run over $1,2,\ldots,n-2$. In view of this expres-

sion, Ω becomes equal to the fundamental form associated with the Fubini–Study metric[6,10] on $\mathbb{C}P^{n-2}$.

D. Rotation and vibration

In this section we return to the rotational and vibrational vectors defined in Sec. II C in connection with the Eckart condition of rotationless constraint. As is observed from (2.25), the Eckart condition is described as the differential equation $\omega = 0$. However, this equation is not completely integrable,[11] because of $\Omega \neq 0$. Alternatively, the assignment $x \to W_{x,\mathrm{vib}}$ is not completely integrable.[11] Accordingly, there are no submanifolds of the center-of-mass system on which every motion is vibrational. Here a curve or a motion is called vibrational if its tangent vectors at every point of the curve are vibrational. This nonintegrability is worth reinvestigation. To this end, we wish to work with the coordinates introduced by (2.18).

The rotational vector field F given by (2.20) is written as

$$F = \frac{\partial}{\partial \theta}. \tag{2.29}$$

Any rotational vector field is a multiple of F. We turn to vibrational vector fields. Let X be a vector field on the internal space M. The horizontal (or vibrational) lift X^* of X is defined as a unique vector field on $\dot{Q}$ which projects to X; $\pi_* X^*_x = X_{\pi(x)}$, $x \in \dot{Q}$, where π_* is the tangent map of π: $\dot{Q} \to M$. For local vector fields $\partial/\partial r$, $\partial/\partial w^a$, and $\partial/\partial \bar{w}^a$ on the open set U_{n-1}, we obtain, using the condition $\omega((\partial/\partial w^a)^*) = 0$, etc.,

$$\begin{aligned}
&\left(\frac{\partial}{\partial r}\right)^* = \frac{\partial}{\partial r}, \\
&\left(\frac{\partial}{\partial w^a}\right)^* = \frac{\partial}{\partial w^a} + \frac{i}{2}\rho^2 \bar{w}^a \frac{\partial}{\partial \theta}, \quad a = 1,\ldots,n-2, \\
&\left(\frac{\partial}{\partial \bar{w}^a}\right)^* = \frac{\partial}{\partial \bar{w}^a} - \frac{i}{2}\rho^2 w^a \frac{\partial}{\partial \theta}, \quad a = 1,\ldots,n-2,
\end{aligned} \tag{2.30}$$

where ρ^2 was defined in (2.18). We notice that every local vibrational vector field on U_{n-1} is expressed as a linear combination of these horizontal vectors.

The Jacobi brackets among these vectors are calculated to give

$$\begin{aligned}
&\left[\frac{\partial}{\partial r}, \left(\frac{\partial}{\partial w^a}\right)^*\right] = \left[\frac{\partial}{\partial r}, \left(\frac{\partial}{\partial \bar{w}^a}\right)^*\right] = 0, \\
&\left[\left(\frac{\partial}{\partial w^a}\right)^*, \left(\frac{\partial}{\partial w^b}\right)^*\right] = \left[\left(\frac{\partial}{\partial \bar{w}^a}\right)^*, \left(\frac{\partial}{\partial \bar{w}^b}\right)^*\right] = 0, \\
&\left[\left(\frac{\partial}{\partial w^a}\right)^*, \left(\frac{\partial}{\partial \bar{w}^b}\right)^*\right] = i\frac{\bar{w}^a w^b - (1 + \Sigma|w^a|^2)\delta_{ab}}{(1 + \Sigma|w^a|^2)^2}\frac{\partial}{\partial \theta}.
\end{aligned} \tag{2.31}$$

These equations show that the assignment $x \to W_{x,\mathrm{vib}}$ is not completely integrable (Frobenius' theorem[11]), and are capable of the following interpretation: The infinitesimal vibrations $(\partial/\partial w^a)^*$ and $(\partial/\partial \bar{w}^b)^*$ are coupled to give rise to the rotational vector field, the right-hand side of the last equation in (2.31). This is a reason why the rotation and vibration cannot be separated to each other. We note also that the right-hand sides of (2.31) give the components of the curvature tensor (2.28);

$$\begin{aligned}
&\Omega\left(\left(\frac{\partial}{w^a}\right)^*, \left(\frac{\partial}{\partial \bar{w}^b}\right)^*\right) \\
&\quad = -\omega\left(\left[\left(\frac{\partial}{\partial w^a}\right)^*, \left(\frac{\partial}{\partial \bar{w}^b}\right)^*\right]\right), \quad \text{etc.}
\end{aligned}$$

This is why the nonintegrability is measured by the curvature.

E. The induced metric on the internal space

Recall that we have the standard flat metric (2.23) on the center-of-mass system Q. This metric is, of course, invariant under the U(1) action and induces a Riemannian metric B on the internal space M as follows: Let X and Y be vector fields on M, and X^* and Y^* their horizontal lifts, respectively. Then the induced metric B is well defined by

$$B_{\pi(x)}(X,Y) = K_x(X^*, Y^*), \quad x \in \dot{Q}. \tag{2.32}$$

Theorem 2.8: The induced metric B on the internal space $\mathbb{R}^+ \times \mathbb{C}P^{n-2}$ is expressed as

$$dr^2 + r^2\, d\sigma^2, \tag{2.33}$$

where $d\sigma^2$ is the Fubini–Study metric on $\mathbb{C}P^{n-2}$.

Proof: To get (2.33), we use the local coordinates (2.18). Inserting (2.18) into (2.23) results in

$$K_x = r^2\omega^2 + dr^2 + r^2\, d\sigma^2, \tag{2.34}$$

where ω is the curvature form given by (2.27) and

$$d\sigma^2 = \frac{(1 + \Sigma|w^a|^2)\Sigma\, dw^a\, d\bar{w}^a - \Sigma\, \bar{w}^a w^b\, dw^a\, d\bar{w}^b}{(1 + \Sigma|w^a|^2)^2}. \tag{2.35}$$

This is known as the Fubini–Study metric on the complex projective space.[6] The expression (2.34) was also obtained by Warner,[12] and reduces, when restricted to $r = 1$, to the one used by Trautman.[9] Recalling that for horizontal vector fields the connection form vanishes, we come to the desired conclusion (2.33).

Remark: For complex-valued vector fields X and Y, we can get a "Hermitian" metric on G on M by setting

$$G_{\pi(x)}(X,Y) = K_x(\bar{X}^*, Y^*). \tag{2.36}$$

For later use, we discusss below the inner product K^*_x defined in the cotangent space $T^*_x(\dot{Q})$. Let K^b_x be the isomorphism of $T_x(\dot{Q})$ to $T^*_x(\dot{Q})$;

$$K^b_x(u) \cdot v = K_x(u,v), \quad u,v \in T_x(\dot{Q}), \tag{2.37}$$

and set $(K^b_x)^{-1} = K^\#_x$. Then the inner product K^*_x is defined for $p,q \in T^*_x(\dot{Q})$ by

$$K^*_x(p,q) = K_x(K^\#_x(p), K^\#_x(q)). \tag{2.38}$$

Like K_x, K^*_x is extended so as to be a symmetric bilinear form on $T^*_x(\dot{Q})^{\mathrm{C}}$.

Let the components of the metric $d\sigma^2$ be denoted by $(g_{a\bar{b}})$, and set $g_{\bar{a}b} = g_{b\bar{a}}$. By $(g^{\bar{a}b})$ and $(g^{a\bar{b}})$ we mean the inverse matrices of $(g_{a\bar{b}})$ and $(g_{\bar{a}b})$, respectively;

$$d\sigma^2 = \frac{1}{2}\sum_{a,b}(g_{a\bar{b}}\, dw^a\, d\bar{w}^b + g_{\bar{a}b}\, d\bar{w}^a\, dw^b),$$

$$\sum_b g^{\bar{a}b} g_{b\bar{c}} = \delta^a_c\,, \quad \sum_b g^{a\bar{b}} g_{\bar{b}c} = \delta^a_c\,.$$

Then one has, from (2.34), (2.36), and (2.38),

$$\begin{aligned} &K^*_x(\omega,\omega) = 1/r^2\,, \\ &K^*_x(dr,dr) = 1 = G^*(dr,dr)\,, \\ &K^*_x(dw^a,d\,\overline{w}^b) = 2g^{a\bar{b}}/r^2 = G^*(d\,\overline{w}^a,d\,\overline{w}^b)\,, \\ &K^*_x(d\,\overline{w}^a,dw^b) = 2g^{\bar{a}b}/r^2 = G^*(dw^a,dw^b)\,, \end{aligned} \tag{2.39}$$

and the other vanishing.

Let f be a complex-valued function on Q. Then its differential df is written as

$$df = \frac{\partial f}{\partial \theta}\,\omega + \frac{\partial f}{\partial r}\,dr + \sum \left(\frac{\partial}{\partial w^a}\right)^* f\,dw^a + \sum \left(\frac{\partial}{\partial\,\overline{w}^a}\right)^* f\,d\,\overline{w}^a\,. \tag{2.40}$$

Using this together with (2.39), we obtain

$$\begin{aligned} K^*_x(\,\overline{df},df) = {} & \frac{1}{r^2}\,\frac{\overline{\partial f}}{\partial\theta}\,\frac{\partial f}{\partial\theta} + \frac{\overline{\partial f}}{\partial r}\,\frac{\partial f}{\partial r} \\ & + \frac{2}{r^2}\sum g^{a\bar{b}} \left(\frac{\partial}{\partial w^a}\right)^* \bar{f} \left(\frac{\partial}{\partial\,\overline{w}^b}\right)^* f \\ & + \frac{2}{r^2}\sum g^{\bar{a}b} \left(\frac{\partial}{\partial\,\overline{w}^a}\right)^* \bar{f} \left(\frac{\partial}{\partial w^b}\right)^* f\,. \end{aligned} \tag{2.41}$$

This is nothing but a kinetic energy density. The first term of the right-hand side gives the rotational energy density, and the rest the vibrational energy density.

III. THE ASSOCIATED COMPLEX LINE BUNDLES

The internal space M may be viewed as the space reduced from the center-of-mass system Q by separating off the rotation angle variable. Accordingly, the conservation of the total angular momentum is brought into effect for describing the internal states of the molecule. Take up the open subset U_{n-1} considered in (2.18). Since U_{n-1} is broken up into a direct product $S^1 \times \pi(U_{n-1})$, any local function on U_{n-1} may be expanded into a Fourier series in the rotation angle, an expansion in the eigenfunctions of the total angular momentum operator. Then the internal state of an assigned momentum eigenvalue will be singled out as a Fourier coefficient of this series. However, this idea can be carried out only locally, because the total space $\dot{Q}$ is not broken up into a product space of S^1 and M. Therefore, for the global description of the internal states, we must pass to complex line bundles over the internal space. We will soon see that the introduction of the complex line bundles is a geometric consequence of the conservation of the total angular momentum.

A. The associated complex line bundles

Fix an integer m and let ρ_m denote the representation of $U(1)$ given by

$$\rho_m(e^{it})\colon\ \zeta \to e^{imt}\zeta\,, \quad \zeta \in \mathbb{C}\,. \tag{3.1}$$

Define a left action of $U(1)$ on $\dot{Q} \times \mathbb{C}$ by

$$(z,\zeta) \to (e^{it}z, e^{imt}\zeta)\,, \tag{3.2}$$

where $z = (z^1,\ldots,z^{n-1})$ is used instead of x. This action defines an equivalence relation in $\dot{Q} \times \mathbb{C}$. Then the quotient manifold, denoted by $\dot{Q} \times_m \mathbb{C}$, is made into a complex line bundle $L_m = (\dot{Q} \times_m \mathbb{C}, \pi_m, M)$ via the commutative diagram

$$\begin{array}{ccc} \dot{Q} \times \mathbb{C} & \xrightarrow{\ q\ } & \dot{Q} \times_m \mathbb{C} \\ {\scriptstyle \mathrm{pr}}\downarrow & & \downarrow{\scriptstyle \pi_m}\,, \\ \dot{Q} & \xrightarrow{\ \pi\ } & M \end{array} \tag{3.3}$$

where pr and q denote the projection onto the first factor and the natural projection, respectively. The L_m is called the complex line bundle associated with the principal fiber bundle $\dot{Q} \to M$. A map $s\colon M \to \dot{Q} \times_m \mathbb{C}$ is called a cross section in L_m if $\pi \circ s = \mathrm{id}_M$, id_M being the identity of M. The space of square integrable cross sections in L_m will serve as the space of internal states of the molecule.

A complex-valued function f on $\dot{Q}$ is called ρ_m equivariant if it satisfies

$$f(e^{it}z, e^{-it}\,\overline{z}) = e^{imt} f(z,\overline{z})\,. \tag{3.4}$$

The ρ_m-equivariant functions are in one-to-one correspondence with the cross sections in L_m.[6] This correspondence is denoted by $q^{\#}_m$. For a cross section s in L_m one then has

$$(q^{\#}_m s)(x) = q_x^{-1}(s(\pi(x)))\,, \quad x \in \dot{Q}\,, \tag{3.5}$$

where $q_x\colon \mathbb{C} \to \pi_m^{-1}(\pi(x))$ denotes the isomorphism restricted from q on fibers; $q_x(\zeta) = [(x,\zeta)]$, $[\ ]$ denoting the equivalence class.

We now take up investigation into what the $q^{\#}_m$ means in quantum mechanics. The total angular momentum operator is defined by $\hat{F} = -iF$, where F is the rotational vector given by (2.20). If we differentiate Eq. (3.4) with respect to t, we get

$$\hat{F} f = mf\,. \tag{3.6}$$

This means that the equivariant function f is an eigenfunction of $\hat{F}$. Thus picking up the ρ_m-equivariant functions out of wave functions and, therefore, introducing the complex line bundles L_m amount to making geometric use of the conservation of the total angular momentum.

We turn to the local description of cross sections and to gauge transformations, in order to understand that cross sections in L_m are closely related with coefficients of the eigenfunction expansion for the total angular momentum operator. Define a local ρ_m-equivariant function ϕ_k on an open subset U_k as follows:

$$\begin{aligned} &\phi_k = (z^k/|z^k|)^m\,, \quad 1 \leqslant k \leqslant n-1\,, \\ &U_k = \{(z^1,\ldots,z^{n-1});\ z^k \neq 0\}\,. \end{aligned} \tag{3.7}$$

On the nonempty intersection $U_k \cap U_j$, one has

$$\phi_k = (z^k |z|^j / z^j |z|^k)^m \phi_j\,. \tag{3.8}$$

Let s_k denote the corresponding local cross section in L_m; $q^{\#}_m s_k = \phi_k$. Suppose a cross section s in L_m is expressed as $f_k s_k = f_j s_j$ on $\pi(U_k \cap U_j)$. Then, put together with (3.8), this expression gives rise to the gauge transformation

$$f_k = (z^j |z^k| / z^k |z^j|)^m f_j \quad \text{on } \pi(U_j \cap U_k)\,. \tag{3.9}$$

We note that, for $k = n-1$, ϕ_k becomes $e^{im\theta}$, an eigenfunction of $\hat{F}$ [see (2.29)], in the coordinates (2.18). For any k, a local cross section $f_k s_k$ corresponds to an eigenstate $f_k\phi_k$ of $\hat{F}$. Thus the f_k, an internal state defined on $\pi(U_k)$, is thought of as a Fourier coefficient of the Fourier series for a local wave function on U_k. The gauge transformation (3.9) thus gives the law of piecing together locally defined internal states f_k. In the literature the coefficients in (3.9) are called transition functions (see Refs. 7 and 8, for example).

B. The linear connection and curvature

The connection defined on the principal fiber bundle $\dot{Q}\to M$ can be carried over into the associated complex line bundle L_m. Let X be a vector field on M, and X^* its horizontal lift. Then for a cross section s in L_m its covariant derivative with respect to X is defined by

$$\nabla_X s = q_m^{\#\,-1} X^*(q_m^{\#} s)\,. \tag{3.10}$$

The operator ∇ is called the linear connection, which is linear in X and s, and satisfies for arbitrary functions f the conditions

$$\nabla_{fX} s = f\nabla_X s\,,\quad \nabla_X(fs) = (Xf)s + f\nabla_X s\,. \tag{3.11}$$

The curvature of ∇ is defined for vector fields X and Y on M by

$$R(X,Y)s = [\nabla_X,\nabla_Y]s - \nabla_{[X,Y]}s\,. \tag{3.12}$$

Theorem 3.1: The curvature of the linear connection on L_m is expressed as

$$R(X,Y) = -im\Omega(X,Y)\,,\quad m\in\mathbb{Z}\,, \tag{3.13}$$

where X and Y are vector fields on M, and Ω is the curvature form on the principal fiber bundle $\dot{Q}\to M$.

Proof: By combining Eqs. (3.10) and (3.12),

$$R(X,Y)s = q_m^{\#\,-1}([X^*,Y^*] - [X,Y]^*)(q_m^{\#} s)\,. \tag{3.14}$$

We use the local coordinates (2.18) for the right-hand side of (3.14). We notice here that for a ρ_m-equivariant function $q_m^{\#}s$ one has, from (2.29) and (3.6),

$$\frac{\partial}{\partial\theta} q_m^{\#} s = imq_m^{\#} s\,. \tag{3.15}$$

We now apply (3.14) together with (2.28), (2.31), and (3.15) to obtain

$$R\left(\frac{\partial}{\partial w^a},\frac{\partial}{\partial \bar{w}^b}\right) = -im\Omega\left(\frac{\partial}{\partial w^a},\frac{\partial}{\partial \bar{w}^b}\right),\ \text{etc.}$$

This completes the proof, because Eq. (3.13) is a tensor equation.

As was stated in Theorem 2.7, the two-form Ω defines a two-form on $\mathbb{C}P^{n-2}$; so does the curvature R. Hence we can show that $[R]/2\pi i$ is an integral cohomology class, following Wells, Jr.,[7] for example; taking up $\mathbb{C}P^1\subset\mathbb{C}P^{n-2}$, as a two-cycle, to be defined in U_{n-1} by $z^1 = \cdots = z^{n-3} = 0$ and $r = 1$, we have, along with $w^{n-2} = w$,

$$R\,|_{CP^1} = m(dw\wedge d\bar{w})/(1+|w|^2)^2\,,$$

which yields, after integration,

$$\frac{1}{2\pi i}\int_{CP^1} R = -m\,. \tag{3.16}$$

This equation may also be viewed as giving the quantized strength of the magnetic field R. For $n = 3$, we have proved the same equation in Ref. 3, in which we interpreted the curvature as a magnetic monopole field of the quantized strength on the internal space $\mathbb{R}^+\times\mathbb{C}P^1\cong\dot{\mathbb{R}}^3$.

The internal motion of the molecule is coupled with the magnetic field R through a locally defined gauge potential. This is observed when we express the connection ∇ in local coordinates (2.18). Take up the horizontal vector fields (2.30) defined on U_{n-1}. Then by definition (3.10) along with (3.15) we obtain

$$\begin{aligned}
&\nabla_r s = \frac{\partial}{\partial r} s\,,\quad \nabla_a s = \left(\frac{\partial}{\partial w^a} + im\frac{i}{2}\rho^2\bar{w}^a\right)s\,,\\
&\nabla_{\bar{a}} s = \left(\frac{\partial}{\partial \bar{w}^a} - im\frac{i}{2}\rho^2 w^a\right)s\,,
\end{aligned} \tag{3.17}$$

where ∇_r, ∇_a, $\nabla_{\bar{a}}$ stand for $\nabla_{\partial/\partial r}$, $\nabla_{\partial/\partial w^a}$, $\nabla_{\partial/\partial\bar{w}^a}$, respectively. These equations prove the above assertion, because

$$m\frac{i}{2}\rho^2\sum(w^a\,d\bar{w}^a - \bar{w}^a\,dw^a) \tag{3.18}$$

is a locally defined gauge potential for mR [see (2.27) and (2.28)].

C. The inner product for cross sections

The inner product for cross sections in L_m should be derived from that for functions on Q. The volume element on M for integration must be reduced as well from the standard one on Q. The configuration space Q_0 has the standard volume element

$$dQ_0 = dy_1\wedge\cdots\wedge dy_n\,,$$

where $dy_k = dy_k^1\wedge dy_k^2$, $k = 1,\ldots,n$. The volume element dV_0 defined by the inner product K is related to dQ_0 by

$$dV_0 = m_1\cdots m_n\,dQ_0\,.$$

According to (2.11), the dV_0 is expressed also as

$$dV_0 = N_0^{-2}\,dB^1\wedge dB^2\wedge dV\,,$$

where N_0 is the normalization constant and

$$dV = dq^1\wedge\cdots\wedge dq^{2(n-1)} \tag{3.19}$$

is the volume element on Q defined by the inner product K. Thus, separating off the center-of-mass coordinates from dQ_0, we obtain the volume element dQ on Q in the form

$$dQ = \mu\,dV\,,\quad \mu = \sum_{k=1}^{k=n} m_k\left[\prod_{k=1}^{k=n} m_k\right]^{-1}. \tag{3.20}$$

To bring out the volume element dM on M, we refer to the expression (2.34) of the standard flat metric K_x on Q. In view of (2.34), we obtain dV (up to sign) in the form

$$\begin{aligned}
dV &= r\omega\wedge dr\wedge r^{2(n-2)}\,dS\\
&= \omega\wedge r^{2n-3}\,dr\wedge dS\,,
\end{aligned} \tag{3.21}$$

where dS is the volume element defined on $\mathbb{C}P^{n-2}$ by the Kahler metric $d\sigma^2$. We may take dS in the coordinates (2.18) as

$$dS = (i/2)^{n-2}J_g\,dw^1\wedge d\bar{w}^1\wedge\cdots\wedge dw^{n-2}\wedge d\bar{w}^{n-2}\,, \tag{3.22}$$

where J_g denotes the determinant of the Hermitian matrix $(g_{a\bar{b}})$, $a,b = 1,...,n-2$, and is calculated as

$$J_g = \det(g_{a\bar{b}}) = \left(1 + \sum |w^a|^2\right)^{-n+1} \quad (n \geqslant 2) . \tag{3.23}$$

A straightforward but long calculation brings about the relation

$$[1/(n-2)!]\Omega^{n-2} = i^{n-2} \det(g_{a\bar{b}}) dw^1 \wedge d\bar{w}^1 \wedge \cdots \wedge dw^{n-2} \wedge d\bar{w}^{n-2} . \tag{3.24}$$

Hence, one has

$$dS = [(n-2)!2^{n-2}]^{-1} \Omega^{n-2} . \tag{3.25}$$

In the coordinates (2.18), the volume element dQ becomes

$$dQ = \mu \, d\theta \wedge r^{2n-3} \, dr \wedge dS . \tag{3.26}$$

Now we turn to the inner product for cross sections. For the cross section s_k, $k = 1,2$, in L_m, the Hermitian metric on L_m is given by

$$(s_1|s_2)(\pi(x)) = \overline{q_m^{\#} s_1(x)} \, q_m^{\#} s_2(x) , \tag{3.27}$$

where the overbar indicates the complex conjugate. The right-hand side of (3.27) is clearly invariant under the U(1) action, so that it depends on $\pi(x)$ only.

Theorem 3.2: The inner product for cross sections s_k, $k = 1,2$, is given by

$$\int_M (s_1|s_2) dM = \int_Q \overline{q_m^{\#} s_1} \, q_m^{\#} s_2 \, dQ , \tag{3.28}$$

where dM is the volume element on M defined, together with (3.25), by

$$dM = 2\pi\mu r^{2n-3} \, dr \wedge dS . \tag{3.29}$$

Proof: For ρ_m-equivariant functions $q_m^{\#} s_k = \phi_k$, $k = 1,2$, one has the inner product, using (3.20) and (3.21),

$$\int_Q \overline{\phi_1} \, \phi_2 \, dQ = \mu \int_{S^1} \int_M \overline{\phi_1} \, \phi_2 \omega \wedge r^{2n-3} \, dr \wedge dS = 2\pi\mu \int_M \overline{\phi_1} \, \phi_2 r^{2n-3} \, dr \wedge dS . \tag{3.30}$$

Here we have used the fact that $\overline{\phi_1}\phi_2$ is constant on each fiber $\pi^{-1}(p) \cong S^1$, $p \in M$. Then Eqs. (3.27) and (3.30) are put together to prove Eq. (3.28). We note in conclusion that the volume element dM on M is not equal to that defined by the induced Riemannian metric (2.33) on M. The latter is written as $r^{2n-4} \, dr \wedge dS = (2\pi\mu r)^{-1} \, dM$.

IV. QUANTUM MECHANICS FOR INTERNAL STATES

We are now in a position to set up quantum mechanics for internal states of the planar molecule. What we have to do is to bring out the internal Hamiltonian operator acting on cross sections from the standard Hamiltonian operator on the center-of-mass system.

A. The Laplacian

We start with the kinetic energy operator. The standard one on the configuration space Q_0 is, of course, given by

$$-\frac{1}{2} \sum_{k=1}^{n} \frac{1}{m_k} \left(\frac{\partial}{\partial y_k}\right)^2 , \tag{4.1}$$

where $\partial/\partial y_k$, $k = 1,...,n$, denote the gradient operators. The operator (4.1) is thought of as $-\frac{1}{2}$ times the Laplacian Δ_0 with respect to the inner product K on Q_0. By Corollary 2.2, Δ_0 turns out to have the form

$$\Delta_0 = N_0^2 \sum_{h=1}^{2} \left(\frac{\partial}{\partial B^h}\right)^2 + \sum_{j=1}^{2(n-1)} \left(\frac{\partial}{\partial q^j}\right)^2 . \tag{4.2}$$

Separating off the center-of-mass coordinates, we have the Laplacian on the center-of-mass system Q in the form

$$\Delta = \sum_{j=1}^{2(n-1)} \left(\frac{\partial}{\partial q^j}\right)^2 = 4 \sum_{k=1}^{n-1} \frac{\partial^2}{\partial z^k \, \partial \bar{z}^k} . \tag{4.3}$$

We should here note that the Δ can be derived from the energy functional by integration by parts. Let f be a wave function on Q. Then its energy density is given by

$$K_x^*(\overline{df}, df) = 2 \sum_{k=1}^{n-1} \left(\frac{\partial \bar{f}}{\partial z^k} \frac{\partial f}{\partial \bar{z}^k} + \frac{\partial \bar{f}}{\partial \bar{z}^k} \frac{\partial f}{\partial z^k}\right) , \tag{4.4}$$

where K_x^* was defined in (2.38). Thus one obtains, by integration by parts,

$$\int_Q K_x^*(\overline{df}, df) \, dQ = -\int_Q \bar{f} \Delta f \, dQ . \tag{4.5}$$

We now wish to express Δ in the coordinates (2.18). Let f have its support in U_{n-1}. Then, from the expressions (2.41) and (3.26) of $K_x^*(\overline{df}, df)$ and dQ, respectively, we obtain, after integration of the left-hand side of (4.5) by parts,

$$\Delta = \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^{2n-3}} \frac{\partial}{\partial r} \left(r^{2n-3} \frac{\partial}{\partial r}\right) + \frac{2}{r^2} \sum_{a,b} \left[\frac{1}{J_g} \left(\frac{\partial}{\partial \bar{w}^a}\right)^* \left(g^{\bar{a}b} J_g \left(\frac{\partial}{\partial w^b}\right)^*\right) + \frac{1}{J_g} \left(\frac{\partial}{\partial w^a}\right)^* \left(g^{a\bar{b}} J_g \left(\frac{\partial}{\partial \bar{w}^b}\right)^*\right)\right] , \tag{4.6}$$

where $J_g = \det(g_{a\bar{b}})$. The first term of the right-hand side represents the rotational energy operator, and the rest of the terms the vibrational energy operator.

B. The internal Hamiltonian operator

The internal Hamiltonian operator acting on the cross sections should be derived from the standard one on Q,

$$H = -\tfrac{1}{2}\Delta + V ,$$

where V is a potential function depending on the internal coordinates only. Let s be a cross section in L_m. Then the internal Hamiltonian operator H_m is defined through

$$\int_M (s|H_m s) \, dM = \int_Q \overline{q_m^{\#} s} \, H(q_m^{\#} s) \, dQ . \tag{4.7}$$

A local expression of H_m is easy to obtain.

Proposition 4.1: In the local coordinates (2.18), the internal Hamiltonian H_m takes the form

$$H_m = -\frac{1}{2}\left[\frac{1}{2^{2n-3}}\frac{\partial}{\partial r}\left(r^{2n-3}\frac{\partial}{\partial r}\right)\right.$$
$$+\frac{2}{r^2}\sum_{a,b}\left(\frac{1}{J_g}\nabla_{\bar a}(g^{\bar a b}J_g\nabla_b)\right.$$
$$\left.\left.+\frac{1}{J_g}\nabla_a(g^{a\bar b}J_g\nabla_{\bar b})\right)\right]+\frac{m^2}{2r^2}+V. \qquad (4.8)$$

Proof: Let s be a local section with compact support in U_{n-1}. We operate a local ρ_m-equivariant function $q_m^{\#}s$ with the Hamiltonian and use the formulas (3.15) and (4.6) together with (3.17) to find (4.8). This ends the proof.

We observe that the Hamiltonian H_m indeed satisfies the principle of minimal coupling, which requires that when the magnetic field mR is turned on, the operator $\partial/\partial w^a$ and $\partial/\partial \overline{w}^a$ should be replaced by the covariant operators ∇_a and $\nabla_{\bar a}$, respectively [see (3.17)]. This coupling is the very influence that the rotation of the molecule produces on the internal motion. The term $m^2/2r^2$ represents the other influence, which shows that a centrifugal potential for the internal motion should be added. We remark further that the same expression as (4.8) is available in the open subset U_k, $1\leqslant k\leqslant n-1$.

To get a global idea of the internal Hamiltonian in a differential geometric setting, we return to the kinetic energy density $K_x^*(\overline{df},df)$. For our present purpose, f has to be a ρ_m-equivariant function. Let $f=q_m^{\#}s$. We are to have a close look at $df=dq_m^{\#}s$. The question that arises is whether one can interchange d and $q_m^{\#}$. To answer this question, we get back to the definition of the connection. An alternative introduction of the connection is made as follows: Let $T^*(M)^{\mathbb{C}}$ be the complexified cotangent bundle. A connection, denoted by d^{∇}, is a $\mathbb{C}$-linear mapping from the space of cross sections in L_m to the space of cross sections in the tensor product bundle $T^*(M)^{\mathbb{C}}\otimes L_m$, which satisfies the Leibnitz formula

$$d^{\nabla}(fs)=df\otimes s+f\,d^{\nabla}s, \qquad (4.9)$$

where f and s are a function on M and a cross section in L_m, respectively. If we apply the formula (4.9) to a vector field X on M and set $d^{\nabla}s(X)=\nabla_X s$, we again find the second of Eqs. (3.11). The first equation of (3.11) is clear, because $d^{\nabla}s$ is required to be an L_m-valued differential form.

We proceed to the next stage to ask if one can extend $q_m^{\#}$ to the cross sections in $T^*(M)^{\mathbb{C}}\otimes L_m$. In other words, how can we define $q_m^{\#}\,d^{\nabla}s$? From definition (3.10), we obtain the equation $q_m^{\#}(d^{\nabla}s(X))=(dq_m^{\#}s)(X^*)$. The right-hand side of this equation implies that the horizontal part of the differential $dq_m^{\#}s$ is picked up. Hence we define d^h as the horizontal part of d; for a function f on Q, one has

$$d^h f=df-(Ff)\omega, \qquad (4.10)$$

where F and ω are the rotational vector field and the connection form, respectively. For F, $d^h f$ vanishes, as seen from (2.25). For vibrational (or horizontal) vector fields Y, one has $d^h f(Y)=df(Y)$.

We now come to the final equation

$$q_m^{\#}d^{\nabla}s=d^h q_m^{\#}s \quad\text{or}\quad d^{\nabla}s=q_m^{\#\,-1}d^h(q_m^{\#}s). \qquad (4.11)$$

This equation solves our question. That is, d and $q_m^{\#}$ do not commute, but the connection leads us to the desirable conclusion (4.11). In the coordinates (2.18), we obtain from (2.40)

$$d^{\nabla}s=dr\otimes\nabla_r s+\sum(dw^a\otimes\nabla_a s+d\overline{w}^a\otimes\nabla_{\bar a}s). \qquad (4.12)$$

Our third task is to define a Hermitian metric on the tensor product bundle $T^*(M)^{\mathbb{C}}\otimes L_m$, which is induced naturally from both the Hermitian metric in $T^*(M)^{\mathbb{C}}$ given by the Riemannian metric on M and the already defined Hermitian metric (3.27) in L_m. In view of (4.11), we define the Hermitian metric on $T^*(M)^{\mathbb{C}}\otimes L_m$ for $d^{\nabla}s_1$ and $d^{\nabla}s_2$ by

$$(d^{\nabla}s_1|d^{\nabla}s_2)=K_x^*(\overline{d^h q_m^{\#}s_1},d^h q_m^{\#}s_2). \qquad (4.13)$$

Note that for horizontal vectors, K_x indeed defines a Riemannian metric on M (Theorem 2.8); so does K_x^* for horizontal covectors. Using the orthogonal decomposition $df=d^h f+(Ff)\omega$, one obtains the kinetic energy density for a ρ_m-equivariant function $q_m^{\#}s$ in the following form:

$$K_x^*(\overline{dq_m^{\#}s},dq_m^{\#}s)=(m^2/r^2)(s|s)+(d^{\nabla}s|d^{\nabla}s). \qquad (4.14)$$

The first term of the right-hand side is the rotational energy density possessed by the molecule of the total angular momentum eigenvalue m, and the last term is the internal energy density coupled with the gauge field mR. In the local coordinates (2.18), definition (4.13) gets easier to understand. From (2.39) and (2.41) we obtain

$$K_x^*(\overline{d^h q_m^{\#}s},d^h q_m^{\#}s)$$
$$=(\nabla_r s|\nabla_r s)+\frac{2}{r^2}\sum[g^{a\bar b}(\nabla_{\bar a}s|\nabla_{\bar b}s)+g^{\bar a b}(\nabla_a s|\nabla_b s)]$$
$$=G^*(dr,dr)(\nabla_r s|\nabla_r s)$$
$$+\sum[G^*(d\overline{w}^a,d\overline{w}^b)(\nabla_{\bar a}s|\nabla_{\bar b}s)$$
$$+G^*(dw^a,dw^b)(\nabla_a s|\nabla_b s)].$$

On the other hand, $(d^{\nabla}s|d^{\nabla}s)$ is written out, from (4.12), as

$$(d^{\nabla}s|d^{\nabla}s)=(dr\otimes\nabla_r s|dr\otimes\nabla_r s)$$
$$+\left(\sum d\overline{w}^a\otimes\nabla_{\bar a}s|\sum d\overline{w}^b\otimes\nabla_{\bar b}s\right)+\cdots. \qquad (4.15)$$

Accordingly, definition (4.13) implies that

$$(d\overline{w}^a\otimes\nabla_{\bar a}s|d\overline{w}^b\otimes\nabla_{\bar b}s)$$
$$=G^*(d\overline{w}^a,d\overline{w}^b)(\nabla_{\bar a}s|\nabla_{\bar b}s),\ \text{etc.}, \qquad (4.16)$$

which serve as a local definition of the inner product for cross sections in $T^*(M)^{\mathbb{C}}\otimes L_m$.

Using (4.14), we obtain the kinetic energy functional for $q_m^{\#}s$ in the form

$$\int_Q K_x^*(\overline{dq_m^{\#}s},dq_m^{\#}s)\,dQ$$
$$=\mu\int_{S^1}\int_M\left(\frac{m^2}{r^2}(s|s)+(d^{\nabla}s|d^{\nabla}s)\right)$$
$$\times\omega\wedge r^{2n-3}\,dr\wedge dS$$
$$=\int_M\left(\frac{m^2}{r^2}(s|s)+(d^{\nabla}s|d^{\nabla}s)\right)dM. \qquad (4.17)$$

Here we have used (3.21), (3.29), and the fact that kinetic energy density is constant on each fiber $\pi^{-1}(p)$ over $p \in M$.

The last step toward a global definition of the internal Hamiltonian is to define a covariant codifferential δ^{∇}, which is dual to d^{∇} and maps the space of cross sections in $T^*(M)^{\mathbb{C}} \otimes L_m$ to the space of cross sections in L_m. The definition depends on the Riemannian metric B and the volume element dM on M. Let s and u be cross sections in L_m and in $T^*(M)^{\mathbb{C}} \otimes L_m$, respectively. Then the covariant codifferential operator δ^{∇} is defined through

$$\int_M (d^{\nabla} s | u)\, dM = \int_M (s | \delta^{\nabla} u)\, dM . \tag{4.18}$$

Using δ^{∇} and d^{∇}, one has the Laplacian operator[7] Δ_m for sections in L_m:

$$\Delta_m = -\delta^{\nabla} \circ d^{\nabla} . \tag{4.19}$$

(The minus sign is for our convenience. In the literature, the Laplace operator is defined to be $\Delta = \delta^{\nabla} \circ d^{\nabla} + d^{\nabla} \circ \delta^{\nabla}$ for bundle-valued p-forms.)

We are in a final position to derive H_m. From definition (4.7), one obtains

$$\begin{aligned}
&\int_Q \overline{q_m^{\#} s}\, H(q_m^{\#} s)\, dQ \\
&= \int_Q \overline{q_m^{\#} s} \left(-\frac{1}{2} \Delta + V \right) (q_m^{\#} s)\, dQ \\
&= \int_Q \left[\frac{1}{2} K_x^{*} \left(\overline{dq_m^{\#} s}, dq_m^{\#} s \right) + \overline{q_m^{\#} s}\, V q_m^{\#} s \right] dQ \\
&= \int_M \left[\frac{1}{2} (d^{\nabla} s | d^{\nabla} s) + \frac{m^2}{2r^2} (s|s) + (s|Vs) \right] dM \\
&= \int_M \left(s \left| \left(-\frac{1}{2} \Delta_m + \frac{m^2}{2r^2} + V \right) s \right. \right) dM .
\end{aligned} \tag{4.20}$$

Thus we have the following.

Theorem 4.2: The internal Hamiltonian operator H_m, acting on the cross sections in the complex line bundle L_m, is expressed as

$$H_m = -\tfrac{1}{2} \Delta_m + m^2/2r^2 + V , \tag{4.21}$$

where Δ_m is the Laplacian operator defined on L_m and given by (4.19).

We remark in conclusion that according to the product structure $M \cong \mathbb{R}^+ \times \mathbb{C}P^{n-2}$ and Proposition 4.1, H_m can be expressed as

$$H_m = -\frac{1}{2} \frac{1}{r^{2n-3}} \frac{\partial}{\partial r} \left(r^{2n-3} \frac{\partial}{\partial r} \right) - \frac{1}{2r^2} \Lambda_m + \frac{m^2}{2r^2} + V , \tag{4.22}$$

where Λ_m is the Laplacian operator defined on the restriction of L_m to the submanifold $r = 1$. This Λ_m is called the Bochner–Laplacian operator, and is studied by Kuwabara.[13]

V. A COMPLETE ORTHONORMAL BASIS

In this section we discuss how one can pick up ρ_m-equivariant functions out of a complete orthonormal system in $L^2(Q)$. To carry out this purpose, we make use of boson calculus for the harmonic oscillator.

A. Picking up ρ_m-equivariant functions

Suppose we have the harmonic oscillator potential

$$V = \frac{1}{2} \sum_{j=1}^{n-1} |z^j|^2 = \frac{1}{2} \sum_{k=1}^{n} m_k (x_k | x_k) . \tag{5.1}$$

Though this potential is not realistic for the molecule, it helps in explaining the complete orthonormal bases in $L^2(Q)$. Let a_k and $a_k^{\dagger}$ denote the annihilation and creation operators, respectively, defined by

$$\begin{aligned}
a_k &= 2^{-1/2} \left(q^k + \frac{\partial}{\partial q^k} \right), \\
a_k^{\dagger} &= 2^{-1/2} \left(q^k - \frac{\partial}{\partial q^k} \right),
\end{aligned} \qquad k = 1,\ldots,2(n-1) . \tag{5.2}$$

Then, as is well-known, the functions

$$[k_1! \cdots k_{2(n-1)}!]^{-1/2} (a_1^{\dagger})^{k_1} \cdots (a_{2(n-1)}^{\dagger})^{k_{2(n-1)}} |0\rangle \tag{5.3}$$

form a complete orthonormal system in $L^2(Q)$, where $|0\rangle$ is the normalized ground state.

In keeping with (2.13), we introduce the annihilation and creation operators

$$\begin{aligned}
A_j &= 2^{-1/2} (a_{2j-1} + i a_{2j}) , \\
B_j &= 2^{-1/2} (a_{2j-1} - i a_{2j}) , \\
A_j^{\dagger} &= 2^{-1/2} (a_{2j-1}^{\dagger} - i a_{2j}^{\dagger}) , \\
B_j^{\dagger} &= 2^{-1/2} (a_{2j-1}^{\dagger} + i a_{2j}^{\dagger}) , \\
&\quad (j = 1,\ldots,n-1) ,
\end{aligned} \tag{5.4}$$

which satisfy the commutation relations

$$\begin{aligned}
&[A_j, A_k^{\dagger}] = \delta_{jk} , \quad [A_j, B_k^{\dagger}] = 0 , \\
&[B_j, B_k^{\dagger}] = \delta_{jk} , \quad [B_j, A_k^{\dagger}] = 0 ,
\end{aligned} \tag{5.5}$$

and the others all vanishing. Since the transformation (5.4) is canonical, the complete orthonormal system (5.3) may be written in the form

$$N_{kl} (A^{\dagger})^k (B^{\dagger})^l |0\rangle , \tag{5.6}$$

where we have used the following abbreviations:

$$\begin{aligned}
N_{kl} &= [k_1! \cdots k_{n-1}! l_1! \cdots l_{n-1}!]^{-1/2} , \\
(A^{\dagger})^k &= (A_1^{\dagger})^{k_1} \cdots (A_{n-1}^{\dagger})^{k_{n-1}} , \\
(B^{\dagger})^l &= (B_1^{\dagger})^{l_1} \cdots (B_{n-1}^{\dagger})^{l_{n-1}} .
\end{aligned} \tag{5.7}$$

We proceed to picking up ρ_m-equivariant functions out of system (5.6). The U(1) action on Q defines a unitary operator U_t in $L^2(Q)$ as usual:

$$(U_t f)(z, \bar{z}) = f(e^{-it} z, e^{it} \bar{z}) . \tag{5.8}$$

Thus, from (3.4), a function f in $L^2(Q)$ is ρ_m-equivariant if and only if

$$U_t f = e^{-imt} f . \tag{5.9}$$

Differentiating Eq. (5.8) with respect to t, we have the generator $-i\hat{F}$ of U_t, where $\hat{F}$ is the total angular momentum operator. Note here that the $\hat{F}$ was defined as a differential operator, but we have used the same letter for the self-adjoint extension of $\hat{F}$. Thus one has

$$U_t = \exp(-it\hat{F}) . \tag{5.10}$$

The total angular momentum operator can be expressed in terms of the operators (5.4) in the form

$$\hat{F}=\sum_{j=1}^{n-1}(-A_j A_j^{\dagger}+B_j B_j^{\dagger}) . \tag{5.11}$$

To verify this is a matter of calculation. Further, by this expression, one obtains

$$[\hat{F},A_k^{\dagger}]=-A_k^{\dagger}, \quad [\hat{F},B_k^{\dagger}]=B_k^{\dagger} .$$

Using these equations together with (5.10), we can show that the unitary operator U_t induces the transformations of $A_k^{\dagger}$ and $B_k^{\dagger}$ in the form

$$\begin{aligned} &U_t A_k^{\dagger} U_t^{-1}=e^{it}A_k^{\dagger}, \\ &U_t B_k^{\dagger} U_t^{-1}=e^{-it}B_k^{\dagger}, \end{aligned} \quad k=1,\ldots,n-1 . \tag{5.12}$$

It is easy to check that both sides of (5.12) satisfy the same differential equations in t. The domain of $A_k^{\dagger}$ and $B_k^{\dagger}$ are supposed, for example, to be the space of such functions as the products of $|0\rangle$ and polynomials in z and $\bar{z}$.

We are now in a position to pick up ρ_m-equivariant functions.

Theorem 5.1: Let

$$E_m=\operatorname{span}\{N_{kl}(A^{\dagger})^k(B^{\dagger})^l|0\rangle ;\ |k|-|l|=-m\}, \tag{5.13}$$

with $|k|=k_1+\cdots+k_{n-1}$, $|l|=l_1+\cdots+l_{n-1}$, being the closed linear subspace of the ρ_m-equivariant functions on Q. Then the Hilbert space $L^2(Q)$ is broken up into the orthogonal direct sum

$$L^2(Q)=\bigoplus_{m=-\infty}^{\infty} E_m . \tag{5.14}$$

Proof: Equation (5.12) and the fact that $U_t|0\rangle=|0\rangle$ imply that

$$U_t(A^{\dagger})^k(B^{\dagger})^l|0\rangle=e^{-it(-|k|+|l|)}(A^{\dagger})^k(B^{\dagger})^l|0\rangle . \tag{5.15}$$

Therefore, from (5.9), one verifies that $(A^{\dagger})^k(B^{\dagger})^l|0\rangle$ is the ρ_m-equivariant if and only if $-|k|+|l|=m$. As the system (5.6) forms a complete orthogonal system, we get Eq. (5.14), as desired.

Remark: In view of (3.28), we observe that E_m is isomorphic with the space of square integrable cross sections in L_m, which is denoted by $q_m^{\#-1}E_m$.

B. Unitary group actions

The planar n-body harmonic oscillator was treated in Ref. 14, using Lie algebraic methods. In this section we wish to study the harmonic oscillator in order to show that $q_m^{\#-1}E_m$ carries unitary representations of $U(n-1)\times U(n-1)$ and of $U(n-1)$, on the analogy of harmonic polynomials in z and $\bar{z}$. We notice first that the harmonic oscillator Hamiltonian is expressed as

$$H=\sum_{j=1}^{n-1}(A_j^{\dagger}A_j+B_j^{\dagger}B_j)+n-1 . \tag{5.16}$$

The eigenspaces of H are clearly spanned by functions (5.6) under conditions $|k|+|l|=N$, where N is non-negative integers. As is well known, each of the eigenspaces is the representation space of a unitary irreducible representation of $U(2(n-1))$.

We restrict the group $U(2(n-1))$ to the subgroups $U(n-1)\times U(n-1)$ and $U(n-1)$. For non-negative integers p and q, set

$$E_{p,q}=\operatorname{span}\{(A^{\dagger})^k(B^{\dagger})^l|0\rangle ;\ |k|=p,\ |l|=q\} . \tag{5.17}$$

The subspace E_m we have discussed is the direct sum of $E_{p,q}$ with $p-q=m$.

Theorem 5.2: The product group $U(n-1)\times U(n-1)$ has a unitary irreducible representation in the space $E_{p,q}$.

Proof: For $(g,h)\in U(n-1)\times U(N-1)$, we define the actions $U_1(g)$ and $U_2(h)$ as follows:

$$\begin{aligned} &U_1(g)A_k U_1(g^{-1})=\sum \overline{g_{jk}}\,A_j , \\ &U_1(g)A_k^{\dagger} U_1(g^{-1})=\sum g_{jk}A_j^{\dagger} , \\ &U_2(h)B_k U_2(h^{-1})=\sum \overline{h_{jk}}\,B_j , \\ &U_2(h)B_k^{\dagger} U_2(h^{-1})=\sum h_{jk}B_j^{\dagger} , \end{aligned} \tag{5.18}$$

where $g=(g_{jk})$ and $h=(h_{jk})$. The tensor product $U_1\otimes U_2$ is then defined to act on $E_{p,q}$ in the form

$$\left(\sum g_{j_1 1}A_{j_1}^{\dagger}\right)^{k_1}\cdots\left(\sum h_{j_1 1}B_{j_1}^{\dagger}\right)^{l_1}\cdots|0\rangle .$$

As is easily verified, the transformations (5.18) are canonical in the sense that the commutation relations are left invariant, so that the norm of $(A^{\dagger})^k(B^{\dagger})^l|0\rangle$ is preserved, and hence $U_1\otimes U_2$ becomes unitary on $E_{p,q}$. Further, since U_1 and U_2 are viewed as unitary irreducible representations[15] in the space of polynomials in $A_j^{\dagger}$ and in $B_j^{\dagger}$, respectively, the tensor product representation $U_1\otimes U_2$ is also irreducible.[16] This ends the proof.

We now proceed to representations of $U(n-1)$. If we replace $\bar{g}$ for h in Eq. (5.18), then we have a Kronecker product representation of $U(n-1)$ in $E_{p,q}$, which is not irreducible. In order to find out invariant subspaces, we consider following two operators invariant under the $U(n-1)$ action,

$$AB=\sum_{j=1}^{n-1}A_jB_j , \quad A^{\dagger}B^{\dagger}=\sum_{j=1}^{n-1}A_j^{\dagger}B_j^{\dagger} . \tag{5.19}$$

Let

$$G_{p,q}=\{f\in E_{p,q} ;\ (AB)f=0\} , \tag{5.20}$$

the kernel of AB: $E_{p,q}\to E_{p-1,q-1}$. Since AB is invariant under the action of $U(n-1)$, $G_{p,q}$ is an invariant subspace of $E_{p,q}$. The $G_{p,q}$ is an analog to the space of harmonic polynomials in z and $\bar{z}$. The following is a key formula to dividing the space $E_{p,q}$ into $U(n-1)$-invariant subspaces.

Proposition 5.3: For integers m, p, and q satisfying $1\leqslant m\leqslant p,q$, and for a function f of $E_{p-m,q-m}$, one has

$$\begin{aligned} AB[(A^{\dagger}B^{\dagger})^m f] &= m(n-2+p+q-m)(A^{\dagger}B^{\dagger})^{m-1}f \\ &\quad +(A^{\dagger}B^{\dagger})^m(AB)f . \end{aligned} \tag{5.21}$$

Proof: We prove this proposition by mathematical induction in m. For $m = 1$ and $f \in E_{p-1,q-1}$, using the commutation relations and the fact that $A_j|0\rangle = B_j|0\rangle = 0$, we can show that

$$(A^\dagger A)f = (p-1)f, \quad (B^\dagger B)f = (q-1)f.$$

Using these equations, we obtain, after calculation,

$$AB((A^\dagger B^\dagger)f) = (A^\dagger B^\dagger)(AB)f + (n+p+q-3)f.$$

Thus the proposition holds for $m = 1$. Now, assuming that (5.21) holds for m, we prove it for $m+1$ with $m+1 \leqslant p,q$. For a function $f \in E_{p-(m+1),q-(m+1)}$, the function $f_1 = (A^\dagger B^\dagger)f$ belongs to $E_{p-m,q-m}$, so that the assumption of induction can be applied for the f_1. Thus one has

$$AB((A^\dagger B^\dagger)^m f_1) = m(n-2+p+q-m)(A^\dagger B^\dagger)^{m-1} f_1 + (A^\dagger B^\dagger)^m (AB) f_1 . \tag{5.22}$$

To dispose of the last term in the right-hand side, we can apply the formula (5.21) with $m = 1$. In fact, as f is looked upon as belonging to $E_{(p-m)-1,(q-m)-1}$, we can apply (5.21) with $m = 1$ and with p and q replaced by $p-m$ and $q-m$, respectively, so that we obtain

$$(AB)f_1 = AB((A^\dagger B^\dagger)f) + (n+p+q-2m-3)f + (A^\dagger B^\dagger)(AB)f. \tag{5.23}$$

Equations (5.22) and (5.23) are put together to yield the formula (5.21) for $m+1$. This completes the proof.

Theorem 5.4: The space $E_{p,q}$ is decomposed into the direct sum

$$G_{p,0} = E_{p,0}, \quad G_{0,q} = E_{0,q}, \qquad E_{p,q} = G_{p,q} \oplus (A^\dagger B^\dagger) E_{p-1,q-1} \quad (p,q \geqslant 1). \tag{5.24}$$

Proof: The first two equalities are clear. We start by proving

$$G_{p,q} \cap (A^\dagger B^\dagger) E_{p-1,q-1} = \{0\}. \tag{5.25}$$

What we have to do is to show that for any nonzero function f of $E_{p-1,q-1}$, $AB((A^\dagger B^\dagger)f)$ never vanishes. Let

$$(A^\dagger B^\dagger)f = (A^\dagger B^\dagger)^m f_0, \tag{5.26}$$

where m is a maximal number such that f_0 no longer has the factor $A^\dagger B^\dagger$. Noting that $f_0 \in E_{p-m,q-m}$, we apply the formula (5.21) to (5.26) to get

$$AB((A^\dagger B^\dagger)f) = m(n-2+p+q-m)(A^\dagger B^\dagger)^{m-1} f_0 + (A^\dagger B^\dagger)^m (AB) f_0 . \tag{5.27}$$

If we had $AB((A^\dagger B^\dagger)f) = 0$, then the right-hand side of (5.27) would vanish, so that f_0 would have the factor $A^\dagger B^\dagger$, contradicting the assumption. The last part of the above reasoning is trivial if $m = 1$. For $m > 1$, we can come to the same conclusion. In fact, by applying the formula (5.21) inductively, we can write out $(AB)^m (A^\dagger B^\dagger)f$ in a similar form to (5.27). This ends the proof of (5.25).

The decomposition (5.24) now becomes easy to prove. Indeed, the dimensions of the both sides of (5.24) are equal, because of (5.25) and of the fact that $G_{p,q}$ is the kernel of the linear map AB: $E_{p,q} \to E_{p-1,q-1}$. Thus we have proved the theorem.

A successive application of (5.24) results in the following.

Corollary 5.5: The space $E_{p,q}$ is broken up into

$$E_{p,q} = G_{p,q} \oplus (A^\dagger B^\dagger) G_{p-1,q-1} \oplus \cdots \oplus (A^\dagger B^\dagger)^r G_{p-r,q-r}, \tag{5.28}$$

where $r = \min(p,q)$.

Here we remark that the direct sum decomposition (5.24) has great resemblance to a decomposition of the space of polynomials in complex variables z and $\bar{z}$. Thinking of $G_{p,q}$ as the space of harmonic polynomials, one finds in Ref. 17 an analog to (5.24). The representation of $U(n-1)$ induced in $G_{p,q}$ is then irreducible and belongs to the representation of highest weight $(p,0,\ldots,0,-q)$. (See Refs. 18 and 19, for example.) The dimensionality of $G_{p,q}$ is computed by (5.24) to give

$$\dim G_{p,q} = \dim E_{p,q} - \dim E_{p-1,q-1} = \frac{(n-2+p+q)(p+n-3)!(q+n-3)!}{(n-2)((n-3)!)^2 p!q!}. \tag{5.29}$$

Further, the decomposition (5.28) reminds us of the Clebsch–Gordan series[15] for a Kronecker product representation. Thus the Kronecker product representation of $U(n-1)$ in $E_{p,q}$ is reduced to its irreducible components.

We now return to the harmonic oscillator to carry out our purpose of showing that the groups $U(n-1) \times U(n-1)$ and $U(n-1)$ act on the cross sections in L_m. The space of ρ_m-equivariant eigenstates of the harmonic oscillator is assigned by the conditions $p+q = N$ and $p-q = -m$, and hence by

$$p = (N-m)/2, \quad q = (N+m)/2 \quad (N \geqslant |m|). \tag{5.30}$$

Thus from Theorem 5.2 and Corollary 5.5, we have the following.

Theorem 5.6: The product group $U(n-1) \times U(n-1)$ has a unitary irreducible representation in the space $q_m^{\#-1} E_{p,q}$ of cross sections in the complex line bundle L_m, where p and q are assigned by (5.30).

Theorem 5.7: The unitary representation of $U(n-1)$ in the space $q_m^{\#-1} E_{p,q}$ of cross sections in L_m, which is induced from that in the above theorem, is reducible to its irreducible components according to the decomposition (5.28), where p and q are assigned by (5.30).

[1] B. T. Sutcliffe, in *Quantum Dynamics of Molecules*, edited by R. G. Wooley (Plenum, New York, 1980).
[2] A. Guichardet, Ann. Inst. H. Poincaré **40**, 329 (1984).
[3] T. Iwai, J. Math. Phys. **28**, 964 (1987).
[4] T. Iwai, J. Math. Phys. **28**, 1315 (1987).
[5] G. Rosensteel, J. Math. Phys. **21**, 924 (1979); D. J. Rowe and G. Rosensteel, Ann. Phys. (NY) **126**, 198 (1980).
[6] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963, 1969), Vols. I and II.
[7] R. O. Wells, Jr., *Differential Analysis on Complex Manifolds* (Springer, New York, 1980).

[8]S. S. Chern, *Complex Manifolds without Potential Theory* (Springer, New York, 1979), 2nd ed.
[9]A. Trautman, Int. J. Theor. Phys. **16**, 561 (1977).
[10]S. Goldberg, *Curvature and Homology* (Dover, New York, 1982).
[11]Y. Matsushima, *Differentiable Manifolds* (Dekker, New York, 1972).
[12]N. P. Warner, Proc. R. Soc. London Ser. A **383**, 207, 217 (1982).
[13]R. Kuwabara, "Spectrum of the Schrödinger operator on the line bundles over $\mathbb{C}P^n$," to appear in Tôhoku Math. J.
[14]E. Chacón, M. Moshinsky, and V. Vanagas, J. Math. Phys. **22**, 605 (1981); E. Chacón, P. Hess, and M. Moshinsky, *ibid.* **25**, 1565 (1984).
[15]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[16]T. Brocker and T. T. Dieck, *Representations of Compact Lie Groups* (Springer, New York, 1985).
[17]M. Berger, P. Gauduchon, and E. Mazet, "Le Spectre d'une Variété Riemannienne," in *Lecture Notes in Mathematics*, Vol. 194 (Springer, Berlin, 1971).
[18]M. Ikeda, Mem. Fac. Eng. Hiroshima Univ. **3**, 17 (1967).
[19]M. Ikeda and N. Seto, Math. Jpn. **13**, 149 (1968).

When is a Hamiltonian system separable?

Ian Marshall
Mathematics Department, University of Manchester Institute of Science and Technology, P. O. Box 88, Manchester M60 1QD, United Kingdom

Stefan Wojciechowski
Mathematics Department, University of Linköping, S-581 83 Linköping, Sweden

(Received 3 November 1987; accepted for publication 27 January 1988)

The Hamiltonian system given by $H = \frac{1}{2}p^2 + V(q)$ with $V \in C^\infty(\mathbb{R}^n)$ is considered. A method for integrating such a system is that of separating the variables in the Hamilton–Jacobi equation. It is known that if such a separation is possible, then it can take place only when the equation is expressed in terms of generalized elliptic coordinates or in a degeneration of these. A criterion is proposed for deciding if separation is possible, and if it is, in which degeneration of elliptic coordinates it takes place.

I. INTRODUCTION

We consider the problem of integrating a Hamiltonian system by separating the variables in the Hamilton–Jacobi equation.

Proving integrability by separating the variables in—and so solving—the Hamilton–Jacobi equation has the advantage (over just the application of Liouville's theorem) that it allows one to directly find a transformation to cyclic variables for the system, and so to solve the equations of motion explicitly. The so-called property of separability (we speak of a "separable Hamiltonian system" or of a "separable potential") was the principal object of inquiry in the late nineteenth and early twentieth centuries for the study of Hamiltonian systems. A review of some important results of the period can be found in Ref. 1. Recently, more results on separability have been found; see Ref. 2–5. The last word, it seems, remains to be said.

Our work can be seen as the conclusion of a program initiated by Stäckel[6] in 1891, that of classifying Hamiltonian systems according to their separability or nonseparability. It is often thought that the question posed in the title of this paper is answered by Stäckel's theorem; in fact this is not the case. Stäckel's theorem gives us no indication of what the variables for separation are, if they exist, or even if indeed such variables do exist. It only tells us if a *given* coordinate system has this property with respect to whichever Hamiltonian system we are considering. A natural Hamiltonian is typically expressed in Cartesian coordinates; our criterion is implemented in Cartesian coordinates.

We present the means for the classification by separability of "natural" Hamiltonian systems (a natural Hamiltonian is one having the form $H = T + V$, where $T = \frac{1}{2}\Sigma_{k=1}^{n} p_k^2$ and V is a function of the position variables q_i only), which describe the motion of particles in a space of zero curvature. In other words, one can take any natural Hamiltonian system and subject it to certain tests enabling one to answer the question, *is the system separable?* And if so, *in which coordinates is it separable?*

For a system with two degrees of freedom the answer to this question was provided by Whittaker in his book.[7] From his account of results of Bertrand and Darboux we are led to formulate a theorem which we call the Bertrand–Darboux theorem. Essentially this tells us that a necessary and sufficient condition for a given natural Hamiltonian $H = \frac{1}{2}(p_x^2 + p_y^2) + V(x, y)$ to be separable in elliptic, polar, parabolic, or Cartesian coordinates is that the expression

$$\begin{aligned}(V_{yy} - V_{xx})(-2axy - b'y - bx + c_1) \\ + 2V_{xy}(ay^2 - ax^2 + by - b'x + c') \\ + V_x(6ay + 3b) + V_y(-6ax - 3b')\end{aligned} \tag{1.1}$$

vanishes for some constants $(a,b,b',c,c',c_1) \neq (0,0,0,c,c,0)$. In addition, this theorem tells us in which particular coordinate system [the characteristic coordinates of (1.1)] the Hamilton–Jacobi equation for H separates, depending on the values of the constants (a,b,b',c,c',c_1).

The second integrable case of the Henon–Heiles system was shown to be separable using this criterion.[8]

As an illustration of the use of this theorem as an effective criterion of separability, let us prove that the three-dimensional nonperiodic Toda lattice Hamiltonian is nonseparable:

$$H = \tfrac{1}{2}(p_1^2 + p_2^2 + p_3^2) + e^{2(q_1 - q_2)} + e^{2(q_2 - q_3)}. \tag{1.2}$$

If we set $x = 1/\sqrt{2}(q_1 - q_2)$, $y = 1/\sqrt{6}(q_1 + q_2 - 2q_3)$, and $z = 1/\sqrt{3}(q_1 + q_2 + q_3)$, we have

$$H = \tfrac{1}{2}(p_x^2 + p_y^2 + p_z^2) + e^{2\sqrt{2}x} - e^{\sqrt{6}y - \sqrt{2}x}, \tag{1.3}$$

and the problem of the separability of the Hamilton–Jacobi equation is reduced to the problem with two degrees of freedom, with

$$\tilde{H}(x, y, p_x, p_y) = \tfrac{1}{2}(p_x^2 + p_y^2) + V(x, y)$$

and

$$V(x, y) = e^{2\sqrt{2}x} - e^{\sqrt{6}y - \sqrt{2}x}. \tag{1.4}$$

Now if we form the expression given in (1.1) for V given by (1.4), we find that in order to make the expression vanish we are forced to choose $(a,b,b',c,c',c_1) = (0,0,0,c,c,0)$. Hence, applying the theorem, the Hamilton–Jacobi equation for $\tilde{H}$ given by (1.4) is nonseparable; this in turn implies the nonseparability of (1.2).

We will formulate a similar condition for a system with arbitrary degrees of freedom.

 0022-2488/88/061338-09$02.50

Iarov-Iarovoi[3] proved, using the zero curvature condition, that if a given natural Hamiltonian system is separable, then it is separable in generalized elliptic coordinates or in some degeneration of elliptic coordinates—for example, in parabolic coordinates.[9] Because of the large number of such degenerations for $n \geqslant 3$, we are unable to provide a straightforward formulation of the criterion of separability in general. We adopt instead the approach of treating elliptic coordinates (nondegenerate case), followed by a discussion of the effect of degenerations.

In the case $n = 2$ the Bertrand–Darboux theorem does constitute the most general criterion of separability, in that it accounts for all degenerations.

For $n = 3$ there are already 11 degenerations, instead of four for $n = 2$ (see Refs. 2 and 10). We are preparing a paper that gives the form of the criterion in each case, and we expect this to largely supersede Eisenhart[10] in its application directly to the solving of problems.

In Sec. II we recall the $n = 2$ case. We then go on to consider separability for arbitrary n. In Sec. III we confine ourselves to the case of separability in generalized elliptic coordinates, and then in Sec. IV we look at what happens in the degenerate cases. Finally, the Appendix contains calculations which, either because they are already available in the literature, or because they are not central to the main arguments, have been relegated to the end.

II. THE BERTRAND–DARBOUX THEOREM

For the Hamiltonian function in $\mathbf{R}^2 \times \mathbf{R}^2$ $H = \frac{1}{2}(p_x^2 + p_y^2) + V(x, y)$, where (p_x, p_y) is the momentum conjugate to (x, y), for some constants $(a,b,b',c,c',c_1) \neq (0,0,0,c,c,0)$, the following conditions are equivalent.

(1) H has an independent integral of the form

$$K = (ay^2 + by + c)p_x^2 + (ax^2 + b'x + c')p_y^2 + (-2axy - b'y - bx + c_1)p_x p_y + k(x, y),$$

where k is some differentiable function: $\mathbf{R} \times \mathbf{R} \to \mathbf{R}$.

(2) V satisfies the differential equation

$$(V_{yy} - V_{xx})(-2axy - b'y - bx + c_1) + 2V_{xy}(ay^2 - ax^2 + by - b'x + c - c') + V_x(6ay + 3b) + V_y(-6ax - 3b') = 0,$$

where subscripts on V mean "differentiate," e.g., $V_x = \partial V/\partial x$.

(3) The system is separable in one of the following orthogonal coordinate systems in the plane:

Cartesian, polar, parabolic, elliptic.

A complete account of the content of this theorem can be found in Ref. 7. See also Ref. 8 or Ref. 11 for the parabolic case.

III. SEPARABILITY IN GENERALIZED ELLIPTIC COORDINATES

The Bertrand–Darboux theorem has an extension to higher dimensions establishing a condition for the separability, in each degeneration of elliptic coordinates, of the Hamilton–Jacobi equation for a natural Hamitonian system. The respective degenerations prove too messy to be incorporated explicitly, and all at once, into a single expression; so we prefer to formulate the theorem as it applies just to separability in undegenerated elliptic coordinates, and to follow this by a breakdown of all the possibilities for the (slightly) different forms the theorem takes in the degenerate cases.

Theorem: Given a natural Hamiltonian system with

$$H = \frac{1}{2}\sum_{k=1}^{n} p_k^2 + V(q), \tag{3.1}$$

the following three conditions are equivalent.

(a) H has n global, independent, involutive integrals of motion having, in some Cartesian frame, the form

$$K_i = \sum_{j \neq i} \frac{l_{ij}^2}{\alpha_i - \alpha_j} + p_i^2 + k_i(q), \quad i = 1,...,n, \tag{3.2}$$

where $\alpha_1,...,\alpha_n$ are n distinct constants. What is important is the differences $\alpha_i - \alpha_j$; we can suppose that $0 = \alpha_1 < \alpha_2 < \cdots < \alpha_n$. $l_{ij} = q_i p_j - q_j p_i$ and k_i are functions of q only.

(b) The potential V satisfies the set of $\frac{1}{2}n(n-1)^2$ equations,

$$(\alpha_i - \alpha_r)^{-1}(q_i^2 V_{rs} - q_i q_r V_{is}) = (\alpha_i - \alpha_s)^{-1}(q_i^2 V_{rs} - q_i q_s V_{ir}), \quad i,r,s \text{ all different},$$

$$(\alpha_i - \alpha_r)^{-1} q_i q_r (V_{ii} - V_{rr}) - \sum_{j \neq i,r} (\alpha_i - \alpha_j)^{-1} q_i q_j V_{jr} + V_{ir}\left[\sum_{j \neq i,r} (\alpha_i - \alpha_j)^{-1} q_j^2 + (\alpha_r - \alpha_i)^{-1}(q_i^2 - q_r^2)\right] + V_{ir} + 3(\alpha_i - \alpha_r)^{-1}(q_r V_i - q_i V_r) = 0, \quad i \neq r. \tag{3.3}$$

Subscripts on V mean "differentiate," e.g., $V_i = \partial V/\partial q_i$.

(c) The Hamilton–Jacobi equation for H, $H(q,p)|_{p = \partial S/\partial q} = E$, is separable in generalized elliptic coordinates $(u_1,...,u_n)$ given by

$$1 + \sum_{k=1}^{n} \frac{q_k^2}{z - \alpha_k} = \prod_{j=1}^{n} (z - u_j) \times \left(\prod_{k=1}^{n} (z - \alpha_k)\right)^{-1} = \frac{U(z)}{A(z)}. \tag{3.4}$$

Remark: For $n = 2$ this theorem is the same as the Bertrand–Darboux theorem with $\alpha = \frac{1}{2}(\alpha_1 - \alpha_2)^{-1}$, $b = b' = c_1 = 0$, $c - c' = 1$.

Proof: The condition $\{H, K_i\} = 0$ yields

$$\sum_{j=1}^{n} k_{i,j} p_j = -2\sum_{j \neq i}^{n} (\alpha_i - \alpha_j)^{-1} \times (l_{ij}(q_i V_j - q_j V_i)) - 2p_i V_i.$$

We construct $\Sigma_{i,j=1}^{n} (z - \alpha_i)^{-1} k_{i,j} p_j$ and take the coefficient of p_r to get

$$\sum_{i=1}^{n} (z - \alpha_i)^{-1} k_{i,r} = -2(z - \alpha_r)^{-1} V_r - 2\sum_{i \neq r}^{n} (z - \alpha_i)^{-1}(z - \alpha_r)^{-1} \times (q_i^2 V_r - q_i q_r V_i), \quad \forall r. \tag{3.5}$$

By differentiating (3.5) we have an expression for $\Sigma_{i=1}^{n} (z - \alpha_i)^{-1} k_{i,rs}$. Setting

$$\sum_{i=1}^{n} (z - \alpha_i)^{-1}(k_{i,rs} - k_{i,sr}) = 0,$$

as of course it should, we obtain

$$\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_i^2((z-\alpha_r)^{-1}-(z-\alpha_s)^{-1})V_{rs}+(z-\alpha_s)^{-1}q_s\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_i\frac{\partial}{\partial q_i}V_r$$
$$-(z-\alpha_r)^{-1}q_r\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_i\frac{\partial}{\partial q_i}V_s+(z-\alpha_r)^{-1}V_{rs}-(z-\alpha_s)^{-1}V_{rs}$$
$$+3(z-\alpha_r)^{-1}(z-\alpha_s)^{-1}(q_sV_r-q_rV_s)=0,\quad \forall z,\quad r,s=1,\ldots,n. \tag{3.6}$$

[Note that from (3.5) we can obtain $k_{i,r}$, and this enables us to check that $\{K_i,K_j\}=0$, $\forall_{ij}$.]

The residues of (3.6) at $z=\alpha_i$, for i,r,s all different and for $i=s\neq r$, respectively, give the equations in (b). That (3.6) can itself be recovered from (b) is also easy to show. Thus (a) $\Rightarrow$ (3.6) $\Leftrightarrow$ (b).

We now show that (3.6) $\Rightarrow$ (c), and hence (b) $\Rightarrow$ (c). We make use of the following identities:

(i) $$\frac{\partial}{\partial q_i}=2\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}\frac{q_i}{\alpha_i-u_k}\frac{\partial}{\partial u_k},$$

(ii) $$\sum_{i=1}(z-\alpha_i)q_i\frac{\partial}{\partial q_i}=2\frac{U(z)}{A(z)}\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}(z-u_k)^{-1}\frac{\partial}{\partial u_k},$$

(iii) $$(z-\alpha_r)^{-1}q_r\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_i\frac{\partial}{\partial q_i}=(z-\alpha_r)^{-1}\frac{U(z)}{A(z)}\frac{\partial}{\partial q_r}$$
$$-2\frac{U(z)}{A(z)}\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}(z-u_k)^{-1}(\alpha_r-u_k)^{-1}\frac{\partial}{\partial u_k},$$

(iv) $$\frac{\partial q_i}{\partial u_k}=\frac{1}{2}\frac{q_i}{u_k-\alpha_i}.$$

To get (i), differentiate the equation

$$1+\sum_{k=1}^{n}\frac{q_k^2}{z-\alpha_k}=\frac{U(z)}{A(z)}$$

with respect to q_k:

$$\Rightarrow\frac{2q_k}{z-\alpha_k}=-\frac{1}{A(z)}U(z)\sum_{j=1}^{n}(z-u_j)^{-1}\frac{\partial u_j}{\partial q_k}.$$

Now keeping u and q fixed, let $z\to u_j$:

$$\Rightarrow\frac{2q_k}{u_j-\alpha_k}=-\frac{U'(u_j)}{A(u_j)}\frac{\partial u_j}{\partial q_k}.$$

Hence

$$\frac{\partial u_j}{\partial q_k}=\frac{2q_k}{\alpha_k-u_j}\frac{A(u_j)}{U'(u_j)},$$

and (i) follows. Identity (ii) comes from (i) together with the equations

$$(z-\alpha_i)^{-1}(\alpha_i-u_k)^{-1}=(z-u_k)^{-1}((z-\alpha_i)^{-1}+(\alpha_i-u_k)^{-1})$$

and

$$\sum_{i=1}^{n}\frac{q_i^2}{\alpha_i-u_k}=1,\quad \forall k.$$

Identity (iii) follows easily from (ii).

At length we obtain that (3.6) is equivalent to

$$-2\frac{U(z)}{A(z)}q_s\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}(\alpha_s-u_k)^{-1}(z-u_k)^{-1}\frac{\partial}{\partial u_k}V_r$$
$$+2\frac{U(z)}{A(z)}q_r\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}(\alpha_r-u_k)^{-1}(z-u_k)^{-1}$$
$$\times\frac{\partial}{\partial u_k}V_s+6(z-\alpha_r)^{-1}(z-\alpha_s)^{-1}$$
$$\times\left[q_s\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}\frac{q_r}{(\alpha_r-u_k)}\frac{\partial V}{\partial u_k}\right.$$
$$\left.-q_r\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}\frac{q_s}{(\alpha_s-u_k)}\frac{\partial V}{\partial u_k}\right]=0 \tag{3.7}$$

(where, remember, $V_r=\partial V/\partial q_r$). For $r\neq s$, keeping q and u fixed, let $z\to u_i$ in (3.7). Using the identities (i)–(iv) we get

$$\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}(\alpha_r-u_k)^{-1}(\alpha_s-u_k)^{-1}$$
$$\times\left[(u_i-u_k)\frac{\partial^2V}{\partial u_i\partial u_k}+\frac{\partial V}{\partial u_k}\right]=0.$$

Now make use of the fact [see (A10) in Appendix A] that

$$\sum_{k=1}^{n}\frac{P(u_k)}{U'(u_k)}=0$$

when $P(z)$ is any polynomial of degree less than $n-1$, and we get

$$\sum_{k=1}^{n}\frac{A(u_k)}{U'(u_k)}(\alpha_r-u_k)^{-1}(\alpha_s-u_k)^{-1}$$
$$\times\left[(u_i-u_k)\frac{\partial^2V}{\partial u_i\partial u_k}-\left(\frac{\partial V}{\partial u_i}-\frac{\partial V}{\partial u_k}\right)\right]=0. \tag{3.8}$$

We read (3.8) in this way: for fixed i and r it is an $(n-1)$-dimensional linear equation of the form $\mathscr{A}^{(r,i)}\mathbf{x}^{(i)}=0$, where $\mathscr{A}^{(r,i)}$ is the $(n-1)\times(n-1)$ matrix with components

$$\mathscr{A}^{(r,i)}{}_{sk} = \frac{A(u_k)}{U'(u_k)}(\alpha_r - u_k)^{-1}(\alpha_s - u_k)^{-1},$$

$$s \neq r, \quad k \neq i, \tag{3.9}$$

and $\mathbf{x}^{(i)}$ is the $(n-1)\times 1$ column vector,

$$x^{(i)}{}_k = \frac{\partial}{\partial u_i}\frac{\partial}{\partial u_k}[(u_i - u_k)V]. \tag{3.10}$$

Unless $\det \mathscr{A}^{(r,i)} \equiv 0$ this equation has the solution $\mathbf{x}^{(i)} = 0$ for all i. Now

$$\det \mathscr{A}^{(r,i)} = \frac{(\alpha_r - u_1)^{-1}}{U'(u_1)} \overset{i}{\vee} \cdots \frac{(\alpha_r - u_n)^{-1}}{U'(u_n)} \times \det\left[\frac{A(u_k)}{\alpha_s - u_k}\right]$$

($\overset{i}{\vee}$ means omit the ith term). As $u_k \to \alpha_k$ for $k \neq r$,

$$\det\left[\frac{A(u_k)}{\alpha_s - u_k}\right] \to \pm \frac{A(u_r)}{\alpha_i - u_r} A'(\alpha_1) \overset{i,r}{\vee} \cdots A'(\alpha_n) \not\equiv 0.$$

Hence $\det \mathscr{A}^{(r,i)} \not\equiv 0$, and therefore $x^{(i)}{}_k = 0$, $\forall i,k$. In other words,

$$\frac{\partial}{\partial u_i}\frac{\partial}{\partial u_k}[(u_i - u_k)V] = 0, \quad \forall i,k. \tag{3.11}$$

The general solution of (3.11) is

$$V = \sum_{j=1}^{n} \frac{f_j(u_j)}{U'(u_j)}, \tag{3.12}$$

and the Hamilton–Jacobi equation is separable (see Appendix A).

In elliptic coordinates the Hamilton–Jacobi equation takes the form

$$-2\sum_{j=1}^{n} \frac{A(u_j)}{U'(u_j)}\left(\frac{\partial S}{\partial u_j}\right)^2 + V = E, \tag{3.13}$$

and is separable whenever V has the form given by (3.12).

(c) $\Rightarrow$ (a): $V(q)$ separable in generalized elliptic coordinates $\Leftrightarrow V$ takes the form

$$V(q(u)) = \sum_{j=1}^{n} \frac{f_j(u_j)}{U'(u_j)}.$$

The Hamilton–Jacobi equation takes the form

$$\sum_{j=1}^{n} \frac{1}{U'(u_j)}\left[-4A(u_j)\left(\frac{\partial S}{\partial u_j}\right)^2 + 2f_j(u_j) - P(u_j)\right] = 0, \tag{3.14}$$

where $P(z) = \eta_1 z^{n-1} + \eta_2 z^{n-2} + \cdots + \eta_n$, $2E = \eta_1$, and can be simultaneously separated by posing

$$S = \sum_{j=1}^{n} S_j(u_j) \tag{3.15}$$

to give

$$-4A(u_j)\left(\frac{\partial S}{\partial u_j}\right)^2 + 2f_j(u_j) - P(u_j) = 0$$

or

$$-4A(u_j)v_j^2 + 2f_j(u_j) - P(u_j) = 0, \quad j = 1,\ldots,n. \tag{3.16}$$

The $\eta_1,\ldots,\eta_n$ are constants of integration that would be fixed by initial conditions in the complete solution. If we set

$$x_j = -4A(u_j)v_j^2 - u_j^n$$

and

$$y_j = 2f_j(u_j) + u_j^n,$$

we can write (3.16) as

$$\mathscr{U}\boldsymbol{\eta} = \mathbf{x} + \mathbf{y}, \tag{3.17}$$

with $\mathscr{U}$ the $n\times n$ matrix, $\mathscr{U}_{ij} = u_i^{n-j}$. Now $\det \mathscr{U}$ is the van der Monde determinant, $\Pi_{i<j}(u_i - u_j)$; this is nowhere identically zero, and hence (3.17) defines the constants $\eta_1,\ldots,\eta_n$ uniquely in terms of u and v, and so in terms of q and p.

We write

$$\boldsymbol{\eta}^x = \mathscr{U}^{-1}\mathbf{x} \text{ and } \boldsymbol{\eta}^y = \mathscr{U}^{-1}\mathbf{y} \text{ so that } \boldsymbol{\eta} = \boldsymbol{\eta}^x + \boldsymbol{\eta}^y. \tag{3.18}$$

We will show that (3.18) together with the mapping $(q_1^2,\ldots,q_n^2) \leftrightarrow (u_1,\ldots,u_n)$ implies

$$\sum_{k=1}^{n} z^{n-k}\eta^x{}_k(u(q),v(q,p)) = A(z)\left[1 + \sum_{i=1}^{n}\frac{R_i}{z-\alpha_i}\right] - z^n, \quad \forall z, \tag{3.19}$$

where

$$R_i = \sum_{j\neq i}(\alpha_i - \alpha_j)^{-1}l_{ij}^2 + p_i^2 + q_i^2.$$

A polynomial of degree ν is uniquely defined by its values at $\nu+1$ distinct points. It is enough then to check (3.19) for the points $z = u_1,\ldots,u_n$:

$$\begin{aligned}
1 + \sum_{i=1}^{n}\frac{R_i}{z-\alpha_i} &= \left(1 + \sum_{k=1}^{n}(z-\alpha_k)^{-1}q_k^2\right)\left(1 + \sum_{k=1}^{n}(z-\alpha_k)^{-1}p_k^2\right) \\
&\quad - \left(\sum_{k=1}^{n}(z-\alpha_k)^{-1}q_k p_k\right)^2 \\
&= \frac{U(z)}{A(z)}\left(1 + \sum_{k=1}^{n}(z-\alpha_k)^{-1}p_k^2\right) \\
&\quad - \left[\frac{1}{2}\frac{d}{dt}\left(\frac{U(z)}{A(z)}\right)\right]^2
\end{aligned}$$

as $p = \dot{q}$;

$$\therefore 1 + \sum_{i=1}^{n}\frac{R_i}{u_j - \alpha_i} = -\frac{1}{4}\frac{U'(u_j)^2}{A(u_j)^2}\dot{u}_j^2 = -4v_j^2. \tag{3.20}$$

Hence

$$\begin{aligned}
&(\text{rhs of } (3.19))|_{z=u_j} \\
&= \left[A(z)\left(1 + \sum_{i=1}^{n}\frac{R_i}{z-\alpha_i}\right) - z^n\right]\Bigg|_{z=u_j} \\
&= -4A(u_j)v_j^2 - u_j^n = x_j,
\end{aligned}$$

but from (3.18) this is just $\mathscr{U}\boldsymbol{\eta}^x$. In turn

$$\mathscr{U}\boldsymbol{\eta}^x = (\text{lhs of } (3.19))|_{z=u_j},$$

so (3.19) is proved.

We know that $\eta^x{}_k + \eta^y{}_k$ is a constant of motion for $k = 1,\ldots,n$, hence

$$\sum_{k=1}^{n} z^{n-k}(\eta^x{}_k + \eta^y{}_k)$$

is a constant of motion for all z. Taking the limit as $z \to \alpha_i$ and using (3.19) which we just proved, we obtain n constants of motion

$$K_i = R_i + k_i(q), \tag{3.21}$$

where

$$k_i = \lim_{z\to\alpha_i} \frac{z-\alpha_i}{A(z)}\left(\sum_{k=1}^{n} z^{n-k}\eta^y{}_k\right)$$

and is easily seen to be independent of p.

Moreover, as

$$\left\{\sum_{k=1}^{n} z^{n-k}\eta_k, \sum_{k=1}^{n} z_2^{n-k}\eta_k\right\}$$

is clearly zero for all choices of z_1 and z_2, we find $\{K_i,K_j\} = 0$, $\forall i,j$.

IV. DEGENERATIONS

We now discuss the possible degenerations of elliptic coordinates on $\mathbb{R}^n$. The systematic description in Ref. 2 is very useful here.

There are four types of degeneration:

$$1 + \sum_{k=1}^{n} \frac{q_k^2}{z-\alpha_k} = \frac{U(z)}{A(z)} \to$$

(i) $$\sum_{k=1}^{n} \frac{q_k^2}{z-\alpha_k} = q^2 \prod_{j=1}^{n-1} (z-w_j)/A(z)$$
$$= q^2 W(z)/A(z)$$
$$(q^2 = q_1^2 + \cdots + q_n^2);$$

(ii) $$\left[1 + \sum_{k=1}^{\nu} \frac{q_k^2}{z-\alpha_k} = \frac{U_1(z)}{A_1(z)},\right.$$
$$\left. 1 + \sum_{k=\nu+1}^{n} \frac{q_k^2}{z-\alpha_k} = \frac{U_2(z)}{A_2(z)}\right],$$

where

$$U_1(z) = \prod_{j=1}^{\nu} (z-u_j), \quad A_1(z) = \prod_{i=1}^{\nu} (z-\alpha_i),$$
$$U_2(z) = \prod_{j=\nu+1}^{n} (z-u_j), \quad A_2(z) = \prod_{i=\nu+1}^{n} (z-\alpha_i);$$

(iii) $$\sum_{k=1}^{n-1} \frac{\phi_k^2}{z-\alpha_k} + z - 2\phi_n = \frac{U(z)}{\mathring{A}(z)},$$
$$\mathring{A}(z) = \prod_{k=1}^{n-1} (z-\alpha_k)$$

(this is the case $\alpha_n \to \infty$);

(iv) $$1 + \sum_{k=3}^{n} \frac{q_k^2}{z-\alpha_k} + \frac{\rho^2}{z-\alpha_2} = \frac{\hat{U}(z)}{A(z)}$$

and

$$q_1 = \rho\cos\theta, \quad q_2 = \rho\sin\theta,$$
$$\hat{U}(z) = \prod_{j=2}^{n} (z-u_j)$$

(this is the case $\alpha_1 - \alpha_2 \to 0$).

The degenerations in (ii) and (iv) can take place as many times as we like—up to $(n-1)$ of course—the other two only once. Also (i) and (iii) cannot be combined with each other whereas both (ii) and (iv) can be combined in various ways with any of the other degenerations. We will examine (i)–(iv) in detail.

(i) Suppose we have

$$1 + \sum_{k=1}^{n} \frac{q_k^2}{z-\alpha_k} = \prod_{j=1}^{n} (z-u_j)$$
$$\times\left(\prod_{k=1}^{n} (z-\alpha_k) = \frac{U(z)}{A(z)}\right)^{-1}.$$

Then

$$\sum_{k=1}^{n} \frac{q_k^2}{z-\alpha_k} = \frac{U(z)-A(z)}{A(z)}.$$

$U(z) - A(z)$ is a polynomial of degree $n-1$, with leading term

$$z^{n-1}\sum_{k=1}^{n} (\alpha_k - u_k) = z^{n-1}\sum_{k=1}^{n} q_k^2,$$

as can be seen by multiplying both sides by $A(z)$.

Thus $U(z) - A(z) = q^2 W(z)$, where $W(z)$ is a polynomial of degree $n-1$ with leading term z^{n-1}. By the fundamental theorem of algebra $W(z)$ can be written

$$W(z) = (z-w_1)\cdots(z-w_{n-1}), \quad \text{for } w_i \in \mathbb{C}. \tag{4.1}$$

It is immediately apparent that in fact all w_i are real for $q_i,\alpha_i \in \mathbb{R}$; we have

$$\sum_{k=1}^{n} \frac{q_k^2}{w_i - \alpha_k} = 0, \quad \forall w_i.$$

Suppose $w_i = \xi + i\eta$ for some i with $\eta \neq 0$. Then we have

$$\sum_{k=1}^{n} ((\xi-\alpha_k)^2 + \eta^2)^{-1} q_k^2 (\xi - \alpha_k - i\eta) = 0. \tag{4.2}$$

The imaginary part of (4.2) is

$$\eta \sum_{k=1}^{n} ((\xi-\alpha_k)^2 + \eta^2)^{-1} q_k^2 = 0$$

—and we have a contradiction. We have then

$$\sum_{k=1}^{n} \frac{q_k^2}{z-\alpha_k} = q^2 \frac{W(z)}{A(z)}, \quad \text{with } w_i \in \mathbb{R}.$$

The theorem cannot be directly translated into the appropriate form for this case, but needs to be developed from the beginning. It is straightforward to do this. We leave it to Appendix B.

(ii) This case is obvious. We have essentially two noninteracting systems.

(iii) Suppose

$$1 + \sum_{k=1}^{n} \frac{q_k^2}{z-\alpha_k} = \prod_{j=1}^{n} (z-u_j)$$
$$\times\left(\prod_{k=1}^{n} (z-\alpha_k) = \frac{U(z)}{A(z)}\right)^{-1}.$$

Let $q_k = \phi_k/\sqrt{\alpha_n}$, $k \neq n$, and $q_n = (\phi_n + \alpha_n)/\sqrt{\alpha_n}$. Multiply both sides by $z - \alpha_n$ and let $\alpha_n \to \infty$. We obtain

$$\sum_{k=1}^{n}\frac{\phi_k^2}{z-\alpha_k}-z+2\phi_n=\frac{U(z)}{\mathring{A}(z)},\tag{4.3}$$

where $\mathring{A}(z)=(z-\alpha_1)\cdots(z-\alpha_{n-1})$.

In this case the theorem is not affected by taking the limit. That is,

$$\lim_{\alpha_n\to\infty}(\mathrm{a})\Rightarrow\lim_{\alpha_n\to\infty}(\mathrm{b})\Rightarrow\lim_{\alpha_n\to\infty}(\mathrm{c})\Rightarrow\lim_{\alpha_n\to\infty}(\mathrm{a})$$

when the limit is taken according to the above prescription. (This has been checked!) We only need to see what happens to (a), (b), and (c) in this limit. We get the following results.

(a) For $i\neq n$,

$$\begin{aligned}K_i&=\frac{\xi_i^2}{\alpha_n}+\frac{1}{\alpha_n^2}\sum_{j\neq i}^{n-1}\frac{(\phi_i\xi_j-\phi_j\xi_i)^2}{\alpha_i-\alpha_j}\\&\quad+\frac{1}{\alpha_n^2}\frac{(\phi_i\xi_n-\phi_n\xi_i-\alpha_n\xi_i)^2}{\alpha_i-\alpha_n}+k_i(q)\\&\quad(\text{where }\xi_i=p_i/\sqrt{\alpha_n},\quad i=1,\ldots,n)\\&=\frac{1}{\alpha_n^2}\sum_{j\neq i}^{n-1}\frac{(\phi_i\xi_j-\phi_j\xi_i)^2}{\alpha_i-\alpha_j}+\frac{1}{\alpha_n^2}\frac{(\phi_i\xi_n-\phi_n\xi_i)^2}{\alpha_i-\alpha_n}\\&\quad+\frac{1}{\alpha_i-\alpha_n}\left(\frac{1}{\alpha_n}(\alpha_i-\alpha_n)+1\right)\xi_i^2\\&\quad-\frac{2\xi_i(\phi_i\xi_n-\phi_n\xi_i)}{\alpha_n(\alpha_i-\alpha_n)}+k_i(q).\end{aligned}$$

Multiply by α_n^2 and let $\alpha_n\to\infty$. Writing

$$\tilde{k}_i(\phi)=\lim_{\alpha_n\to\infty}\alpha_n^2k_i\left(\frac{\phi_1}{\sqrt{\alpha_n}},\ldots,\frac{\phi_n+\alpha_n}{\sqrt{\alpha_n}}\right),$$

we get

$$\tilde{K}_i=\sum_{j\neq i}^{n-1}(\alpha_i-\alpha_j)^{-1}\lambda_{ij}^2+2\xi_i\lambda_{in}-\alpha_i\xi_i^2+\tilde{k}_i(\phi),\tag{4.4}$$

$$\begin{aligned}K_n&=\frac{\xi_n^2}{\alpha_n}+\frac{1}{\alpha_n^2}\sum_{j=1}^{n-1}\frac{(\phi_n\xi_j-\phi_j\xi_n+\alpha_n\xi_j)^2}{\alpha_n-\alpha_j}+k_n(q)\\&=\frac{\xi_n^2}{\alpha_n}+\frac{1}{\alpha_n^2}\sum_{j=1}^{n-1}\frac{\lambda_{nj}^2}{\alpha_n-\alpha_j}+\sum_{j=1}^{n-1}\frac{\xi_j^2}{\alpha_n-\alpha_j}\\&\quad+\frac{2}{\alpha_n}\sum_{j=1}^{n-1}\frac{\xi_j\lambda_{nj}}{\alpha_n-\alpha_j}+k_n(q).\end{aligned}$$

Multiply by α_n and let $\alpha_n\to\infty$:

$$\tilde{K}_n=\sum_{i=1}^{n}\xi_i^2+\tilde{k}_n(\phi),\tag{4.5}$$

where

$$\tilde{k}_n(\phi)=\lim_{\alpha_n\to\infty}\alpha_nk_n\left(\frac{\phi_1}{\sqrt{\alpha_n}},\ldots,\frac{\phi_n+\alpha_n}{\sqrt{\alpha_n}}\right).$$

$\tilde{K}_n$ must be the Hamiltonian of the theorem as it has a natural form:

$$0=\{\tilde{K}_n,H\}\Rightarrow\{T,V-\tfrac{1}{2}\tilde{k}_n\}=0\Rightarrow\tilde{k}_n=2V,$$

to within a constant.

(b) Applying the limit $\alpha_n\to\infty$ in the same way to the equations in part (b) of the theorem, we obtain

$$\begin{aligned}&\phi_i(V_{ii}-V_{nn})+3V_i+(2\phi_n+\alpha_i)V_{in}+\sum_{j\neq i}^{n-1}(\alpha_i-\alpha_j)^{-1}(\phi_i\phi_jV_{jn}-\phi_j^2V_{in})=0,\quad i=1,\ldots,n-1,\\&(\alpha_i-\alpha_r)^{-1}(\phi_i^2V_{rs}-\phi_i\phi_rV_{is})=(\alpha_i-\alpha_s)^{-1}(\phi_i^2V_{rs}-\phi_i\phi_sV_{ir})=0,\quad i,r,s=1,\ldots,n-1\text{ all different},\\&(\alpha_i-\alpha_r)^{-1}\phi_i\phi_r(V_{rr}-V_{ii})+(\alpha_i-\alpha_r)^{-1}(\phi_i^2-\phi_r^2)V_{ir}+\phi_iV_{nr}\\&\quad+\left(2\phi_n+\alpha_i-\sum_{j\neq i,r}^{n-1}(\alpha_i-\alpha_j)^{-1}\phi_j^2\right)V_{ir}+\sum_{j\neq i,r}^{n-1}(\alpha_i-\alpha_j)^{-1}\phi_i\phi_jV_{jr}\\&\quad+3(\alpha_i-\alpha_r)(\phi_iV_r-\phi_rV_i)=0,\quad i,r=1,\ldots,n-1,\quad i\neq r.\end{aligned}\tag{4.6}$$

(c) This translates to the following: "The Hamilton–Jacobi equation is separable in parabolic coordinates given by

$$\sum_{k=1}^{n-1}\frac{\phi_k^2}{z-\alpha_k}+z-2\phi_n=\frac{U(z)}{\mathring{A}(z)}."$$

(iv) In this case the limit is $\alpha_1-\alpha_2\to0$.

Consider the n integrals of the theorem,

$$K_i=\sum_{j\neq i}\frac{l_{ij}^2}{\alpha_i-\alpha_j}+p_i^2+k_i(q),\quad i=1,\ldots,n.$$

As in Ref. 5 we take instead the integrals

$$K=K_1+K_2,\quad(\alpha_2-\alpha_2)(K_1-K_2),\quad K_i,\quad i\geqslant3,$$

and then let $\alpha_1-\alpha_2\to0$. This gives us, respectively, K, l_{12}, K_i $(i\geqslant3)$—n integrals in involution.

We have an obvious symmetry in the system: invariance with respect to rotations in the (q_1,q_2) plane. This means we can reduce the system in the usual way by one degree of freedom before applying the theorem. In the framework of the Hamilton–Jacobi equation the reduction procedure is just precisely that of separating off one variable. We illustrate the procedure.

Let

$$\begin{aligned}&q_1=\rho\cos\theta,\quad q_2=\rho\sin\theta,\\&p_1=\dot{q}_1=\dot{\rho}\cos\theta-\rho^2\dot{\theta}\sin\theta,\\&p_2=\dot{q}_2=\dot{\rho}\sin\theta-\rho^2\dot{\theta}\cos\theta.\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial H}{\partial\dot{\rho}}&=\frac{\partial}{\partial\dot{\rho}}\left(\frac{1}{2}p_1^2+\frac{1}{2}p_2^2\right)\\&=\frac{\partial}{\partial\dot{\rho}}\left(\frac{1}{2}\dot{\rho}^2+\frac{1}{2}\rho^2\dot{\theta}^2\right)=\dot{\rho},\end{aligned}$$

$$\frac{\partial H}{\partial \dot{\theta}} = \rho^2 \dot{\theta}.$$

Let the momentum conjugate to ρ be ξ and the momentum conjugate to θ be η; then $\xi = \dot{\rho}$, $\eta = \rho^2\dot{\theta}$. In (ρ,θ,ξ,η) variables l_{12} is just η. Hence θ is a cyclic variable and η is a constant of motion.

So we have

$$H = \tfrac{1}{2}(\xi^2 + p_3^2 + \cdots + p_n^2) + V(\rho,q_3,\ldots,q_n) + \eta/\rho^2$$
$$= \tfrac{1}{2}(\xi^2 + p_3^2 + \cdots + p_2^2) + \tilde{V}(\rho,q_3,\ldots,q_n). \quad (4.7)$$

Look at the constant $K = K_1 + K_2$:

$$l_{1j}^2 + l_{2j}^2 = (\rho p_j - \xi q_j)^2 + \eta^2 q_j^2/\rho^2$$

in the new variables, so

$$K = \sum_{j=3}^{n} (\alpha_2 - \alpha_j)^{-1}(\rho p_j - \xi q_j)^2 + \xi^2 + \frac{\eta^2}{\rho^2} + k_1(q) + k_2(q) + \frac{\eta^2}{\rho^2}\sum_{j=3}^{n}(\alpha_2 - \alpha_j)^{-1} q_j^2.$$

Note that as θ is cyclic all k_i must be independent of θ $(i = 3,\ldots,n)$ as must $k_1 + k_2$.

Set $\phi_2 = \rho$, $\phi_j = q_j$, $j \geqslant 3$, $\xi_2 = \xi$, $\xi_j = p_j$, $j \geqslant 3$; then we have $n-1$ integrals in involution K_i $(i = 2,\ldots,n)$ of the form

$$K_i = \sum_{j \neq i} (\alpha_i - \alpha_j)^{-1}(\xi_i\phi_j - \xi_j\phi_i)^2 + \xi_i^2 + \chi_i(\phi), \quad (4.8)$$

all commuting with the Hamiltonian

$$H = \tfrac{1}{2}(\xi_2^2 + \cdots + \xi_n^2) + \tilde{V}(\phi_2,\ldots,\phi_n). \quad (4.9)$$

We are now able to immediately apply the theorem.

APPENDIX A: ELLIPTIC COORDINATES ON $\mathbb{R}^n$ AND SEPARABILITY

The coordinates $(q_1,\ldots,q_n)$ are Euclidean (or Cartesian) coordinates on $\mathbb{R}^n$.

Ordinary generalized elliptic coordinates are defined by the transformation[4]

$$1 + \sum_{k=1}^{n} \frac{q_k^2}{z - \alpha_k} = \prod_{j=1}^{n}(z - u_j) \times \left(\prod_{k=1}^{n}(z - \alpha_k) = \frac{U(z)}{A(z)}\right)^{-1}. \quad (A1)$$

This transformation is defined with respect to the constants $(\alpha_1,\ldots,\alpha_n) \in \mathbb{R}^n$. We must assume in (A1) that all α's are different.

The metric on $\mathbb{R}^n$ is given by $ds^2 = \Sigma_{i=1}^n dq_i^2$. Let us see how we write this metric in terms of $\mathbf{u}$ variables.

We have

$$q_k^2 = \lim_{z \to \alpha_k}(z - \alpha_k)\frac{U(z)}{A(z)} = \frac{U(\alpha_k)}{A'(\alpha_k)}$$
$$\Rightarrow 2q_k dq_k = -\frac{U(\alpha_k)}{A'(\alpha_k)}\sum_{j=1}^{n}(\alpha_k - u_j)^{-1}\,du_j,$$

so

$$dq_k^2 = \frac{1}{4}\frac{U(\alpha_k)^2}{A'(\alpha_k)^2}\frac{1}{q_k^2} \times \sum_{i,j=1}^{n}(\alpha_k - u_i)^{-1}(\alpha_k - u_j)^{-1}\,du_i\,du_j$$
$$= \frac{1}{4}q_k^2 \sum_{i,j=1}^{n}(\alpha_k - u_i)^{-1}(\alpha_k - u_j)^{-1}\,du_i\,du_j.$$

Now use

$$(\alpha_k - u_i)^{-1}(\alpha_k - u_j)^{-1} = (u_i - u_j)^{-1}[(\alpha_k - u_i)^{-1} - (\alpha_k - u_j)^{-1}], \quad \text{for } i \neq j,$$

and

$$\sum_{k=1}^{n}\frac{q_k^2}{u_i - \alpha_k} = -1,$$

which comes directly from (A1), so that

$$\sum_{k=1}^{n} dq_k^2 = \frac{1}{4}\sum_{i=1}^{n} du_i^2 \sum_{k=1}^{n} q_k^2(\alpha_k - u_j)^{-2}.$$

Now observe that

$$\sum_{k=1}^{n}\frac{q_k^2}{(\alpha_k - u_i)^2} = -\left[\frac{d}{dz}\sum_{k=1}^{n}\frac{q_k^2}{(z - \alpha_k)^2}\right]\Bigg|_{z=u_i} = -\frac{U'(u_i)}{A(u_i)}.$$

Thus

$$ds^2 = -\frac{1}{4}\sum_{i=1}^{n}\frac{U'(u_i)}{A(u_i)}\,du_i^2. \quad (A2)$$

We use (A2) to transform a Hamiltonian function of the form $H = \frac{1}{2}p^2 + V(q)$ to (u,v) variables. We must find the canonical momentum v. We have

$$H = \frac{1}{2}\sum_{k=1}^{n}\left(\frac{dq_k}{dt}\right)^2 + V(q).$$

We assume that V can be written in terms of u so that $\tilde{V}(u) = V(q)$ when u and q are, respectively, the elliptic and Cartesian coordinates of the same point in $\mathbb{R}^n$. Then

$$H = -\frac{1}{8}\sum_{i=1}^{n}\frac{U'(u_i)}{A(u_i)}\left(\frac{du_i}{dt}\right)^2 + \tilde{V}(u).$$

Now $v = \partial H/\partial \dot{u}$, so

$$v_i = -\frac{1}{4}\frac{U'(u_i)}{A(u_i)}\dot{u}_i \Rightarrow \frac{U'(u_i)}{A(u_i)}\dot{u}_i^2 = 16\frac{A(u_i)}{U'(u_i)}v_i^2.$$

Thus

$$H = -2\sum_{i=1}^{n}\frac{A(u_i)}{U'(u_i)}v_i^2 + \tilde{V}(u). \quad (A3)$$

The Hamilton–Jacobi equation for H expressed in terms of the coordinates $\mathbf{u}$ is

$$-2\sum_{i=1}^{n}\frac{A(u_i)}{U'(u_i)}\left(\frac{\partial S}{\partial u_i}\right)^2 + \tilde{V}(u) = E. \quad (A4)$$

If we apply Stäckel's theorem to (A3) we see that $\tilde{V}$ must have the form

$$\tilde{V}(u)=\sum_{i=1}^{n}\frac{f_i(u_i)}{U'(u_i)}. \tag{A5}$$

If we use the identity

$$\sum_{i=1}^{n}\frac{P(u_i)}{U'(u_i)}=\eta_1 \tag{A6}$$

for any polynomial P of the form $P(z)=\eta_1 z^{n-1}+\eta_2 z^{n-2}+\cdots+\eta_n$, then if V has the form in (A5), (A4) can be separated by posing $S(u)=\Sigma_{i=1}^{n}S_i(u_i)$ to give

$$-2A(u_i)\left(\frac{\partial S}{\partial u_i}\right)^2+f_i(u_i)-P(u_i)=0, \tag{A7}$$

where $\eta_1=E$.

To prove (A6), consider the contour integral (with the same U as above)

$$\frac{1}{2\pi i}\oint_C\frac{z^s\,dz}{U(z)}\,dz, \quad \text{for fixed } (u_1,\ldots,u_n). \tag{A8}$$

If C encloses all the u_i this is just

$$\sum_{j=1}^{n}\frac{u_j^s}{U'(u_j)}.$$

The integral is also equal to minus the residue at infinity of $z^s/U(z)$; this residue is equal to the coefficient of z^{-1} in the expansion

$$\frac{z^s}{U(z)}=z^{s-n}\left[1+z^{-1}\sum_{i=1}^{n}u_i+z^{-2}\sum_{i=1}^{n}u_iu_j+\cdots\right]; \tag{A9}$$

$$\therefore\ \frac{1}{2\pi i}\oint_C\frac{z^s\,dz}{U(z)}\,dz=\begin{cases}0, & \text{for } s<n-1,\\ 1, & \text{for } s=n-1,\end{cases} \tag{A10}$$

and so on.

APPENDIX B: CASE (i) OF SEC. IV

The conditions (a), (b), (c) become the following, respectively.

(ã) The following are n global (nonindependent), involutive integrals:

$$K_j=\sum_{k\neq j}\frac{l_{jk}^2}{\alpha_j-\alpha_k}+k_j(q), \quad j=1,\ldots,n,$$

for some functions $k_1,\ldots,k_n$ of q only; $\alpha_1,\ldots,\alpha_n$ are n distinct constants.

(b̃) The potential V satisfies the set of $\frac{1}{2}n(n-1)^2$ equations,

$$\begin{aligned}&(\alpha_j-\alpha_r)^{-1}(q_j^2V_{rs}-q_jq_rV_{js})\\&\quad=(\alpha_j-\alpha_s)^{-1}(q_j^2V_{rs}-q_jq_sV_{jr}),\\&(\alpha_r-\alpha_s)^{-1}q_rq_s(V_{rr}-V_{ss})+(\alpha_r-\alpha_s)^{-1}(q_s^2-q_r^2)V_{rs}\\&\quad+\sum_{j\neq r,s}(\alpha_r-\alpha_j)^{-1}q_j^2V_{rs}-q_r\sum_{j\neq r,s}(\alpha_r-\alpha_j)^{-1}q_jV_{js}\\&\quad+3(\alpha_r-\alpha_s)^{-1}(q_sV_r-q_rV_s)=0.\end{aligned}$$

(c̃) The Hamilton–Jacobi equation for H is separable in spherical-elliptic coordinates $(q^2,w_1,\ldots,w_{n-1})$ given by

$$\begin{aligned}\sum_{k=1}^{n}\frac{q_k^2}{z-\alpha_k}&=q^2\prod_{j=1}^{n-1}(z-w_j)\left(\prod_{k=1}^{n}(z-\alpha_k)\right)^{-1}\\&=q^2\frac{W(z)}{A(z)}.\end{aligned}$$

Proof, (ã)⇒(b̃): In the same way as before, we construct $\Sigma_{i=1}^{n}(z-\alpha_i)^{-1}k_{i,r}$ and put $\Sigma_{i=1}^{n}(z-\alpha_i)^{-1}(k_{i,rs}-k_{i,sr})=0$. We get

$$\begin{aligned}&3(z-\alpha_r)^{-1}(z-\alpha_s)^{-1}(q_sV_r-q_rV_s)-(z-\alpha_r)^{-1}q_r\\&\quad\times\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_iV_{is}+(z-\alpha_s)^{-1}q_s\\&\quad\times\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_iV_{ir}+V_{rs}\sum_{i=1}^{n}(z-\alpha_i)^{-1}q_i^2\\&\quad\times\left((z-\alpha_r)^{-1}-(z-\alpha_s)^{-1}\right)=0.\end{aligned}$$

As before we let $z\to\alpha_i$, $i\neq r,s$, and then $z\to\alpha_r$ to get the equations in (b̃).

To go now to (c̃) we use

$$\begin{aligned}\frac{\partial}{\partial q_k}&=2q_k\frac{\partial}{\partial q^2}+2q_kq^{-2}\sum_{i=1}^{n-1}(\alpha_k-w_i)^{-1}\\&\quad\times\frac{A(w_i)}{W'(w_i)}\frac{\partial}{\partial w_i},\\\frac{\partial q_k}{\partial w_i}&=-\frac{1}{2}\frac{q_k}{\alpha_k-w_i},\\\sum_{k=1}^{n}q_k(z-\alpha_k)^{-1}\frac{\partial}{\partial q_k}&\\&=2q^2\frac{W(z)}{A(z)}\frac{\partial}{\partial q^2}+2\frac{W(z)}{A(z)}\sum_{i=1}^{n-1}\frac{A(w_i)}{W'(w_i)}\\&\quad\times(z-w_i)^{-1}\frac{\partial}{\partial w_i},\\q_r(z-\alpha_r)^{-1}\sum_{k=1}^{n}q_k(z-\alpha_k)^{-1}\frac{\partial}{\partial q_k}&\\&=q^2(z-\alpha_r)^{-1}\frac{W(z)}{A(z)}\frac{\partial}{\partial q_r}-2q_r\frac{W(z)}{A(z)}\\&\quad\times\sum_{i=1}^{n-1}\frac{A(w_i)}{W'(w_i)}(z-w_i)^{-1}(\alpha_r-w_i)^{-1}\frac{\partial}{\partial w_i}.\end{aligned}$$

In the same way as before, using now

$$\sum_{i=1}^{n-1}\frac{Q(w_i)}{W'(w_i)}=1$$

for any polynomial Q of degree $n-2$, we get this time

$$\begin{aligned}&\sum_{j\neq i}^{n-1}\frac{A(w_j)}{W'(w_j)}(\alpha_r-w_j)^{-1}(\alpha_s-w_j)^{-1}\\&\quad\times\left[(w_i-w_j)\frac{\partial^2V}{\partial w_i\partial w_j}-\frac{\partial V}{\partial w_i}-\frac{\partial V}{\partial w_j}\right]\\&\quad+\frac{\partial V}{\partial w_i}+q^2\frac{\partial^2V}{\partial w_i\,\partial q^2}=0.\end{aligned} \tag{B1}$$

Now $V(q^2,\mathbf{w})$ is separable in $(q^2,\mathbf{w})$ coordinates $\Leftrightarrow V$ has the form

$$V=\hat{V}(q^2)+q^{-2}\tilde{V}(\mathbf{w}),$$

where

$$\tilde{V}(\mathbf{w}) = \sum_{i=1}^{n-1} \frac{f_i(w_i)}{W'(w_i)}.$$

The way to read (B1), fixing r and i, is

$$\begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathscr{B}^{(r,i)} \end{pmatrix} \begin{pmatrix} x_0^{(i)} \\ \mathbf{x}^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix},$$

where $\mathscr{B}^{(r,i)}$ is the $(n-2)\times(n-2)$ matrix with components

$$\mathscr{B}^{(r,i)}{}_{sj} = \frac{A(w_j)}{W'(w_j)} (\alpha_r - w_j)^{-1}(\alpha_s - w_j)^{-1},$$

$$s,j = 1,\ldots,n-1, \quad s \neq r, \quad j \neq i,$$

$$x_0^{(i)} = \frac{\partial}{\partial w_i}\left(V + q^2 \frac{\partial V}{\partial q^2}\right),$$

and $\mathbf{x}^{(i)}$ is the $(n-2)$ component column vector,

$$x^{(i)}{}_j = \frac{\partial}{\partial w_i} \frac{\partial}{\partial w_i} ((w_i - w_j)V),$$

$$j = 1,\ldots,n-1, \quad j \neq i.$$

It is easy to check that $\det \mathscr{B}^{(r,i)} \neq 0$. Therefore, for any i, $x_0^{(i)} = 0$ and $x^{(i)}{}_j = 0$, $j = 1,\ldots,n-1$. We find that the general solution to (B1) is $V = \hat{V}(q^2) + q^{-2}\tilde{V}(\mathbf{w})$, where $\tilde{V}(\mathbf{w})$ has the form

$$\tilde{V}(\mathbf{w}) = \sum_{i=1}^{n-1} \frac{f_i(w_i)}{W'(w_i)}.$$

Proof, (c̃)⇒(ã): This time the separated equations are $\mathscr{W}\boldsymbol{\eta} = \mathbf{x} + \mathbf{y}$, where

$$x_i = -4A(w_i)\chi_i^2 - \varkappa w_i^{n-1}$$

and

$$y_i = f_i(w_i) + \varkappa w_i^{n-1}.$$

$\mathscr{W}_{ij} = w_i^{n-1-j}$, $\varkappa$ is as yet unspecified, and $\boldsymbol{\chi}$ is the momentum conjugate to $\mathbf{w}$. We find $\det \mathscr{W} \neq 0$. Set $\boldsymbol{\eta}^x = \mathscr{W}^{-1}\mathbf{x}$, $\boldsymbol{\eta}^y = \mathscr{W}^{-1}\mathbf{y}$. Let

$$\sum_{j=1}^{n-1} z^{n-1-j}\eta^x{}_j = A(z)L(z) - \varkappa z^{n-1}. \tag{B2}$$

The rhs of (B2) must be a polynomial of degree $n-2$, so $\varkappa$ will be chosen to be the coefficient of z^{n-1} in $A(z)L(z)$, which we suppose to be a polynomial of degree $n-1$. Claim that

$$L = \sum_{i=1}^{n} (z - \alpha_i)^{-1} \sum_{j \neq 1} (\alpha_i - \alpha_j)^{-1} l_{ij}^2. \tag{B3}$$

Then $A(z)L(z) - \varkappa z^{n-1}$ is a polynomial of degree $n-2$, and so is uniquely determined by giving its value at $n-1$ different points. Thus to prove the validity of (B2)⇒(B3) we can prove (B3)⇒(B2).

We find that the rhs of (B3) is

$$q^2 \frac{W(z)}{A(z)} \sum_{j=1}^{n} (z - \alpha_j)^{-1} p_j^2 - \left[\frac{1}{2}\frac{d}{dt}\left(q^2 \frac{W(z)}{A(z)}\right)\right]^2,$$

and if we evaluate this function at $z = w_i$ we get

$$-\frac{1}{4} q^4 \frac{W'(w_i)^2}{A(w_i)^2} \dot{w}_i^2 = -4\chi_i^2.$$

If then $L(z)$ is given by (B3), $A(w_i)L(w_i) = -4A(w_i)\chi_i^2$ as required, so that (B3)⇒(B2). $\varkappa$ = coefficient of z^{n-1} in $A(z)L(z)$. This is

$$\sum_{i \neq j} (\alpha_i - \alpha_j)^{-1} l_{ij}^2,$$

which is zero by the antisymmetry of $(\alpha_i - \alpha_j)^{-1} l_{ij}^2$.

We have

$$\sum_{j=1}^{n} z^{n-1-j}\eta^x{}_j = A(z)L(z), \quad \forall z,$$

with $L(z)$ given by (B3). As before we recover the constants

$$K_i = \sum_{i \neq j} (\alpha_i - \alpha_j)^{-1} l_{ij}^2 + k_i(q), \quad i = 1,\ldots,n.$$

[1] P. Havas, J. Math. Phys. **16**, 1461 (1975).
[2] E. G. Kalnins and W. Miller, Jr., J. Math. Phys. **27**, 1721 (1986).
[3] M. P. Iarov-Iarovoi, J. Appl. Math. Mech. **27**, 1499 (1964).
[4] S. Wojciechowski, Phys. Scr. **31**, 433 (1985).
[5] I. Marshall, Phys. Scr. **32**, 565 (1985).
[6] P. Stäckel, *Integration der Hamilton–Jacobi'schen gleichungen mittelst Separation der Variabeln* (Habilitationsschrift, Halle, 1891).
[7] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U.P., Cambridge, 1937), §§152 and 153.
[8] S. Wojciechowski, Phys. Lett. A **100**, 277 (1984).
[9] S. Wojciechowski, "Three families of integrable one-particle potentials," Universita di Roma, Dipartimento di Fisica preprint 440, February 1985.
[10] L. P. Eisenhart, Phys. Rev. **45**, 427 (1934); **74**, 87 (1948).
[11] A. Ankiewicz and C. Pask, J. Phys. A **16**, 4203 (1983).

Generalized eigenfunction expansions for scattering in inhomogeneous three-dimensional media

James H. Rose and Margaret Cheney[a)]
Center for Nondestructive Evaluation, Iowa State University, Ames, Iowa 50011

(Received 9 December 1987; accepted for publication 10 February 1988)

An infinite number of ways are developed for representing a function in terms of the (generalized) eigenfunctions of a three-dimensional scattering problem and simple known auxiliary functions. The freedom represented by this variety of expansions arises from the causal nature of the wave equations considered. The new expansions are shown to generalize both the Fourier and Radon transforms. An application of the new expansions to the inverse scattering problem is given. It is shown (under some restrictions) that the scattering amplitude and potential are related via one of the generalized transforms.

I. INTRODUCTION

The expansion of a function in terms of (generalized) eigenfunctions is very useful in scattering theory.[1,2] In this paper, we will exhibit an infinite number of related ways a function can be expanded in terms of the eigenfunctions and simple known auxiliary functions. These results hold for scattering solutions of a three-dimensional generalization of the Stürm–Liouville equation.[3,4] In particular our results are valid for the variable-velocity wave equation, Schrödinger's equation, and the acoustic wave equation with variable velocity and density.

Eigenfunction expansions for these equations are not new. For Schrödinger's equation and the wave equation, eigenfunction expansions were first found by Ikebe.[5] His methods were used by Schulenberger and Wilcox[6] to obtain eigenfunction expansions for a number of wave equations occurring in classical physics.

Our work differs from this previous work in a number of ways. First, we use completely different techniques. We extend Ikebe's result by using methods that do not depend on self-adjointness. We are thus able to avoid the vector formulation used by Schulenberger and Wilcox. Second, our results differ from previous ones in that we obtain eigenfunction expansions that contain arbitrary parameters. We show that a certain choice of these parameters yields expansions that are particularly useful in inverse scattering. Finally, we obtain generalizations of the Radon transform[7] as well.

The usefulness of the eigenfunction expansions for inverse scattering can be illustrated by considering the Schrödinger equation. For this equation it is well-known that the potential and the Born (weak scattering) approximation to the scattering amplitude are related by a Fourier transform.[1,2] That is, they are related by an expansion in free-space solutions of the wave equation. The new eigenfunction expansions allow us to generalize this relation. In particular it will be shown that the *exact* scattering amplitude and the potential are a generalized transform pair. A convenient representation of the potential in terms of the data (the scattering amplitude) and the eigenfunctions results.

The structure of this paper is as follows. It has two main parts. The first part (Secs. II–V) considers a fairly general form of the wave equation, which is assumed to have no bound states. The second part (Sec. VI), which deals with the Schrödinger equation, allows bound states. In Sec. II we introduce the wave equation, establish notation, and give a brief review of some needed elements of scattering theory. In Sec. III we prove our basic result, a representation of the delta function in terms of the eigenfunctions and simple known auxiliary functions. This provides us with an infinite variety of ways of expanding a known function. Reductions to known expansions are shown. Next, in Sec. IV, we discuss the sense in which our results generalize the Fourier and Radon transforms. In Sec. V a transform particularly suited to inverse scattering theory is developed. Section VI derives the generalized eigenfunction expansions for Schrödinger's equation when bound states are possible.

II. THE WAVE EQUATION, NOTATION, AND REVIEW OF SCATTERING THEORY

The wave equation treated in Secs. II–V of this paper has the form

$$(\nabla^2 - V(\mathbf{x})\omega^2 - q(\mathbf{x}) + \omega^2)\psi(\omega,\mathbf{x}) = 0. \tag{2.1}$$

Here $\mathbf{x}$ is a coordinate in R^3; ω, which denotes an angular frequency, is a real scalar; and ∇^2 is the Laplacian with respect to $\mathbf{x}$. The solution $\psi(\omega,\mathbf{x})$ is called the wave field. The scatterer is described by the two real functions $V(\mathbf{x})$ and $q(\mathbf{x})$. Precise conditions on $V(\mathbf{x})$ and $q(\mathbf{x})$ will be given later. Suffice it to say here that (1) $V(\mathbf{x})$ and $q(\mathbf{x})$ decay sufficiently rapidly to zero for large $x = |\mathbf{x}|$; (2) $1 - V$ and q are non-negative and $1 - V$ is bounded away from zero; (3) Eq. (2.1) has no bound states (i.e., solutions with rapid spatial falloff). We note that $V(\mathbf{x})$ is related to the local velocity of wave propagation, $c(\mathbf{x})$, by $V(\mathbf{x}) = 1 - c^{-2}(\mathbf{x})$. Since V and q go to 0 for large x, one can think of the scatterer as a relatively localized disturbance in an otherwise uniform medium. This background medium has velocity 1.

Equation (2.1) can be brought into a form which is a three-dimensional generalization of the Sturm–Liouville

[a)] On leave of absence from Department of Mathematics, Duke University, Durham, North Carolina 27706.

 0022-2488/88/061347-09$02.50

equation. In particular we make the formal substitution $\psi(\omega,\mathbf{x}) = \rho^{-1/2}(\mathbf{x})p(\omega,\mathbf{x})$ in (2.1). The following wave equation is obtained:

$$\nabla\cdot[(1/\rho)]\nabla p] + \kappa(\mathbf{x})\omega^2 p - Q(\mathbf{x})p = 0. \tag{2.2}$$

Here $\kappa(\mathbf{x}) = c^{-2}(\mathbf{x})\rho^{-1}(\mathbf{x})$, while

$$Q(\mathbf{x}) = [\tfrac{3}{4}|\nabla\rho|^2/\rho - \tfrac{1}{2}(\nabla^2\rho/\rho) - q(\mathbf{x})]/\rho(\mathbf{x}). \tag{2.3}$$

If one assumes that $\rho(\mathbf{x})$ is twice differentiable, positive, and bounded, then $\kappa(\mathbf{x})$ and $q(\mathbf{x})$ are defined. Finally we note that Eq. (2.2) has just the form of the Sturm–Liouville equation except that ∇ and $\nabla\cdot$ have replaced the derivatives with respect to distance in that equation.

A variety of physical problems can be modeled by Eq. (2.1). First, if one sets $q = 0$ and uses $V(\mathbf{x}) = 1 - c^{-2}(\mathbf{x})$, the wave equation

$$(\nabla^2 + \omega^2 c^{-2}(\mathbf{x}))\psi = 0 \tag{2.4}$$

is obtained. This equation is commonly used as a scalar wave approximation to a large number of physical problems. Similarly if one sets $V(\mathbf{x}) = 0$ in Eq. (2.1), the Schrödinger equation is obtained. Finally the acoustic wave equation[8] can be obtained by rewriting (2.1) as (2.2) and choosing $q(\mathbf{x})$ such that $Q(\mathbf{x}) = 0$. For the acoustic wave equation, p is interpreted as the excess pressure, $\kappa(\mathbf{x})$ as the compressibility, and $\rho(\mathbf{x})$ as the density. Both $\rho(\mathbf{x})$ and $\kappa(\mathbf{x})$ are assumed to approach constants ρ_0 and $\kappa_0 = 1/\rho_0$ for sufficiently large $\mathbf{x}$.

In order to define scattering solutions of (2.1), we will need the Green's functions

$$G_0^{\pm}(\omega,\mathbf{z}) = -(4\pi|\mathbf{z}|)^{-1}\exp(\pm i\omega|\mathbf{z}|). \tag{2.5}$$

Here $+$ and $-$ refer to radiation and incoming boundary conditions, respectively.

We will be interested in scattering solutions that correspond to an incident plane wave $\exp(i\omega\hat{e}\cdot\mathbf{x})$, where $\hat{e}$ is a unit vector denoting the direction of incidence. We define these solutions of (2.1) by the Lippman–Schwinger equation[2]

$$\psi^{\pm}(\omega,\hat{e},\mathbf{x}) = \exp(i\omega\hat{e}\cdot\mathbf{x}) + \int G_0^{\pm}(\omega,\mathbf{x}-\mathbf{y}) \times [q(\mathbf{y}) + \omega^2 V(\mathbf{y})]\psi^{\pm}(\omega,\hat{e},\mathbf{y})d^3\mathbf{y}. \tag{2.6}$$

It has been shown by Agmon[9,10] that Eq. (2.6) always has a unique solution provided that the following conditions hold:

(a) q and V are real,

(b) $q + \omega^2 V$ is locally L^2 in $\mathbf{x}$,

(c) $q(\mathbf{x}) + \omega^2 V(\mathbf{x}) = \mathcal{O}(x^{-5/2-\epsilon})$ at infinity for some $\epsilon > 0$,

(d) $q(\mathbf{x})$ has no zero-energy bound or half-bound states.

Condition (d) is relevant only in solving (2.6) with $\omega = 0$. We will refer to hypotheses (a)–(c) together as hypothesis A. Finally we note that

$$\psi^-(\omega,\hat{e},\mathbf{x}) = \psi^+(-\omega,-\hat{e},\mathbf{x}). \tag{2.6'}$$

We will also need the fully interacting Green's functions which are solutions of

$$(\nabla^2 - \omega^2 V(\mathbf{x}) - q(\mathbf{x}) + \omega^2)G^{\pm}(\omega,\mathbf{x},\mathbf{y}) = \delta^3(\mathbf{x}-\mathbf{y}). \tag{2.7}$$

These are related to the $G_0^{\pm}$ by[2]

$$G^{\pm}(\omega,\mathbf{x},\mathbf{y}) = G_0^{\pm}(\omega,\mathbf{x},\mathbf{y}) + \int d^3\mathbf{z}\, G_0^{\pm}(\omega,\mathbf{x},\mathbf{z})[\omega^2 V(\mathbf{z}) + q(\mathbf{z})]G^{\pm}(\omega,\mathbf{y},\mathbf{z}). \tag{2.8}$$

Equation (2.8), which is similar to (2.6), also has a unique solution under hypothesis A. This uniqueness of solutions implies that

$$G^+(-\omega,\mathbf{x},\mathbf{y}) = G^-(\omega,\mathbf{x},\mathbf{y}). \tag{2.9}$$

We will also need the fact that $G^+(\omega = 0,\mathbf{x},\mathbf{y}) = G^-(\omega = 0,\mathbf{x},\mathbf{y})$. This is obtained by substituting (2.5) in (2.8) and setting $\omega = 0$:

$$G^{\pm}(0,\mathbf{x},\mathbf{y}) = -(4\pi|\mathbf{x}-\mathbf{y}|)^{-1} - \int d^3\mathbf{z}(4\pi|\mathbf{x}-\mathbf{z}|)^{-1}q(\mathbf{z})G^{\pm}(0,\mathbf{y},\mathbf{z}). \tag{2.10}$$

Since (2.10) has a unique solution, it follows that $G^+(0,\mathbf{x},\mathbf{y}) = G^-(0,\mathbf{x},\mathbf{y})$.

The causal properties of the Green's functions are an essential feature of our derivation. These are most easily seen by taking the "time-domain" Fourier transform[11] and defining

$$\hat{G}^{\pm}(t,\mathbf{x},\mathbf{y}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega t}G^{\pm}(\omega,\mathbf{x},\mathbf{y}). \tag{2.11}$$

(Note that we use the caret to denote both unit vectors and time-domain Fourier transforms. The meaning is always clear from the context.) It can be shown[12] under hypothesis B, below, that $\hat{G}$ satisfies a certain Cauchy problem with initial data supported at $\mathbf{x} = \mathbf{y}$ at time $t = 0$. The domain of dependence properties (causality) for the wave equation then implies that for $c_m = \sup c(\mathbf{x})$,

$$\hat{G}^+(t,\mathbf{x},\mathbf{y}) = 0, \quad \text{for } t < |\mathbf{x}-\mathbf{y}|/c_m. \tag{2.12a}$$

Therefore, since $G^+(-t,\mathbf{x},\mathbf{y}) = G^-(t,\mathbf{x},\mathbf{y})$, we also have

$$\hat{G}^-(t,\mathbf{x},\mathbf{y}) = 0, \quad \text{for } t > -|\mathbf{x}-\mathbf{y}|/c_m. \tag{2.12b}$$

Hypothesis B is

(a) $q(\mathbf{x})$ is non-negative and bounded;

(b) $c^{-2}(\mathbf{x}) = 1 - V(\mathbf{x})$ is positive, bounded, and bounded away from zero (by c_m^{-2});

(c) $V(\mathbf{x})$ has two continuous derivatives.

Finally we note that if V and q have compact support, Eq. (2.6) can be expanded as[5]

$$\psi(\omega,\hat{e},\mathbf{x}) = \exp(i\omega\hat{e}\cdot\mathbf{x}) + A(\omega,\hat{e}',\hat{e})\exp(i\omega x)x^{-1} + o(x^{-1}). \tag{2.13}$$

The symbol $\hat{e}'$ is defined by $\hat{e}' = \mathbf{x}/x$ and denotes the direction of scattering. The function A is called the scattering amplitude. It is given by

$$A(\omega,\hat{e}',\hat{e}) = -(4\pi)^{-1}\int \exp(-i\omega\hat{e}'\cdot\mathbf{y}) \times [\omega^2 V(\mathbf{y}) + q(\mathbf{y})]\psi^+(\omega,\hat{e},\mathbf{y})d^3\mathbf{y}. \tag{2.14}$$

We note that A obeys reciprocity[2]; that is,

$$A(\omega,\hat{e}',\hat{e}) = A(\omega,-\hat{e},-\hat{e}'). \tag{2.15}$$

The scattering amplitude can also be obtained from "near-

field" measurements of ψ as follows.[13] Suppose V and q have compact support, and let Ω be a surface enclosing this support. Then

$$A(\omega,\hat{e}',\hat{e}) = -(4\pi)^{-1}\int_{\Omega}\exp(-i\omega\hat{e}'\cdot\mathbf{x}) \times[\hat{n}\cdot\nabla\psi^{+}(\omega,\hat{e},\mathbf{x}) + ik\hat{e}\cdot\hat{n}\psi^{+}(\omega,\hat{e},\mathbf{x})]d^{2}\mathbf{x}, \tag{2.16}$$

where $\hat{n}$ denotes the outward unit normal to Ω.

III. THE ORTHOGONALITY RELATIONS

This section contains the main results and their proofs. It is shown that a function $\varphi(\mathbf{x})$ can be expanded in terms of the eigenfunctions and simple known auxiliary functions. Our strategy is as follows. We start with Eq. (3.1), which is known for Schrödinger's equation, and show that it applies to our more general wave equation (2.1). Then we show that Eq. (3.1) *together with causality* yields a representation of $G^{+}(\omega=0,\mathbf{x},\mathbf{y})$ in terms of an integral over $\exp[i\omega\tau(\mathbf{x},\mathbf{y})]$ and a product of two eigenfunctions. Here $\tau(\mathbf{x},\mathbf{y})$ is, within certain constraints, at the disposal of the reader. This is the origin of the infinite number of expansions, one for each choice of $\tau(\mathbf{x},\mathbf{y})$. The auxiliary functions mentioned above are determined by the choice of τ. Next we obtain $\delta^{3}(\mathbf{x}-\mathbf{y})$ by operating on $G^{+}(\omega=0,\mathbf{x},\mathbf{y})$ with $[\nabla^{2}-q(\mathbf{x})]$. Finally we integrate $\delta^{3}(\mathbf{x}-\mathbf{y})$ with the test function $\varphi(\mathbf{x})$. The result is the expansion of φ in terms of the eigenfunctions and the auxiliary functions. These basic results are given by Corollary 3.3 in the frequency domain and by Corollary 3.4 in the time domain.

Proposition 3.1: Suppose hypotheses A and B are satisfied. Then the following equation holds in the distribution sense:

$$-8\pi^{2}(i\omega)^{-1}[G^{+}(\omega,\mathbf{x},\mathbf{y}) - G^{-}(\omega,\mathbf{x},\mathbf{y})] = \int_{S^{2}}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^{2}\hat{e}. \tag{3.1}$$

(The asterisk denotes complex conjugate.) In other words, if $\varphi\in C_{0}^{\infty}$ (infinitely differentiable functions with compact support),

$$\frac{-8\pi^{2}}{i\omega}\int[G^{+}(\omega,\mathbf{x},\mathbf{y}) - G^{-}(\omega,\mathbf{x},\mathbf{y})]\varphi(\mathbf{y})d^{3}\mathbf{y} = \int_{S^{2}}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\int\psi^{\pm *}(\omega,\hat{e},\mathbf{y})\varphi(\mathbf{y})d^{3}\mathbf{y}\,d^{2}\hat{e}.$$

Proof: Equation (3.1) holds for the Schrödinger equation.[14] We consider a Schrödinger equation with a parameter l:

$$[\nabla^{2}+\omega^{2}-l^{2}V(\mathbf{x})-q(\mathbf{x})]\psi = 0.$$

The corresponding ψ and G now depend on l. For each l and ω, (3.1) holds. Upon choosing $l=\omega$, we obtain (3.1) where ψ and G corresond to Eqs. (2.6) and (2.8). Q.E.D.

Theorem 3.2: Suppose hypotheses A and B are satisfied. Then for any τ satisfying $|\tau|\leqslant|\mathbf{x}-\mathbf{y}|/c_{m}$ and for $\mathbf{x}\neq\mathbf{y}$,

$$G^{+}(0,\mathbf{x},\mathbf{y}) = -(16\pi^{3})^{-1}\int_{-\infty}^{\infty}e^{i\omega\tau}\times\int_{S^{2}}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^{2}\hat{e}\,d\omega. \tag{3.2}$$

Equation (3.2) should be interpreted as a function of $\mathbf{y}$ in the distribution sense by

$$\int G^{+}(0,\mathbf{x},\mathbf{y})\varphi(\mathbf{y})d^{3}\mathbf{y} = -(16\pi^{3})^{-1}\int_{-\infty}^{\infty}\int_{S^{2}}\psi^{\pm}(\omega,\hat{e},\mathbf{x}) \times\int\psi^{\pm *}(\omega,\hat{e},\mathbf{y})e^{i\omega\tau(\mathbf{x},\mathbf{y})}\varphi(\mathbf{y})d\mathbf{y}\,d^{2}\hat{e}\,d\omega, \tag{3.3}$$

where $\varphi\in C_{0}^{\infty}$.

Proof: We consider the following distribution in τ:

$$Q(\tau,\mathbf{x},\mathbf{y}) = -(16\pi^{3})^{-1}\int_{-\infty}^{\infty}e^{i\omega\tau} \times\int_{S^{2}}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^{2}\hat{e}\,d\omega.$$

It operates on a test function η in C_{0}^{∞} by

$$\int_{-\infty}^{\infty}Q(\tau,\mathbf{x},\mathbf{y})\eta(\tau)d\tau = -(16\pi^{3})^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i\omega\tau} \times\int_{S^{2}}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^{2}\hat{e}\,d\omega\,\eta(\tau)d\tau. \tag{3.4}$$

In the right-hand side of (3.4), we use (3.1):

$$\int_{-\infty}^{\infty}Q(\tau,\mathbf{x},\mathbf{y})\eta(\tau)d\tau = (2\pi)^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i\omega\tau}(i\omega)^{-1} \times[G^{+}(\omega,\mathbf{x},\mathbf{y}) - G^{-}(\omega,\mathbf{x},\mathbf{y})]d\omega\,\eta(\tau)d\tau. \tag{3.5}$$

In the right-hand side of (3.5), we write

$$G(\omega,\mathbf{x},\mathbf{y}) = \int_{-\infty}^{\infty}e^{i\omega t}\hat{G}(t,\mathbf{x},\mathbf{y})dt. \tag{3.6}$$

We also interchange the order of integration in (3.5) by using the distributional definition of the Fourier transform.[15,16] The right-hand side of (3.5) is then

$$(2\pi)^{-1}\int_{-\infty}^{\infty}(i\omega)^{-1} \times\int_{-\infty}^{\infty}e^{i\omega t}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y}) - \hat{G}^{-}(t,\mathbf{x},\mathbf{y})]dt\,\check{\eta}(\omega)d\omega, \tag{3.7}$$

where

$$\check{\eta}(\omega) = \int_{-\infty}^{\infty}e^{i\omega\tau}\eta(\tau)d\tau. \tag{3.8}$$

Again we use the distributional definition of the Fourier transform in (3.7) to interchange the order of integration:

$$(2\pi)^{-1}\int_{-\infty}^{\infty}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y}) - \hat{G}^{-}(t,\mathbf{x},\mathbf{y})] \times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{i\omega(t+\tau)}(i\omega)^{-1}\eta(\tau)d\tau\,d\omega\,dt. \tag{3.9}$$

In (3.9) we must choose a regularization of $(i\omega)^{-1}$, but the choice we make does not matter in the end. For convenience,

choose the principal value. We then do the ω integral first (again using the distributional definition of the Fourier transform). The ω integral is equal[16] to $\pi\,\mathrm{sgn}(t+\tau)$. Thus (3.9) is equal to

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y})-\hat{G}^{-}(t,\mathbf{x},\mathbf{y})]\times\mathrm{sgn}(t+\tau)\eta(\tau)d\tau\,dt$$
$$=\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\tau}^{\infty}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y})-\hat{G}^{-}(t,\mathbf{x},\mathbf{y})]dt\,\eta(\tau)d\tau$$
$$-\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{-\tau}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y})-\hat{G}^{-}(t,\mathbf{x},\mathbf{y})]dt\times\eta(\tau)d\tau. \tag{3.10}$$

Next we use causality (2.12), which implies

$$\int_{-\tau}^{\infty}\hat{G}^{-}(t,\mathbf{x},\mathbf{y})dt=0,\quad\text{for}\quad -\frac{|\mathbf{x}-\mathbf{y}|}{c_m}<-\tau, \tag{3.11}$$

and

$$\int_{-\infty}^{-\tau}\hat{G}^{+}(t,\mathbf{x},\mathbf{y})dt=0,\quad\text{for}\quad -\tau<\frac{|\mathbf{x}-\mathbf{y}|}{c_m}. \tag{3.12}$$

These facts imply that

$$\int_{-\tau}^{\infty}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y})-\hat{G}^{-}(t,\mathbf{x},\mathbf{y})]dt=\int_{-\infty}^{\infty}\hat{G}^{+}(t,\mathbf{x},\mathbf{y})dt=G^{+}(0,\mathbf{x},\mathbf{y}) \tag{3.13}$$

and

$$\int_{-\infty}^{-\tau}[\hat{G}^{+}(t,\mathbf{x},\mathbf{y})-\hat{G}^{-}(t,\mathbf{x},\mathbf{y})]dt=\int_{-\infty}^{\infty}\hat{G}^{-}(t,\mathbf{x},\mathbf{y})dt=G^{-}(0,\mathbf{x},\mathbf{y}). \tag{3.14}$$

We use these facts in (3.10) and (3.5), noting that $G^{+}(0,\mathbf{x},\mathbf{y})=G^{-}(0,\mathbf{x},\mathbf{y})$:

$$\int_{-\infty}^{\infty}Q(\tau,\mathbf{x},\mathbf{y})\eta(\tau)d\tau=\int_{-\infty}^{\infty}G^{+}(0,\mathbf{x},\mathbf{y})\eta(\tau)d\tau, \tag{3.15}$$

where we have restricted the test function η to have support in $[-|\mathbf{x}-\mathbf{y}|/c_m,|\mathbf{x}-\mathbf{y}|/c_m]$. Q.E.D.

The following corollaries are our basic results; they give an infinite number of ways of expanding $\varphi(\mathbf{x})$. We note that the new freedom represented by these expansions arises because causality was satisfied in deriving (3.2) for any $\tau(\mathbf{x},\mathbf{y})$ such that $-|\mathbf{x}-\mathbf{y}|/c_m\leqslant\tau\leqslant|\mathbf{x}-\mathbf{y}|/c_m$. As we will see in the next section, the ordinary eigenfunction expansion is obtained by setting $\tau=0$.

Corollary 3.3: Suppose Hypotheses A and B are satisfied. Then for $|\tau|\leqslant|\mathbf{x}-\mathbf{y}|/c_m$,

$$\delta(\mathbf{x}-\mathbf{y})=-(16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}(\nabla^2-q)e^{i\omega\tau(\mathbf{x},\mathbf{y})}\times\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega. \tag{3.16}$$

The operator ∇^2-q may operate in either the $\mathbf{x}$ variable or the $\mathbf{y}$ variable. In terms of its action on a test function $\varphi\in C_0^{\infty}$, (3.16) is

$$\varphi(\mathbf{x})=-(16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}(\nabla^2-q)\psi^{\pm}(\omega,\hat{e},\mathbf{x})\times\int\psi^{\pm *}(\omega,\hat{e},\mathbf{y})e^{i\omega\tau(\mathbf{x},\mathbf{y})}\varphi(\mathbf{y})d^3\mathbf{y}\,d^2\hat{e}\,d\omega. \tag{3.17}$$

Proof: Applying $\nabla_y^2-V(\mathbf{y})$ to both sides of (3.2), we have

$$\delta(\mathbf{x}-\mathbf{y})=-(16\pi^3)^{-1}(\nabla_y^2-q(\mathbf{y}))\times\int_{-\infty}^{\infty}\int_{S^2}e^{i\omega\tau(\mathbf{x},\mathbf{y})}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega.$$

This must be interpreted by integrating against a test function $\varphi\in C_0^{\infty}$:

$$\varphi(\mathbf{x})=-(16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\times\int e^{i\omega\tau(\mathbf{x},\mathbf{y})}\psi^{\pm *}(\omega,\hat{e},\mathbf{y})(\nabla^2-q)\varphi(\mathbf{y})d^3\mathbf{y}\,d^2\hat{e}\,d\omega$$
$$=-(16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\times\int(\nabla^2-q)[e^{i\omega\tau(\mathbf{x},\mathbf{y})}\psi^{\pm *}(\omega,\hat{e},\mathbf{y})]\times\varphi(\mathbf{y})d^3\mathbf{y}\,d^2\hat{e}\,d\omega.$$

This proves (3.16) when ∇^2-q operates in the $\mathbf{y}$ variable.

To show that (3.16) holds when ∇^2-q operates in $\mathbf{x}$, we first note that the reciprocity relation $G(\omega,\mathbf{x},\mathbf{y})=G(\omega,\mathbf{y},\mathbf{x})$ implies that

$$[\nabla_x^2-q(\mathbf{x})]G(0,\mathbf{x},\mathbf{y})=\delta(\mathbf{x}-\mathbf{y})=[\nabla_y^2-q(\mathbf{y})]G(0,\mathbf{x},\mathbf{y}).$$

We can, therefore, apply $\nabla_x^2-q(\mathbf{x})$ to both sides of (3.2). We let the resulting distribution operate on test functions of the form $\varphi_1(\mathbf{x})\varphi_2(\mathbf{y})$ $(\varphi_1,\varphi_2\in C_0^{\infty})$. Thus we obtain

$$\int\varphi_1(\mathbf{x})\varphi_2(\mathbf{x})d^3\mathbf{x}=\iint\delta(\mathbf{x}-\mathbf{y})\varphi_1(\mathbf{x})\varphi_2(\mathbf{y})d^3\mathbf{x}\,d^3\mathbf{y}$$
$$=(-16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}\int e^{i\omega\tau(\mathbf{x},\mathbf{y})}\times\psi^{\pm}(\omega,\hat{e},\mathbf{x})(\nabla^2-q)\varphi_1(\mathbf{x})d^3\mathbf{x}\times\int\psi^{\pm *}(\omega,\hat{e},\mathbf{y})\varphi_2(\mathbf{y})d^3\mathbf{y}\,d^2\hat{e}\,d\omega. \tag{3.18}$$

In the right-hand side of (3.18), we let ∇^2-q operate on $e^{i\omega\tau(\mathbf{x},\mathbf{y})}\psi^{\pm}(\omega,\hat{e},\mathbf{x})$. We then apply the resulting formula to test functions φ_1 which approximate a delta function. Q.E.D.

Corollary 3.4: Suppose hypotheses A and B are satisfied. Then for $|\tau|\leqslant|\mathbf{x}-\mathbf{y}|/c_m$,

$$\delta(\mathbf{x}-\mathbf{y})=-(8\pi^2)^{-1}\int_{-\infty}^{\infty}\int_{S^2}(\nabla^2-q)\times u^{\pm}(t-\tau(\mathbf{x},\mathbf{y}),e,\mathbf{x})u^{\pm}(t,\hat{e},\mathbf{y})d^2\hat{e}\,dt, \tag{3.19}$$

where ∇^2-q may operate in either $\mathbf{x}$ or $\mathbf{y}$. Acting on a test function $\varphi\in C_0^{\infty}$, (3.19) is

$$\varphi(\mathbf{x}) = -(8\pi^2)^{-1}\int_{-\infty}^{\infty}\int_{S^2}(\nabla^2 - q)u^{\pm}(t-\tau(\mathbf{x},\mathbf{y}),e,\mathbf{x}) \times\int u^{\pm}(t,\hat{e},\mathbf{y})\varphi(\mathbf{y})d^3\mathbf{y}\,d^2\hat{e}\,dt. \tag{3.20}$$

Proof: In (3.2), we substitute

$$\psi^{\pm}(\omega,\hat{e},\mathbf{x}) = \int_{-\infty}^{\infty} e^{i\omega t}u^{\pm}(t,\hat{e},\mathbf{x})dt.$$

This results in

$$G^{+}(0,\mathbf{x},\mathbf{y}) = -(16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2} e^{i\omega\tau} \times\int_{-\infty}^{\infty} e^{i\omega t}u^{\pm}(t,\hat{e},\mathbf{x})dt\,\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega$$

$$= -(8\pi^2)^{-1}\int_{S^2}\int_{-\infty}^{\infty} u^{\pm}(t,\hat{e},\mathbf{x}) \times u^{\pm}(t+\tau,\hat{e},\mathbf{y})dt\,d^2\hat{e},$$

where $|\tau|\leqslant|\mathbf{x}-\mathbf{y}|/c_m$. We obtain (3.19) as in the proof of Corollary 3.3. Q.E.D.

IV. EIGENFUNCTION EXPANSIONS

Functions will be expanded in terms of eigenfunctions of the scattering problem and auxiliary functions determined from $\exp[i\omega\tau(\mathbf{x},\mathbf{y})]$. The representation of the delta function given in Eq. (3.16) is our starting point. Results will first be given for the frequency domain and then for the time domain. In the course of Sec. IV A, we will show that the frequency-domain expansions generalize the idea of the Fourier transform. Similarly Sec. IV B discusses the generalization of the Radon transform.

A. Frequency domain expansions

Equation (3.16) can be rewritten as

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1}(\nabla^2 - q(\mathbf{x})) \times\int_{-\infty}^{\infty} d\omega\int_{S^2} d^2\hat{e}\,\psi^{\pm}(\omega,\hat{e},\mathbf{x})\tilde{\varphi}^{\pm}(\omega,\hat{e},\mathbf{x}), \tag{4.1a}$$

where

$$\tilde{\varphi}^{\pm}(\omega,\hat{e},\mathbf{x}) = \int d^3\mathbf{y}\,\psi^{\pm *}(\omega,\hat{e},\mathbf{y})\varphi(\mathbf{y})e^{i\omega\tau(\mathbf{x},\mathbf{y})}. \tag{4.1b}$$

We remind the reader that it is essential that $\tau(\mathbf{x},\mathbf{y})$ be chosen so that $-|\mathbf{x}-\mathbf{y}|/c_m < \tau(\mathbf{x},\mathbf{y}) < |\mathbf{x}-\mathbf{y}|/c_m$. Equations (4.1a) and (4.1b) represent a rather general expansion. However, its utility in this form is not immediately clear since the expansion coefficients $\tilde{\varphi}^{\pm}$ depend in general on $\mathbf{x}$.

The nature of this difficulty can be seen by considering the Fourier transform, which, as we will see later, is a special case of (4.1a) and (4.1b). One great virtue of the Fourier transform is that it allows one to express a function $f(\mathbf{x})$ either in the $\mathbf{x}$ variable, or in its conjugate variable $\mathbf{q}$, $\hat{f}(\mathbf{q})$. As we have just noted Eqs. (4.1a) and (4.1b) lack this property for a general choice of $\tau(\mathbf{x},\mathbf{y})$. However, we will show that if $\tau(\mathbf{x},\mathbf{y})$ is chosen to have the following separable form:

$$\tau(\mathbf{x},\mathbf{y}) = -a(\mathbf{x}) + a(\mathbf{y}), \tag{4.2}$$

then this virtue of the Fourier transform is retained in our expansions. Here $a(\mathbf{x})$ is a real-valued function that has two bounded derivatives and is such that $-|\mathbf{x}-\mathbf{y}|/c_m < \tau(\mathbf{x},\mathbf{y}) < |\mathbf{x}-\mathbf{y}|/c_m$.

Substitution of (4.2) in (4.1) yields

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1}(\nabla^2 - q(\mathbf{x}))\int_{-\infty}^{\infty} d\omega \times\int_{S^2} d^2\hat{e}\,\psi^{\pm}(\omega,\hat{e},\mathbf{x})\exp[-i\omega a(\mathbf{x})]\tilde{\varphi}^{\pm}(\omega,\hat{e}) \tag{4.3a}$$

and

$$\tilde{\varphi}^{\pm}(\omega,\hat{e}) = \int d^3\mathbf{y}\,\psi^{\pm *}(\omega,\hat{e},\mathbf{y})\exp[i\omega a(\mathbf{y})]\varphi(\mathbf{y}). \tag{4.3b}$$

Thus Eqs. (4.3a) and (4.3b) allow us to express φ in terms of $\mathbf{x}$ or $\omega,\hat{e}$.

Particular choices of $a(\mathbf{x})$ lead to interesting results. We will consider three cases: (1) $a=0$; (2) $a(\mathbf{x}) = \hat{e}'\cdot\mathbf{x}/c_m$; and (3) $a(\mathbf{x}) = x/c_m$. The first case allows us to connect our results (1) to the well-known eigenfunction expansion method for Schrödinger's equation and (2) to the Fourier transform. The second case generates an expansion, which, as we will show in the next section, is of interest in inverse scattering. The third case was included because of its simple and beautiful form.

Case (1): a=0

Equations (4.3a) and (4.3b) become

$$\varphi(\mathbf{x}) = (16\pi^3)^{-1}c^{-2}(\mathbf{x})\int_{-\infty}^{\infty}\omega^2\,d\omega \times\int_{S^2} d^2\hat{e}\,\psi^{\pm}(\omega,\hat{e},\mathbf{x})\tilde{\varphi}^{\pm}(\omega,\hat{e}), \tag{4.4a}$$

and

$$\tilde{\varphi}^{\pm}(\omega,\hat{e}) = \int d^3\mathbf{y}\,\psi^{\pm *}(\omega,\hat{e},\mathbf{y})\varphi(\mathbf{y}). \tag{4.4b}$$

Here we have rewritten Eq. (4.1) with $\tau=0$ and used the wave equation (2.1) to evaluate $(-\Delta+q)\psi^{\pm}$. Equations (4.4a) and (4.4b) can be specialized to the Schrödinger equation case by setting $c(\mathbf{x})=1$. With this assumption, Eqs. (4.4a) and (4.4b) become the standard eigenfunction expansion[1,2] used in quantum theory. If we further set $q(\mathbf{x})=0$, then the solutions $\psi^{\pm}(\omega,\hat{e},\mathbf{x})$ become the free-space plane-wave solutions $\psi^{\pm}=\exp(i\omega\hat{e}\cdot\mathbf{x})$. Substitution of this result in (4.4a) and (4.4b) yields the usual Fourier transform. Thus we have shown that the Fourier transform is a special case of Eqs. (4.4a) and (4.4b), and consequently of Eqs. (4.1).

Returning to Eqs. (4.4a) and (4.4b), we note that if the velocity is variable, then a weighting factor of $c^{-2}(\mathbf{x})$ appears. This factor appears in an unsymmetrical fashion in (4.4a). However, like the $(2\pi)^{-1}$ that appears in the Fourier transform, this weighting factor can be arbitrarily written in either Eq. (4.4a) or (4.4b). For example, the following are equivalent to (4.4):

$$\varphi(\mathbf{x}) = (16\pi^3)^{-1} c^{-1}(\mathbf{x}) \int_{-\infty}^{\infty} \omega^2 \, d\omega \times \int_{S^2} d^2\hat{e}\, \psi^{\pm}(\omega,\hat{e},\mathbf{x}) \tilde{\varphi}^{\pm}(\omega,\hat{e}) \tag{4.5a}$$

and

$$\tilde{\varphi}^{\pm}(\omega,\hat{e}) = \int d^3\mathbf{y}\, c^{-1}(\mathbf{y}) \psi^{\pm *}(\omega,\hat{e},\mathbf{y}) \varphi(\mathbf{y}). \tag{4.5b}$$

Similarly $c^{-2}(\mathbf{y})$ could have appeared in Eq. (4.5b) while no powers of $c(\mathbf{x})$ appeared in Eq. (4.5a).

The acoustic wave equation was obtained from (2.1) by setting $\psi^{\pm} = \rho^{-1/2} p^{\pm}$ and using $\rho\kappa = c^{-2}$. Here κ is compressibility, p is the excess pressure, and ρ is the density. With these substitutions, Eqs. (4.4a) and (4.4b) become

$$\varphi(\mathbf{x}) = (16\pi^3)^{-1} \kappa(\mathbf{x}) \int_{-\infty}^{\infty} d\omega \times \int_{S^2} d^2\hat{e}\, p^{\pm}(\omega,\hat{e},\mathbf{x}) \tilde{\varphi}^{\pm}(\omega,\hat{e}) \tag{4.6a}$$

and

$$\tilde{\varphi}^{\pm}(\omega,\hat{e}) = \int d^3\mathbf{y}\, p^{\pm *}(\omega,\hat{e},\mathbf{y}) \varphi(\mathbf{y}). \tag{4.6b}$$

Equations (4.4)–(4.6) are new so far as we know.

Case (2): $a(\mathbf{x}) = \hat{e}' \cdot \mathbf{x}/c_m$

Here $\hat{e}'$ is a unit vector. This choice is, as we will see, interesting for the inverse scattering problem. Equation (4.3a) becomes

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1}(\nabla^2 - q(\mathbf{x})) \int_{-\infty}^{\infty} d\omega \times \int_{S^2} d^2\hat{e}\, \psi^{\pm}(\omega,\hat{e},\mathbf{x}) \exp\left(-i\omega\hat{e}' \cdot \frac{\mathbf{x}}{c_m}\right) \tilde{\varphi}^{\pm}(\omega,\hat{e},\hat{e}'), \tag{4.7a}$$

and (4.3b) becomes

$$\tilde{\varphi}^{\pm}(\omega,\hat{e},\hat{e}') = \int d^3\mathbf{y}\, \psi^{\pm *}(\omega,\hat{e},\mathbf{y}) \exp\left(i\omega\hat{e}' \cdot \frac{\mathbf{y}}{c_m}\right) \varphi(\mathbf{y}). \tag{4.7b}$$

Equation (4.7b) transforms a function of three variables, $\varphi(\mathbf{y})$, into a function of five variables, $\tilde{\varphi}(\omega,\hat{e},\hat{e}')$. However, as we see in the next section, this is precisely what is needed for the inverse problem.

Case (3): $a(\mathbf{x}) = x/c_m$

Equation (4.3a) becomes

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1}(\nabla^2 - q(\mathbf{x})) \int_{-\infty}^{\infty} d\omega \times \int_{S^2} d^2\hat{e}\, \psi^{\pm}(\omega,\hat{e},\mathbf{x}) \exp\left(-\frac{i\omega x}{c_m}\right) \tilde{\varphi}^{\pm}(\omega,\hat{e}), \tag{4.8a}$$

and (4.3b) becomes

$$\tilde{\varphi}^{\pm}(\omega,\hat{e}) = \int d^3\mathbf{y}\, \psi^{\pm *}(\omega,\hat{e},\mathbf{y}) \exp\left(\frac{i\omega y}{c_m}\right) \varphi(\mathbf{y}). \tag{4.8b}$$

The utility of (4.8a) and (4.8b) is not immediately clear. However, their simplicity commend them to our attention.

B. Time-domain expansions

The expansions given above can also be usefully expressed in the time domain. One simply Fourier transforms the appropriate equations with respect to frequency. The time-domain formulation is a generalization of the Radon transform.

We start with the transform of Eqs. (4.3a) and (4.3b) where $\tau(\mathbf{x},\mathbf{y}) = -a(\mathbf{x}) + a(\mathbf{y})$. The result is

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1}(\nabla^2 - q(\mathbf{x})) \int_{S^2} d^2\hat{e} \times \int_{-\infty}^{\infty} d\tau\, u^{\pm}(\tau + a(\mathbf{x}),\hat{e},\mathbf{x}) \Phi^{\pm}(\tau,\hat{e}) \tag{4.9a}$$

and

$$\Phi^{\pm}(\tau,\hat{e}) = \int d^3\mathbf{y}\, \varphi(\mathbf{y}) u^{\pm}(\tau + a(\mathbf{y}),\hat{e},\mathbf{y}). \tag{4.9b}$$

Particular choices of a lead to useful results.

Case (1): $a = 0$

Equations (4.9a) and (4.9b) become

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1} c^{-2}(\mathbf{x}) \int_{S^2} d^2\hat{e}' \times \int_{-\infty}^{\infty} d\tau \frac{\partial^2 u^{\pm}}{\partial\tau^2}(\tau,\hat{e},\mathbf{x}) \Phi^{\pm}(\tau,\hat{e}) \tag{4.10a}$$

and

$$\Phi^{\pm}(\tau,\hat{e}) = \int d^3\mathbf{y}\, \varphi(\mathbf{y}) u^{\pm}(\tau,\hat{e},\mathbf{y}). \tag{4.10b}$$

Here we have used the time-domain form of the wave equation [Eq. (2.1)],

$$\nabla^2 u(t,\mathbf{x}) - c^{-2}(\mathbf{x}) \frac{\partial^2}{\partial t^2} u(t,\mathbf{x}) - q(\mathbf{x}) u(t,\mathbf{x}) = 0, \tag{4.11}$$

in obtaining (4.10a) from (4.9a).

Equations (4.10a) and (4.10b) can now be shown to reduce to the Radon transform as a special case. Let us consider Eqs. (4.10a) and (4.10b) with $V(\mathbf{x})$ and $q(\mathbf{x})$ set equal to 0. The solutions of the Lippmann–Schwinger equation are then plane waves; in the time domain they are given by $\delta(t - \hat{e}\cdot\mathbf{x})$. Substitution of these delta-function plane waves in (4.10a) and (4.10b) yields[11]

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1} \int_{S^2} d^2\hat{e} \times \int_{-\infty}^{\infty} d\tau \frac{\partial^2 \delta}{\partial\tau^2}(\tau - \hat{e}\cdot\mathbf{x}) \Phi^{\pm}(\tau,\hat{e}) \tag{4.12a}$$

and

$$\Phi^{\pm}(\tau,\hat{e}) = \int d^3\mathbf{y}\, \varphi(\mathbf{y}) \delta(\tau - \hat{e}\cdot\mathbf{y}). \tag{4.12b}$$

Case (2): $a(\mathbf{x}) = \hat{e}' \cdot \mathbf{x}/c_m$

Equations (4.7a) and (4.7b) become, after a time-domain transform,

$$\varphi(\mathbf{x}) = -(16\pi^3)^{-1}[\nabla^2 - q(\mathbf{x})]\int_{S^2} d^2\hat{e} \times \int_{-\infty}^{\infty} d\tau\, u^{\pm}\left(\tau + \hat{e}'\cdot\frac{\mathbf{x}}{c_m}, \hat{e}, \mathbf{x}\right)\Phi^{\pm}(\tau,\hat{e},\hat{e}') \tag{4.13a}$$

and

$$\Phi^{\pm}(\tau,\hat{e},\hat{e}') = \int d^3\mathbf{y}\, u^{\pm}\left(\tau + \hat{e}'\cdot\frac{\mathbf{y}}{c_m}, \hat{e}, \mathbf{y}\right)\varphi(\mathbf{y}). \tag{4.13b}$$

Case (3): $a=(\mathbf{x})x/c_m$

Time-domain equations are easily obtained by inserting this form of a in Eqs. (4.9a) and (4.9b).

V. APPLICATIONS TO INVERSE SCATTERING

Formulas representing the properties of a scatterer in terms of the scattering amplitude and the wave field are important in inverse scattering theory. Such formulas have recently been emphasized by Newton[17,18] and by Rose and Cheney.[19] In this section we will derive representations of the potential for two special cases of (2.1): (1) the variable velocity wave equation [$q=0$, $c(\mathbf{x})\leqslant 1$]; and (2) Schrödinger's equation [$V=0$, $c(\mathbf{x})=1$]. Then it will be shown that the scattering amplitude and the potential are, for these special cases, transform pairs as in Eqs. (4.7a) and (4.7b).

We start with the variable velocity wave equation ($q=0$) with the further constraints that the velocity is everywhere less than or equal to 1 ($c_m=1$), and that $V\in C_0^\infty$. Equations (4.7a) and (4.7b) become, after replacing $\varphi(\mathbf{x})$ by $V(\mathbf{x})$,

$$V(\mathbf{x}) = -(16\pi^3)^{-1}\nabla^2\int_{-\infty}^{\infty} d\omega \int_{S^2} d^2\hat{e}\, \psi^-(\omega,\hat{e},\mathbf{x}) \times \exp(-i\omega\hat{e}'\cdot\mathbf{x})\tilde{V}^-(\omega,\hat{e},\hat{e}') \tag{5.1a}$$

and

$$\tilde{V}^-(\omega,e,\hat{e}') = \int d^3\mathbf{y}\, \psi^{-*}(\omega,\hat{e},\mathbf{y})\exp(i\omega\hat{e}'\cdot\mathbf{y})V(\mathbf{y}). \tag{5.1b}$$

We then use Eq. (2.6′) and the fact that $\psi^*(\omega)=\psi(-\omega)$ to obtain

$$\tilde{V}^-(\omega,\hat{e},\hat{e}') = \int d^3\mathbf{y}\, \psi^+(\omega,-\hat{e},\mathbf{y})e^{i\omega\hat{e}'\cdot\mathbf{y}}V(y). \tag{5.2}$$

We now note, by comparing Eqs. (2.14) and (5.2), that

$$\tilde{V}^-(\omega,\hat{e},\hat{e}') = -4\pi\omega^{-2}A(\omega,-\hat{e}',-\hat{e}). \tag{5.3}$$

Next, using reciprocity [Eq. (2.15)], Eq. (5.3) becomes

$$\tilde{V}^-(\omega,\hat{e},\hat{e}') = -4\pi\omega^{-2}A(\omega,\hat{e},\hat{e}'). \tag{5.4}$$

Thus the transform of $V(\mathbf{x})$ can be written immediately in terms of the scattering amplitude. A representation of the potential in terms of the scattering amplitude and wavefield can be obtained by substituting (5.4) in (5.1a). One finds

$$V(\mathbf{x}) = -(4\pi^2)^{-1}\nabla^2\int_{-\infty}^{\infty} d\omega \times \int_{S^2} d^2\hat{e}\, \psi^{+*}(\omega,-\hat{e},\mathbf{x})e^{-i\omega\hat{e}'\cdot\mathbf{x}}\frac{A(\omega,\hat{e},\hat{e}')}{\omega^2}. \tag{5.5}$$

Equations (5.4) and (5.5) show that $V(\mathbf{x})$ and $-4\pi A(\omega)/\omega^2$ are transform pairs. We note that Eq. (5.5) is an essential part of the self-consistent equation approach to inverse scattering recently proposed by Rose and Cheney.[19]

A representation of the potential can also be found for Schrödinger's equation [$V=0 \Rightarrow c(\mathbf{x})=1$]. Again we assume that $q\in C_0^\infty$. This representation follows from Eqs. (4.7a) and (4.7b) by taking $\varphi(\mathbf{x})$ to be $q(\mathbf{x})$:

$$q(\mathbf{x}) = -(16\pi^3)^{-1}[\nabla^2 - q(\mathbf{x})]\int_{-\infty}^{\infty} d\omega \times \int_{S^2} d\hat{e}\, \psi^-(\omega,\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x})\tilde{q}^-(\omega,\hat{e},\hat{e}') \tag{5.6a}$$

and

$$\tilde{q}^-(\omega,\hat{e},\hat{e}') = -4\pi A(\omega,\hat{e},\hat{e}'). \tag{5.6b}$$

In obtaining (5.6b) we followed the line of reasoning from (5.1) to (5.4) but replaced $\omega^2 V$ by q. Substitution of (5.6b) in (5.6a) yields

$$q(\mathbf{x}) = -(4\pi^2)^{-1}(\nabla^2 - q(\mathbf{x}))\int_{-\infty}^{\infty} d\omega \times \int_{S^2} d^2\hat{e}\, \psi^-(\omega,\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x})A(\omega,\hat{e},\hat{e}'). \tag{5.7}$$

Solving (5.7) for $q(\mathbf{x})$ yields

$$q(\mathbf{x}) = \frac{-(4\pi^2)^{-1}\nabla^2\int_{-\infty}^{\infty} d\omega \int_{S^2} d^2\hat{e}\, \psi^{+*}(\omega,-\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x})A(\omega,\hat{e},\hat{e}')}{1-(4\pi^2)^{-1}\int_{-\infty}^{\infty} d\omega \int_{S^2} d^2\hat{e}\, \psi^{+*}(\omega,-\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x})A(\omega,\hat{e},\hat{e}')}. \tag{5.8}$$

Alternatively one can use Eq. (5.7) and $(\nabla^2 - q)\psi = -\omega^2\psi$ to obtain

$$q(\mathbf{x}) = (2\pi^2)^{-1}\hat{e}'\cdot\nabla\int_{-\infty}^{\infty} d\omega\, i\omega\int_{S^2} d^2\hat{e}\, \psi^{+*}(\omega,-\hat{e},\mathbf{x}) \times \exp(-i\omega\hat{e}'\cdot\mathbf{x})A(\omega,\hat{e},\hat{e}'). \tag{5.9}$$

Here we have used

$$(\Delta - q(\mathbf{x}))\exp(-i\omega\hat{e}'\cdot\mathbf{x})\psi^{+*}(\omega,-\hat{e},\mathbf{x}) = -2i\omega\hat{e}\cdot\nabla[\exp(-i\omega\hat{e}'\cdot\mathbf{x})\psi^{+*}(\omega,-\hat{e},\mathbf{x})]. \tag{5.10}$$

Equation (5.9) is Newton's representation[17,18] of the potential for Schrödinger's equation.

VI. SCHRÖDINGER EQUATION WITH BOUND STATES

In this section we treat the Schrödinger equation

$$[\nabla^2 + \omega^2 - q(\mathbf{x})]\psi(\omega,\mathbf{x}) = 0, \tag{6.1}$$

where $\nabla^2 - q$ is now allowed to have negative eigenvalues. We assume that the eigenvalues are $-\kappa_n^2$, $n = 1,...,N$; we will denote the corresponding eigenfunctions (bound states) by $\psi_n^b(\mathbf{x})$, $n = 1,...,N$, $b = 1,2,...,m_n$. Here m_n denotes the multiplicity of the nth bound state. When there are bound states, the usual eigenfunction expansion[2] can be written in the form

$$\delta(\mathbf{x}-\mathbf{y}) = (16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x}) \times\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,\omega^2\,d\omega + \sum_{n,b}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y}); \tag{6.2}$$

this implies that the Green's function can be written[2]

$$G^{\pm}(\omega,\mathbf{x},\mathbf{y}) = (16\pi^3)^{-1}\int_{-\infty}^{\infty}\int_{S^2}\frac{\psi^{\pm}(\omega',\hat{e},\mathbf{x})\psi^{\pm *}(\omega',\hat{e},\mathbf{y})}{\omega^2-\omega'^2\pm i0}d^2\hat{e} \times\omega'^2\,d\omega' + \sum_{n,b}\frac{\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})}{\omega^2+\kappa_n^2}. \tag{6.3}$$

Equations (6.2) and (6.3) hold under the following conditions:

(a) q is real,

(b) q is locally L^2,

(c) $q(x) = \mathcal{O}(x^{-5/2-\epsilon})$ at infinity for some $\epsilon > 0$.

We will refer to hypotheses (a), (b), and (c) together as hypothesis C.

We will also need the assumption that for $\omega = 0$, Eq. (6.1) has no solutions that decay at infinity in x; that is, there are no $\omega = 0$ bound or half-bound states. As in our earlier arguments, we will need causality. This is more complicated in the presence of bound states, as we see from the following proposition.

Proposition 6.1: Suppose hypothesis C holds and q has no $\omega = 0$ bound or half-bound states. Then the time-domain Green's functions $\hat{G}^{\pm}$, which are defined by (2.11), satisfy

$$\hat{G}^{+}(t,\mathbf{x},\mathbf{y}) - \sum_{n,b}(2\kappa_n)^{-1}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\exp(\kappa_n t) = 0, \quad \text{for} \quad t < |\mathbf{x}-\mathbf{y}| \tag{6.4}$$

and

$$\hat{G}^{-}(t,\mathbf{x},\mathbf{y}) - \sum_{n,b}(2\kappa_n)^{-1}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\exp(-\kappa_n t) = 0, \quad \text{for} \quad t > -|\mathbf{x}-\mathbf{y}|. \tag{6.5}$$

Proof: We begin with the frequency-domain Green's functions, which are defined by the $V = 0$ version of (2.8). We consider the ω in (2.8) to be a complex variable. If we multiply Eq. (2.8) by $|q(\mathbf{x})|^{1/2}\exp(-i\omega|\mathbf{x}-\mathbf{y}|)$, we see from the analytic Fredholm theorem[20,21] that $\exp(-i\omega|\mathbf{x}-\mathbf{y}|)G^{+}(\omega,\mathbf{x},\mathbf{y})$ is meromorphic in the upper half-plane and goes to $(-4\pi|\mathbf{x}-\mathbf{y}|)^{-1}$ as $\operatorname{Im}\omega\to\infty$. By arguments similar to those of Ref. 22, it can be shown that the difference $\exp(-i\omega|\mathbf{x}-\mathbf{y}|)G^{+}(\omega,\mathbf{x},\mathbf{y}) + 4\pi|\mathbf{x}-\mathbf{y}|)^{-1}$ goes to zero as $|\omega|\to\infty$. The Fourier transform

$$\int e^{-i\omega s}[e^{-i\omega|\mathbf{x}-\mathbf{y}|}G^{+}(\omega,\mathbf{x},\mathbf{y}) + (4\pi|\mathbf{x}-\mathbf{y}|)^{-1}]d\omega,$$

for $s < 0$, is, therefore, equal to $2\pi i$ times the sum of the residues. The residues can be calculated from Eq. (6.3); the result is (6.4). A similar argument yields (6.5). Q.E.D.

Theorem 6.2: Suppose hypothesis C holds and q has no $\omega = 0$ bound or half-bound states. Then for any τ satisfying $|\tau| \leqslant |\mathbf{x}-\mathbf{y}|$ and for $\mathbf{x} \neq \mathbf{y}$,

$$G^{+}(0,x,y) = -(16\pi^3)^{-1}\int_{-\infty}^{\infty}e^{i\omega\tau} \times\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega + \sum_{n,b}\kappa_n^{-2}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\cosh(\kappa_n\tau). \tag{6.6}$$

Proof: As in the proof of Theorem 3.2, we begin with Eq. (3.1). We Fourier transform it into the time domain, obtaining

$$\hat{G}^{+}(t,\mathbf{x},\mathbf{y}) - \hat{G}^{-}(t,\mathbf{x},\mathbf{y}) = -(16\pi^3)^{-1}\int_{-\infty}^{\infty}\exp(-i\omega t)i\omega \times\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega. \tag{6.7}$$

We now need a step analogous to separating the supports of $\hat{G}^{+}$ and $\hat{G}^{-}$. To do this, we must use Proposition 6.1. We add the term

$$\sum_{n,b}\kappa_n^{-1}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\frac{\exp(-\kappa_n t)-\exp(\kappa_n t)}{2}$$

to both sides of (6.7). On the left-hand side of the resulting equation, we now have expressions (6.4) and (6.5). We can then separate the supports of these expressions by multiplying by the Heaviside function $H(t-\tau)$, where $|\tau| \leqslant |\mathbf{x}-\mathbf{y}|$. We thus obtain

$$\hat{G}^{+}(t,\mathbf{x},\mathbf{y}) = H(t-\tau)\Big[-(16\pi^3)^{-1}\int_{-\infty}^{\infty}\exp(-i\omega t)i\omega \times\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega - \sum_{n,b}\kappa_n^{-1}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\sinh(\kappa_n t)\Big] + \sum(2\kappa_n)^{-1}\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\exp(\kappa_n t). \tag{6.8}$$

Finally, we transform back into the frequency domain and evaluate the result at $\omega = 0$; this gives us (6.6). Q.E.D.

Corollary 6.3: Suppose hypothesis C holds and q has no $\omega = 0$ bound or half-bound states. Then for any τ satisfying $|\tau| \leqslant |\mathbf{x}-\mathbf{y}|$,

$$\delta(\mathbf{x}-\mathbf{y}) = -(16\pi^3)^{-1}\int_{-\infty}^{\infty}(\nabla^2-q)e^{i\omega\tau(\mathbf{x},\mathbf{y})} \times\int_{S^2}\psi^{\pm}(\omega,\hat{e},\mathbf{x})\psi^{\pm *}(\omega,\hat{e},\mathbf{y})d^2\hat{e}\,d\omega + \sum_{n,b}\kappa_n^{-2}(\nabla^2-q)\psi_n^b(\mathbf{x})\psi_n^b(\mathbf{y})\cosh\kappa_n\tau(\mathbf{x},\mathbf{y}). \tag{6.9}$$

Again various choices of $\tau(\mathbf{x},\mathbf{y})$ can be used to illustrate the new expansion. First we set $\tau = 0$ in (6.9). Then one finds Eq. (6.2), the usual form of the completeness relation for Schrödinger's equation.

A form useful for scattering theory results from setting $\tau = \hat{e}'\cdot(\mathbf{y}-\mathbf{x})$ in (6.9) where $\hat{e}'$ is a unit vector. This can be seen by multiplying both sides of the resulting equation by $q(\mathbf{y})$ and integrating. After using

$$(\nabla^2 - q(\mathbf{x}))\psi(\omega,\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x}) = -2i\omega\hat{e}'\cdot\nabla[\psi(\omega,\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x})]$$

in the resulting equation, one obtains

$$q(\mathbf{x}) = \hat{e}'\cdot\nabla\left\{(2\pi^2)^{-1}\int_{-\infty}^{\infty} d\omega\, i\omega \times \int_{S^2} d^2\hat{e}\, \psi^{+\,*}(\omega,-\hat{e},\mathbf{x})\exp(-i\omega\hat{e}'\cdot\mathbf{x})A(\omega,\hat{e},\hat{e}') + \sum_{n,b}\frac{\psi_n^b(\mathbf{x})}{\kappa_n}\left[y_n^b(\hat{e}')e^{-\kappa_n\hat{e}'\cdot\mathbf{x}} - y_n^b(-\hat{e}')e^{\kappa_n\hat{e}'\cdot\mathbf{x}}\right]\right\}. \tag{6.10}$$

Here

$$y_n^b(\hat{e}') = \int d^3\mathbf{y}\, e^{\kappa_n\hat{e}'\cdot\mathbf{y}} q(\mathbf{y})\psi_n^b(\mathbf{y}).$$

Equation (6.10) is Newton's representation[17] for the potential of Schrödinger's equation in the presence of bound states. This representation plays an essential role in an inverse scattering method suggested in the same reference.

Note added in proof: A brief announcement of this work with extensions to the near-field problem has recently appeared in Ref. 23.

ACKNOWLEDGMENTS

The work of one of us (J. H. R.) was supported by the University/Industry Center for Nondestructive Evaluation at Iowa State University and was performed at Ames Laboratory. Ames Laboratory is operated for the U. S. Department of Energy by Iowa State University under Contract No. W-7405-ENG-82. The work of the other author (M. C.) was supported, in part, by the U.S. Office of Naval Research under Young Investigator Award N000-14-85-K0224.

[1] E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961).
[2] R. G. Newton, *Scattering Theory of Waves and Particles* (Springer, New York, 1982), 2nd ed.
[3] P. R. Wallace, *Mathematical Analysis of Physical Problems* (Dover, New York, 1984).
[4] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill, New York, 1955).
[5] T. Ikebe, Arch. Rat. Mech. Anal. **5**, 1 (1960); Publ. RIMS Kyoto Univ. **8**, 579 (1973).
[6] J. R. Schulenberger and C. H. Wilcox, Arch. Rat. Mech. Anal. **46**, 280 (1972).
[7] S. R. Deans, *The Radon Transform and Some of Its Applications* (Wiley, New York, 1983).
[8] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[9] S. Agmon, Ann. Sc. Norm. Sup. Pisa, Cl. Sci. Ser. IV **2**, 151 (1975).
[10] M. Reed and B. Simon, *Methods of Modern Mathematical Physics; IV, Analysis of Operators* (Academic, New York, 1978).
[11] J. H. Rose, M. Cheney, and B. DeFacio, J. Math. Phys. **25**, 2995 (1984).
[12] M. Cheney and E. Somersalo (unpublished).
[13] A. J. Devaney and G. Beylkin, Ultrason. Imaging **6**, 181 (1984).
[14] Reference 2, p. 286.
[15] V. S. Vladimirov, *Generalized Functions in Mathematical Physics* (MIR, Moscow, 1979).
[16] I. M. Gelfand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1964), Vol. 1.
[17] R. G. Newton, Phys. Rev. Lett. **53**, 1863 (1984).
[18] R. G. Newton, Phys. Rev. Lett. **43**, 541 (1979).
[19] J. H. Rose and M. Cheney, Phys. Rev. Lett. **59**, 954 (1987).
[20] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I. Functional Analysis* (Academic, New York, 1972).
[21] M. Reed and B. Simon, *Methods of Modern Mathematical Physics III. Scattering Theory* (Academic, New York, 1979).
[22] R. G. Newton, J. Math. Phys. **21**, 1698 (1980).
[23] M. Cheney and J. H. Rose, Phys. Rev. Lett. **60**, 1221 (1988).

Interference in the radiation of two relativistic charges

R. Hojman
Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile and Centro de Física El Trauco, Fanor Velasco 41 D, Santiago, Chile

G. Palma
Departamento de Física, Facultad de Ciencias, Universidad de Santiago de Chile, Casilla 5659, Correo 2, Santiago, Chile and Centro de Estudios Científicos de Santiago, Av. Presidente Errázuriz 3132, Santiago, Chile

L. Peña
Departamento de Física, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

R. Tabensky
Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Casilla 487-3 Santiago, Chile

(Received 7 April 1987; accepted for publication 13 August 1987)

The radiation produced by two interacting relativistic point charges is considered. Under an appropriate approximation the radiation of the system as a whole (interference effects included) is expressed as a function of the relative velocity.

I. INTRODUCTION

In this paper the radiation produced by the interference of two interacting relativistic point charges is found both exactly and within an approximation valid under specific requirements.

In Sec. II the geometrical background is presented. Section III is devoted to getting the exact expression for the radiated momentum (interference effects included) for a system of two interacting relativistic charges moving arbitrarily. An approximation is introduced in Sec. IV. Its validity depends on the relative distance and relative velocity of the considered particles. In Sec. V a couple of examples are exhibited. The last section is conclusions. An Appendix includes some useful expressions and the main steps for performing some calculations.

II. NEWMANN–PENROSE RETARDED COORDINATES

In order to define the Newmann–Penrose retarded coordinates it is necessary to have a world line $x^\mu = z^\mu(\tau)$ in Minkowski space. One can then define four coordinates $(\tau,\kappa,\theta,\phi)$ associated with any point x^μ in Minkowski space as follows. The retarded time τ is the solution of the equation

$$[x^\mu - z^\mu(\tau)]\{x_\mu - z_\mu(\tau)\} = 0\,, \tag{2.1}$$

where

$$z_\mu(\tau) \equiv \eta_{\mu\nu} z^\nu(\tau) \tag{2.2}$$

with

$$\eta_{\mu\nu} \equiv \mathrm{diag}(1,-1,-1,-1)\,. \tag{2.3}$$

The coordinate κ is defined by

$$\kappa = v_\mu(\tau)[x^\mu - z^\mu(\tau)] \tag{2.4}$$

with

$$v_\mu(\tau) \equiv \dot{z}_\mu = \frac{dz_\mu}{d\tau}\,. \tag{2.5}$$

Geometrically, κ represents the spatial distance between x^μ and $z^\mu(\tau)$ as measured from a frame whose temporal axis is $v_\mu(\tau)$ (τ is the retarded proper time of x^μ).

Finally if we define

$$[x^\mu - z^\mu(\tau)]_\perp \equiv (\eta^{\mu\nu} - t^\mu t^\nu)[x_\nu - z_\nu(\tau)]\,, \tag{2.6}$$

then θ and ϕ are the polar angles of the three-vector $[x^\mu - z^\mu(\tau)]_\perp$ referred to an arbitrary inertial system whose temporal axis has the direction of a unitary vector t^μ.

III. TWO-PARTICLE SYSTEMS

Let us consider two (pointlike) charges q_1 and q_2. The electromagnetic field $F_i^{\mu\nu}$ $(i = 1,2)$ constructed from the Lienard–Wiechert potentials for q_i is given by[1]

$$F_i^{\mu\nu} = \frac{q_i}{\sqrt{4\pi}}\left[K_i^{[\mu}\dot{v}_i^{\nu]}/\kappa_i + (1 - \kappa_i(\dot{v}_i\cdot K_i))(K_i^{[\mu}v_i^{\nu]}/\kappa_i^2)\right], \tag{3.1}$$

where

$$a^{[\mu}b^{\nu]} = a^\mu b^\nu - a^\nu b^\mu\,, \tag{3.2}$$

$$(a\cdot b) = a^\mu b_\mu\,, \tag{3.3}$$

and

$$K^\mu = \kappa^{-1}[x^\mu - z^\mu(\tau)] \tag{3.4}$$

is a lightlike vector.

The total electromagnetic field produced by the considered system is

$$F^{\mu\nu} = F_1^{\mu\nu} + F_2^{\mu\nu}\,. \tag{3.5}$$

The energy momentum tensor $T^{\mu\nu}$ related to an electromagnetic field $F^{\mu\nu}$ is

$$T^{\mu\nu} = F^{\mu\sigma}F_\sigma{}^\nu + \tfrac{1}{4}\eta^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}\,. \tag{3.6}$$

Of course

$$T^{\mu\nu}{}_{,\nu} = 0 \tag{3.7}$$

 0022-2488/88/061356-05$02.50

for any field which satisfies Maxwell equations.

Inserting in Eq. (3.6) the decomposition (3.5) for $F^{\mu\nu}$, $T^{\mu\nu}$ can be split as follows:

$$T^{\mu\nu} = T^{\mu\nu}_{\text{Mix}} + T_1{}^{\mu\nu} + T_2{}^{\mu\nu}, \tag{3.8}$$

where

$$T^{\mu\nu}_{\text{Mix}} = F_1{}^{\mu\sigma}F_{2\sigma}{}^{\nu} + F_2{}^{\mu\sigma}F_{1\sigma}{}^{\nu} + \tfrac{1}{2}\eta^{\mu\nu}F_1{}^{\alpha\beta}F_{2\alpha\beta}, \tag{3.9}$$

$$T_i{}^{\mu\nu} = F_i{}^{\mu\sigma}F_{i\sigma}{}^{\nu} + \tfrac{1}{4}\eta^{\mu\nu}F_i{}^{\alpha\beta}F_{i\alpha\beta}, \quad i = 1,2. \tag{3.10}$$

As usual,[2] the momentum radiated $P^{\mu}{}_{R}$ is defined as

$$P^{\mu}{}_{R} = -\int_{\Sigma} T^{\mu\nu}\, d\Sigma_{\nu}, \tag{3.11}$$

where Σ is a timelike surface that will eventually be considered to tend to infinity.

Equation (3.8) allows us to write

$$P^{\mu}{}_{R} = P^{\mu}_{R\,\text{Mix}} + P^{\mu}_{R\,1} + P^{\mu}_{R\,2}, \tag{3.12}$$

where

$$P^{\mu}_{R\,\text{Mix}} = -\int_{\Sigma^{\infty}} T^{\mu\nu}_{\text{Mix}}\, d\Sigma_{\nu}, \tag{3.13}$$

$$P^{\mu}_{Ri} = -\int_{\Sigma^{\infty}} T_i{}^{\mu\nu}\, d\Sigma_{\nu}, \quad i = 1,2. \tag{3.14}$$

The integrals for P^{μ}_{Ri} can be partially performed to get more compact expressions for them,[3]

$$P^{\mu}_{Ri} = -\frac{2}{3}q_i^2 \int_{-\infty}^{\infty} \dot{v}_i{}^2 v_i{}^{\mu} d\tau_i. \tag{3.15}$$

On the other hand

$$P^{\mu}_{R\,\text{Mix}} = -\int_{\Sigma} \lim_{\substack{\kappa_1\to\infty \\ \kappa_2\to\infty}} (T^{\mu\nu}_{\text{Mix}}\, d\Sigma_{\nu}). \tag{3.16}$$

The detailed shape of Σ is not important because of Gauss' theorem and Eq. (3.7). Thus we are allowed to choose to our convenience any surface that makes calculations simpler. We choose to work with a surface defined by

$$\kappa_1 = \text{const} \tag{3.17}$$

(Bhabha tube for particle q_1).[4]

For such a choice

$$d\Sigma_{\mu} = [(1 - \kappa_1(K_1\cdot\dot{v}_1))K_{1\mu} - v_{1\mu}]\kappa_1^2\, d\tau_1\, d\Omega_1^0. \tag{3.18}$$

It is found that

$$\begin{aligned}\lim_{\substack{\kappa_1\to\infty \\ \kappa_2\to\infty}} &(T^{\mu\nu}_{\text{Mix}}\, d\Sigma_{\nu}) \\ &= -(q_1q_2/2\pi)\,[\rho^3\{(v_2\cdot a_1)(a_2\cdot K_1) \\ &\quad - (v_1\cdot v_2)(a_2\cdot K_1)(a_1\cdot K_1)\} \\ &\quad + \rho^2\{(v_1\cdot a_2)(a_1\cdot K_1) - (a_1\cdot a_2)\}] \\ &\quad \times K_1{}^{\mu}\, d\tau_1\, d\Omega_1^0,\end{aligned} \tag{3.19}$$

where

$$\rho \equiv \kappa_1/\kappa_2, \tag{3.20}$$

and

$$a_i{}^{\mu} \equiv \dot{v}_i{}^{\mu}, \quad i = 1,2. \tag{3.21}$$

It is worth noting that all quantities related to q_i are evaluated at $\bar{\tau}_i$ defined as the retarded proper time on the ith world line of the observation point x^{μ}.

It follows that

$$P^{\mu}_{R\,\text{mix}} = 2q_1q_2 \int \frac{d\Omega_1^0}{4\pi} d\tau_1\, K_1{}^{\mu}\chi, \tag{3.22}$$

where

$$\begin{aligned}\chi = {} & \rho^3\{(v_2\cdot a_1)(a_2\cdot K_1) - (v_1\cdot v_2)(a_2\cdot K_1)(a_1\cdot K_1)\} \\ & + \rho^2\{(v_1\cdot a_2)(a_1\cdot K_1) - (a_1\cdot a_2)\}.\end{aligned} \tag{3.23}$$

IV. AN APPROXIMATION

The integrals involved in Eq. (3.22) can be performed only if the explicit expressions for both world lines are given. Even in such a case the calculations may become extremely cumbersome.

However, it is possible to go one step further if it happens that the time t_l that a light signal takes going back and forth from q_2 to q_1 is much smaller than some characteristic time t_c of the interaction. The condition $t_l \ll t_c$ is fulfilled when the particles are close to each other and their relative velocity

$$r = [1 - (v_1\cdot v_2)^{-2}]^{1/2} \tag{4.1}$$

is small compared with the speed of light.

Under the above conditions $v_2{}^{\mu}$ and $a_2{}^{\mu}$ are slowly varying functions of θ_1 and ϕ_1 over the interval t_l.

Then, as an approximation for evaluating the angular integrals we consider $v_2{}^{\mu}$ and $a_2{}^{\mu}$ to be independent of θ_1 and ϕ_1.

Thus

$$P^{\mu}_{R\,\text{Mix}} = 2q_1q_2 \int d\tau_1 W^{\mu}, \tag{4.2}$$

where

$$\begin{aligned}W^{\mu} = {} & -(a_1\cdot a_2)\, J_2{}^{\mu} + (v_1\cdot a_2)a_{1\nu}J_2{}^{\mu\nu} \\ & + (v_2\cdot a_1)a_{2\nu}\, J_3{}^{\mu\nu} - (v_1\cdot v_2)a_{1\nu}a_{2\sigma}\, J_3{}^{\mu\nu\sigma}\end{aligned} \tag{4.3}$$

and

$$J_n{}^{\alpha_1\alpha_2\cdots\alpha_s} = \int \frac{d\Omega_1^0}{4\pi}\rho^n K_1{}^{\alpha_1}K_1{}^{\alpha_2}\cdots K_1{}^{\alpha_s}. \tag{4.4}$$

The integrals (4.4) can be evaluated in the described approximation and it is found that (see the Appendix)

$$\begin{aligned}W^{\mu} = {} & [(v_1\cdot a_2)(v_2\cdot a_1)f_1(r) + (a_1\cdot a_2)\, g_1(r)]v_1{}^{\mu} \\ & + [(v_1\cdot a_2)(v_2\cdot a_1)\, f_2(r) + (a_1\cdot a_2)\, g_2(r)]v_2{}^{\mu} \\ & + (v_1\cdot a_2)h_1(r)a_1{}^{\mu} + (v_2\cdot a_1)h_2(r)a_2{}^{\mu},\end{aligned} \tag{4.5}$$

where f_i, g_i, and h_i $(i = 1,2)$ are functions of the relative velocity r whose explicit expressions are given in the Appendix.

V. EXAMPLES

A. The neutral particle

First, to illustrate the plausibility of the exact and the approximate formulas [(3.22) and (4.5), respectively] we will consider the radiation of interference produced by a neutral system composed by two charges $+q$ and $-q$ travel-

ing together over the same world line. Of course, the result should be such as to have null total radiated momentum.

In this case

$$v_1 = v_2 \equiv v, \quad a_1 = a_2 \equiv a, \quad \rho = 1 . \tag{5.1}$$

The exact expression (3.23) for χ now reduces to

$$\chi = -a^2 - (K\cdot a)^2 \tag{5.2}$$

and from Eq. (4.3)

$$W^\mu = -a^2 J_0^\mu - a_\nu a_\lambda J_0^{\mu\nu\lambda} . \tag{5.3}$$

Using Eqs. (A10) and (A12)

$$W^\mu = -a^2 v^\mu - a^2 I_0^\mu - a_\nu a_\lambda I_0^{\mu\nu\lambda} - a_\nu a_\lambda v^\mu I_0^{\nu\lambda} . \tag{5.4}$$

From (A14), (A18), (A19), and (A37),

$$I_0^\mu = 0 , \tag{5.5}$$

$$I_0^{\mu\nu} = C_0 h_1^{\mu\nu} , \tag{5.6}$$

$$I_0^{\mu\nu\lambda} = 0 . \tag{5.7}$$

Consequently

$$W^\mu = -(1 + C_0) a^2 v^\mu . \tag{5.8}$$

Equation (A37) implies

$$W^\mu = -\tfrac{2}{3} a^2 v^\mu . \tag{5.9}$$

Thus

$$P^\mu_{R\,\mathrm{Mix}} = \frac{4}{3} q^2 \int d\tau\, a^2 v^\mu , \tag{5.10}$$

which cancels exactly the radiation produced for both particles considered independently. Therefore the total radiated momentum is zero as expected.

On the other hand, the approximate equation (4.5) now reads

$$W^\mu = [g_1(0) + g_2(0)] a^2 v^\mu . \tag{5.11}$$

If the limits of $g_1(r)$ and $g_2(r)$ [see (A36)–(A38)] are adequately evaluated when r tends to zero then

$$W^\mu = -\tfrac{2}{3} a^2 v^\mu , \tag{5.12}$$

which again leads to the correct result.

It is worth mentioning that the approximation is exact for the described situation ($r = 0$, $z_1^\mu = z_2^\mu$) because its accuracy increases as r and the relative distance ($|z_1^\mu - z_2^\mu|$) get smaller.

In the same context it is very simple to regain the usual results for the radiation of an arbitrary charge q by considering it as the superposition of two charges αq and $(1-\alpha)q$.

B. The positronium

Let us now consider a system consisting of two particles $(+q, -q)$ each of them revolving around the other with constant angular velocity ω (the system can be thought of as a rotating dipole).

We will calculate the radiated momentum during one period T assuming that the motion is nearly circular with radius R.

The world lines for one period are described by

$$\begin{aligned} z_1^\mu &= (ct, R\cos\omega t, R\sin\omega t, 0) , \\ z_2^\mu &= (ct, -R\cos\omega t, -R\sin\omega t, 0) . \end{aligned} \tag{5.13}$$

Therefore

$$\begin{aligned} v_1^\mu &= \alpha(c, -R\omega\sin\omega t, R\omega\cos\omega t, 0) , \\ v_2^\mu &= \alpha(c, R\omega\sin\omega t, -R\omega\cos\omega t, 0) , \end{aligned} \tag{5.14}$$

and

$$\begin{aligned} a_1^\mu &= \alpha^2(0, -R\omega^2\cos\omega t, -R\omega^2\sin\omega t, 0) , \\ a_2^\mu &= \alpha^2(0, R\omega^2\cos\omega t, R\omega^2\sin\omega t, 0) , \end{aligned} \tag{5.15}$$

where

$$\alpha = \frac{dt}{d\tau_1} = \left[1 - \left(\frac{R\omega}{c}\right)^2\right]^{-1/2} . \tag{5.16}$$

The adopted approximation allows us to evaluate the dynamical quantities related to $+q$ at the same coordinate time t as those related to $-q$.

Then

$$(v_1\cdot a_2) = (v_2\cdot a_1) = 0 , \tag{5.17}$$

$$\epsilon \equiv (a_1\cdot a_2) = R^2\omega^4/[1 - (R\omega/c)^2]^2 , \tag{5.18}$$

$$\gamma \equiv (v_1\cdot v_2) = c^2[1 + (R\omega/c)^2]/[1 - (R\omega/c)^2] . \tag{5.19}$$

The momentum radiated by interference from $t = 0$ to $t = T$ is

$$\Delta P^\mu_{R\,\mathrm{Mix}} = -2q^2\left[\epsilon\left\{ g_1(r)\int_0^T v_1^\mu\, d\tau_1 + g_2(r)\int_0^T v_2^\mu\, d\tau_2\right\}\right] , \tag{5.20}$$

where use has been made of the fact that r is a constant [$g_1(r)$ and $g_2(r)$ are given in (A38)]. But

$$\int_0^T v_1^\mu\, d\tau_1 = \int_0^T v_2^\mu\, d\tau_2 = \frac{2\pi}{\omega} l^\mu , \tag{5.21}$$

where $l^\mu = (1,0,0,0)$.

Therefore

$$\Delta P^\mu_{R\,\mathrm{mix}} = [g_1(r) + g_2(r)] 4\pi q^2 \times\{R^2\omega^3/[1 - (R\omega/C)^2]^2\} l^\mu . \tag{5.22}$$

Analogously

$$\Delta P^\mu_{R1} = \Delta P^\mu_{R2} = (4\pi/3) q^2\{R^2\omega^3/[1 - (R\omega/C)^2]^2\} l^\mu . \tag{5.23}$$

Finally, from (5.22) and (5.23)

$$\Delta P^\mu_R = 4\pi q^2\{R^2\omega^3/[1 - (R\omega/C)^2]^2\} \times [g_1(r) + g_2(r) + \tfrac{2}{3}] l^\mu . \tag{5.24}$$

VI. CONCLUSIONS

As is well known, the radiated momentun of a single charge can be expressed as an integral over its history. On the other hand, for a system of two interacting pointlike charges, apart from the four-momenta associated to each particle an interference term appears. As we have shown, the exact expression for the momentum radiated by interference involves multiple integrals depending on the detailed shape of both world lines. To gain a deeper insight we have developed an easier to handle and more manageable approximate expression which—as in the single particle case—contains an integral over the past of only one of the particles.

Besides a trivial example (i.e., a neutral particle considered as the coalescence of two equal but opposite charges) to test the plausibility of the results, the total momentum radiated by a positroniumlike configuration was calculated as a function of the relative velocity r. As it can be seen, the r-dependent factor $g_1(r)+g_2(r)-\frac{2}{3}$ is concave and decreases monotonically from $\frac{4}{3}$ (for $r=0$) to $\frac{2}{3}$ (for $r=1$). (Nevertheless, the approximation is valid only for $r\leqslant 1$.)

ACKNOWLEDGMENTS

We thank Dr. J. Zanelli for careful critical reading of the manuscript. We are also indebted to Dr. R. Benguria for enlightening discussions and suggestions.

APPENDIX: EVALUATION OF THE INTEGRALS

We want to evaluate

$$J_n^{\alpha_1\alpha_2\cdots\alpha_s}\equiv\int\frac{d\Omega_1^0}{4\pi}\rho^n K_1^{\alpha_1}K_1^{\alpha_2}\cdots K_1^{\alpha_s}$$

$$n=0,1,2,\ldots,\quad s=1,2,\ldots, \tag{A1}$$

when $\kappa_1\to\infty$ and where

$$\rho=\kappa_1/\kappa_2\underset{\kappa_1\to\infty}{\to}1/(K_1\cdot v_2)\,. \tag{A2}$$

Let us introduce

$$K_{1\perp}^{\mu}=K_1^{\mu}-v_1^{\mu}\,. \tag{A3}$$

Trivially, $K_{1\perp}$ satisfies

$$K_{1\perp}\cdot v_1=0\,, \tag{A4}$$

$$K_{1\perp}\cdot K_1=-1\,, \tag{A5}$$

$$K_{1\perp}\cdot K_{1\perp}=-1\,, \tag{A6}$$

which shows that $K_{1\perp}$ is a spacelike vector. Define

$$I_n\equiv\int\frac{d\Omega_1^0}{4\pi}\rho^n \tag{A7}$$

and

$$I_n^{\alpha_1\alpha_2\cdots\alpha_s}=\int\frac{d\Omega_1^0}{4\pi}\rho^n K_{1\perp}^{\alpha_1}K_{1\perp}^{\alpha_2}\cdots K_{1\perp}^{\alpha_s}\,. \tag{A8}$$

It is easy to prove that

$$J_n=I_n\,, \tag{A9}$$

$$J_n^{\mu}=I_n^{\mu}+v_1^{\mu}I_n\,, \tag{A10}$$

$$J_n^{\mu\nu}=I_n^{\mu\nu}+v_1^{(\nu}I_n^{\mu)}+v_1^{\mu}v_1^{\nu}I_n\,, \tag{A11}$$

$$J_n^{\mu\nu\lambda}=I_n^{\mu\nu\lambda}+v_1^{(\mu}I_n^{\nu\lambda)}+v_1^{(\mu}v_1^{\nu}I_n^{\lambda)}+v_1^{\mu}v_1^{\nu}v_1^{\lambda}I_n\,. \tag{A12}$$

It will be shown that to evaluate $I_n^{\alpha_1\alpha_2\cdots\alpha_s}$ it is only necessary to know I_n. In fact, let us first consider I_n^{μ}.

Obviously, from its definition

$$I_n^{\mu}v_{1\mu}=0\,. \tag{A13}$$

Then I_n^{μ} is a linear combination of vectors orthogonal to $v_{1\mu}$.

On the other hand I_n^{μ} can only depend on v_2^{μ}. Thus we can write

$$I_n^{\mu}=\int\frac{d\Omega_1^0}{4\pi}K_{1\perp}^{\mu}\rho^n=A_n v_{2\perp}^{\mu}\,, \tag{A14}$$

where

$$v_{2\perp}^{\mu}=h_1^{\mu\nu}v_{2\nu}=v_2^{\mu}-\gamma v_1^{\mu} \tag{A15}$$

with

$$h_1^{\mu\nu}=\eta^{\mu\nu}-v_1^{\mu}v_1^{\nu} \tag{A16}$$

and

$$\gamma=(v_1\cdot v_2)\,. \tag{A17}$$

Analogously

$$I_n^{\alpha\beta}\equiv\int\frac{d\Omega_1^0}{4\pi}\rho^n K_{1\perp}^{\alpha}K_{1\perp}^{\beta}=B_n v_{2\perp}^{\alpha}v_{2\perp}^{\beta}+C_h h_1^{\alpha\beta}\,, \tag{A18}$$

$$I_n^{\alpha\beta\gamma}\equiv\int\frac{d\Omega_1^0}{4\pi}\rho^n K_{1\perp}^{\alpha}K_{1\perp}^{\beta}K_{1\perp}^{\gamma}$$

$$=D_n v_{2\perp}^{\alpha}v_{2\perp}^{\beta}v_{2\perp}^{\gamma}+E_n h_1^{(\alpha\beta}v_{2\perp}^{\gamma)}\,. \tag{A19}$$

By multiplying (A14) by v_{21}^{α} and remembering that the approximation allows us to bypass the integration sign it is found that

$$\gamma I_n-I_{n-1}=\gamma^2r^2A_n\,. \tag{A20}$$

Use has been made of the fact that

$$(v_{2\perp}\cdot v_{2\perp})=-\gamma^2r^2\,, \tag{A21}$$

$$(K_{1\perp}\cdot v_{2\perp})=1/\rho-\gamma\,. \tag{A22}$$

Equation (A20) indicates that A_n can be constructed from I_n and I_{n-1}.

By multiplying (A18) times $v_{2\perp}^{\alpha}v_{2\perp}^{\beta}$

$$\gamma^2I_n-2\gamma I_{n-1}+I_{n-2}=\gamma^2r^2(\gamma^2r^2B_n-C_n)\,. \tag{A23}$$

Contracting α and β in Eq. (A18)

$$I_n=\gamma^2r^2B_n-3C_n\,. \tag{A24}$$

The last two equations constitute a system that can be solved for B_n and C_n in terms of I_n, I_{n-1}, and I_{n-2}.

To obtain D_n and E_n it is possible to mimic the above procedure and the resulting system is

$$\gamma^3I_n-3\gamma^2I_{n-1}+3\gamma I_{n-2}-I_{n-3}$$

$$=\gamma^4r^4(\gamma^2r^2D_n-3E_n)\,, \tag{A25}$$

$$\gamma I_n-I_{n-1}=\gamma^2r^2(\gamma^2r^2D_n-5E_n)\,. \tag{A26}$$

Thus D_n and E_n can be expressed in terms of I_n, I_{n-1}, I_{n-2}, and I_{n-3}.

The key point is then to evaluate I_n. Its value should not depend on the frame of reference.

Let us choose the system attached to q_1. Then

$$v_1^{\mu}=(1,0,0,0)\,. \tag{A27}$$

The z direction can be chosen to coincidence with the spatial component of v_2^{μ}.

Therefore

$$v_2^{\mu}=(\gamma,0,0,\sqrt{\gamma^2-1})\,. \tag{A28}$$

Also

$$K_1^{\mu}=(1,\sin\theta_1\cos\phi_1,\sin\theta_1\sin\phi_1,\cos\theta_1)\,. \tag{A29}$$

In the limit $\kappa_1\to\infty$,

$$\rho = 1/\gamma(1 - rx_1) \tag{A30}$$

with

$$x_1 = \cos\theta_1 . \tag{A31}$$

But

$$d\Omega_1^0 = -d\phi_1\, dx_1 . \tag{A32}$$

Consequently, after performing the angular integration

$$I_n = \frac{1}{2}\int_{-1}^{1} dx \frac{1}{\gamma^n(1-rx)^n} . \tag{A33}$$

This integral can be easily done, giving

$$I_1 = (1/2\gamma r)\ln[(1+r)/(1-r)] , \tag{A34}$$

$$I_n = [1/2\gamma^n r(1-n)][(1+r)^{1-n} - (1-r)^{1-n}] , \quad n \neq 1 . \tag{A35}$$

It is useful to exhibit the explicit expression of I_n for the various values of n needed in the calculations

$$I_{-2} = \tfrac{1}{3}(4\gamma^2 - 1), \quad I_{-1} = \gamma, \quad I_0 = 1 , \qquad I_1 = (1/2\gamma_r)\ln[(1+r)/(1-r)], \quad I_2 = 1, \quad I_3 = \gamma . \tag{A36}$$

Solving Eqs. (A23), (A24) and (A25), (A26) for B_n, C_n and D_n, E_n, respectively, and using the above results, the following list can be elaborated:

$$A_0 = 0 ,$$

$$A_1 = \frac{1-r^2}{r^2}\left[\frac{1}{2r}\ln\frac{1+r}{1-r} - 1\right] ,$$

$$A_2 = \frac{1}{\gamma r^2}\left[1 - \frac{1-r^2}{2r}\ln\frac{1+r}{1-r}\right] ,$$

$$A_3 = 1 ,$$

$$B_1 = \frac{1-r^2}{2\gamma r^4}\left[\frac{3-r^2}{2r}\ln\frac{1+r}{1-r} - 3\right] ,$$

$$B_2 = \frac{1-r^2}{r^2}\left[\frac{3(1-r^2)}{r^2}\left(1 - \frac{1}{2r}\ln\frac{1+r}{1-r}\right) + 1\right] ,$$

$$B_3 = \frac{1}{\gamma r^2}\left[\frac{3(1-r^2)}{2r^2}\left(\frac{1-r^2}{2r}\ln\frac{1+r}{1-r} - 1\right) + 1\right] , \tag{A37}$$

$$C_0 = -\tfrac{1}{3} ,$$

$$C_1 = -\tfrac{1}{2}A_2 ,$$

$$C_2 = -A_1 ,$$

$$C_3 = -\tfrac{1}{2}A_2 ,$$

$$E_1 = -\tfrac{1}{6}B_2 ,$$

$$E_2 = -B_1 ,$$

$$E_3 = -\tfrac{1}{2}B_2 ,$$

$$D_n = [(1-r^2)/r^2][A_n + 5E_n] .$$

Furthermore, the functions (f,g,h) appearing in Eq. (4.5) are given by

$$\begin{aligned} f_1(r) &= -\gamma B_2 + A_2 + 3\gamma^2 B_3 \\ &\quad - C_3 - 3\gamma A_3 + I_3 - \gamma^3 D_3 + \gamma E_3 , \\ f_2(r) &= B_2 - 2\gamma B_3 + A_3 + \gamma^2 D_3 , \\ g_1(r) &= \gamma A_2 - I_2 + \gamma^2 E_3 - \gamma C_3 , \\ g_2(r) &= -A_2 - \gamma E_3 , \\ h_1(r) &= C_2 + \gamma^2 E_3 - \gamma C_3 , \\ h_2(r) &= C_3 - \gamma E_3 , \end{aligned} \tag{A38}$$

and as can be seen they are combinations of A_n, B_n, C_n, D_n, E_n, and I_n; so their explicit form as functions of r flows directly when expressions (A37) are replaced in (A38).

[1] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).
[2] C. Teitelboim, Phys. Rev. D **1**, 1572 (1970).
[3] R. Tabensky and D. Villarroel, J. Math. Phys. **16**, 1380 (1975).
[4] H. J. Bhabha, Proc. R. Soc. London Ser. A **172**, 384 (1939).

Four-momentum and angular momentum in classical electrodynamics

Abraham Lozada and P. L. Torres
Departamento de Física, Facultad de Ciencias, Universidad Central de Venezula, A. P. 20513, Caracas 1020-A, Venezuela

(Received 17 September 1987; accepted for publication 13 January 1988)

The total four-momentum P^μ and the total angular momentum tensor $M^{\lambda\mu}$ for a classical point charged particle and its field are considered. The standard definition of P^μ and $M^{\lambda\mu}$ is used as the integrals of their respective densities over a spacelike hyperplane orthogonal to the four-velocity of the particle. It is shown that the integral defining P^μ exists and is a conserved quantity, but that the integral defining $M^{\lambda\mu}$ does not exist. The last result is due to the asymptotic behavior of the Coulomb field at spatial infinity.

I. INTRODUCTION

For a point charged particle in classical electrodynamics, given an arbitrary energy momentum tensor $\Theta_A^{\mu\nu}$, and its corresponding angular momentum tensor density $M_A^{\lambda\mu\nu}$,

$$M_A^{\lambda\mu\nu} \equiv x^\lambda \Theta_A^{\mu\nu} - x^\mu \Theta_A^{\lambda\nu}, \tag{1.1}$$

the standard definitions for the four-momentum P_A^μ and the angular momentum tensor $M_A^{\lambda\mu}$ that have been considered in the literature[1–8] (for a review on this and related matters, see Ref. 8) are

$$P_A^\mu(\tau) = \int_\sigma \Theta_A^{\mu\nu}\, d\sigma_\nu, \tag{1.2}$$

$$M_A^{\lambda\mu}(\tau) = \int_\sigma M_A^{\lambda\mu\nu}\, d\sigma_\nu, \tag{1.3}$$

where the integration is performed on the spacelike hyperplane σ, which cuts the electron world line (EWL) orthogonally at the point $z(\tau)$, with τ being the proper time at which these quantities are evaluated. Here, $d\sigma_\nu = -v_\nu(\tau) d^3\sigma$.

In this paper we consider some facts that have been overlooked in the literature, and that are related to the infinite range of the electromagnetic field. Specifically, we study the integrals (1.2) and (1.3) corresponding to the total four-momentum P^μ and the total angular momentum tensor $M^{\lambda\mu}$, for a classical point charged particle in an external electromagnetic field. We obtain the following.

(i) The integral (1.3) for the total angular momentum density $M^{\lambda\mu\nu}$ is not absolutely convergent. It happens that the value of the integral depends on the way that the infinity is reached. In other words, the improper Riemann integral (1.3) does *not* exist for $M^{\lambda\mu\nu}$.

This fact is due to the behavior of the Coulomb field at spatial infinity.

(ii) The integral (1.2) for the total energy momentum tensor $\Theta^{\mu\nu}$ is absolutely convergent. That is, the improper Reimann integral (1.2) exists for $\Theta^{\mu\nu}$ and hence its value does not depend on the way that the infinity is reached.

(iii) We show explicitly that P^μ is a conserved quantity. That is, if a conservation law in differential form is assumed for $\Theta^{\mu\nu}$ (i.e., $\partial_\nu \Theta^{\mu\nu} = 0$, everywhere in space-time), it follows that the tensor $\Theta^{\mu\nu}$ has a good behavior at spatial infinity in order to define a globally conserved quantity by (1.2).

(iv) Also, through (1.3) we evaluate $M_{\text{mix}}^{\lambda\mu}$, the part of the total angular momentum tensor associated with the interaction of the particles and the external field. This integral is absolutely convergent due to the asymptotic behavior assumed for the external field.

The infinities associated with P^μ and $M^{\lambda\mu}$ (because of the divergences of $\Theta^{\mu\nu}$ at the EWL) will be dealt with by the standard[1,2,8,9] mass renormalization procedure. We want to stress that the nonexistence of the integral for $M^{\lambda\mu\nu}$ pointed out in (i) is *not* related to the divergences of the theory at the EWL, but rather to the behavior of the Coulomb field at spatial infinity.

With respect to the notation, the metric tensor g will have signature $+2$, and the speed of the light is taken as 1. When it will be convenient, scripts on vectors, and tensors, will be omitted and scalar products will be indicated by a dot. Parentheses $(\cdot,\cdot)$ or brackets $[\cdot,\cdot]$ will denote symmetrization or antisymmetrization, respectively, of the enclosed indices (without a factor of one-half). The EWL is $z(\tau)$, where τ is the proper time, and $v(\tau) \equiv v$ ($v^2 = -1$) and $a(\tau) \equiv a$ ($v \cdot a = 0$) are the four-velocity and acceleration, respectively. The components of the total electromagnetic energy tensor for a point charged particle and an external electromagnetic field are

$$\Theta_{\text{elm}}^{\mu\nu} = (1/4\pi)(F^{\mu\alpha}F_\alpha^\nu - \tfrac{1}{4} g^{\mu\nu} F^{\alpha\beta} F_{\alpha\beta}), \tag{1.4}$$

where

$$F^{\mu\nu} = F_{\text{ret}}^{\mu\nu} + F_{\text{ext}}^{\mu\nu}. \tag{1.5}$$

The electromagnetic field tensor $F_{\text{ret}}^{\mu\nu}$ is the one corresponding to the retarded Lienard–Wiechert potential. The nonsingular external electromagnetic field $F_{\text{ext}}^{\mu\nu}$ satisfies Maxwell's equations for vacuum and vanishes at spatial infinity and at the remote past.[10] Specifically, we shall suppose that $F_{\text{ext}}^{\mu\nu} = 0$, $\forall \tau \in (-\infty, \tau_0]$, where τ_0 is an arbitrary proper time (as small as we want), which means that the particle is free $\forall \tau \in (-\infty, \tau_0]$. This is a very mild simplifying assumption that can even be relaxed.

Corresponding to the superposition shown in (1.5) we obtain (in obvious notation) using (1.4),

$$\Theta_{\text{elm}} = \Theta_{\text{ret}} + \Theta_{\text{mix}} + \Theta_{\text{ext}}. \tag{1.6}$$

Retarded coordinates will be used here (see, e.g., Ref. 8). Then for any space-time point x, we define

 0022-2488/88/061361-07$02.50

$$R \equiv x - z(\tau), \quad R^2 = 0 (R^0 > 0),$$
$$\rho \equiv -v \cdot R, \quad u \equiv (R/\rho) - v.$$

In these coordinates,

$$\Theta_{\text{ret}} = \frac{e^2}{4\pi}\left(\tfrac{1}{2}g + vv - uu\right)\frac{1}{\rho^4} + \frac{e^2}{4\pi}\left(a - u\, a\cdot u, \frac{R}{\rho}\right)\frac{1}{\rho^3} + \frac{e^2}{4\pi}[a^2 - (a\cdot u)^2]\frac{RR}{\rho^4}. \tag{1.7}$$

The material properties of the particle will be characterized by the standard bare energy momentum tensor $K_0^{\mu\nu}$, which is given by

$$K_0^{\mu\nu} = \int m_0 v^\mu(\tau) v^\nu(\tau) \delta[x - z(\tau)]\, d\tau, \tag{1.8}$$

where m_0 is the electron bare mass.

Hence, for $K_0^{\mu\nu}$ and its associated angular momentum tensor density $M_0^{\lambda\mu\nu}$, we obtain its corresponding bare four-momentum P_0^μ and bare angular momentum $M_0^{\lambda\mu}$:

$$P_0^\mu(\tau) \equiv P_0^\mu \equiv \int_\sigma K_0^{\mu\nu}\, d\sigma_\nu = m_0 v^\mu(\tau), \tag{1.9}$$

$$M_0^{\lambda\mu}(\tau) \equiv M_0^{\lambda\mu} \equiv \int_\sigma M_0^{\lambda\mu\nu}\, d\sigma_\nu = [z^\lambda(\tau), m_0 v^\mu(\tau)]. \tag{1.10}$$

The evaluation of P_{mix}^μ has been considered by Tabensky,[10] who obtains

$$P_{\text{mix}}^\mu(\tau) \equiv \int_\sigma \Theta_{\text{mix}}^{\mu\nu}\, d\sigma_\nu = -e\int_{-\infty}^{\tau} F_{\text{ext}}^{\mu\nu}(z(\tau'))v_\nu(\tau')\, d\tau'. \tag{1.11}$$

Following a procedure similar to Tabensky's, we obtain (see Appendix A) that

$$M_{\text{mix}}^{\lambda\mu}(\tau) \equiv \int_\sigma M_{\text{mix}}^{\lambda\mu\nu}\, d\sigma_\nu = -e\int_{-\infty}^{\tau} [z^\lambda, F_{\text{ext}}^{\mu\nu}(z(\tau'))v_\nu]\, d\tau'. \tag{1.12}$$

Since $\partial_\nu \Theta_{\text{ext}}^{\mu\nu} = 0$, in all space-time, and $\Theta_{\text{ext}}^{\mu\nu}$ has nice properties at spatial infinity, P_{ext}^μ and $M_{\text{ext}}^{\lambda\mu}$ are independently conserved quantities. Hence in the following we shall exclude the P_{ext}^μ and $M_{\text{ext}}^{\lambda\mu}$ parts in the total four-momentum P^μ and in the total angular momentum tensor $M^{\lambda\mu}$, respectively. That is, $P^\mu = P_0^\mu + P_{\text{ret}}^\mu + P_{\text{mix}}^\mu$ and $M^{\lambda\mu} = M_0^{\lambda\mu} + M_{\text{ret}}^{\lambda\mu} + M_{\text{mix}}^{\lambda\mu}$.

The plan of this paper is as follows. In Sec. II, we show that the integral (1.3) for $M^{\lambda\mu\nu}$ is not absolutely convergent but that the integral (1.2) for $\Theta^{\mu\nu}$ is. In Sec. III, we show that P^μ is a conserved quantity and also present a way for obtaining a conserved quantity from $M^{\lambda\mu\nu}$. In Sec. IV, we discuss the obtained results. In Appendix A, we prove (1.12) and, in Appendix B, some relevant integrals are evaluated.

II. EXISTENCE OF THE INTEGRALS DEFINING P^μ AND $M^{\lambda\mu}$

In this section we want to show that even though P^μ as defined by (1.2) exists, the integral (1.3) defining $M^{\lambda\mu}$ *does not exist.*

Let us start by remembering some facts about multiple improper Riemann integrals with continuous integrand in unbounded regions.[11,12] Let σ be the unbounded integration region, and $\sigma_n \subset \sigma$, a sequence of bounded regions such that $\sigma_n \to \sigma$ (cf. Refs. 11 and 12). If for any such sequence σ_n, the limit when $n \to \infty$ of the integrals evaluated on σ_n exists (is finite) and is *independent* of the choice of $\{\sigma_n\}$, then the limit is said to be the improper integral on σ. We also have the following theorems[11,12]: (a) the integral on σ exists if and only if the integral of the absolute value of the integrand exists (i.e., if and only if the integral is absolutely convergent); and (b) the integral of a positive function exists if and only if it exists for an arbitrary sequence $\{\sigma_n\}$ of bounded regions.

In order to study the existence of the integrals (1.2) and (1.3) for $\Theta^{\mu\nu}$ and $M^{\lambda\mu\nu}$, respectively, we shall consider, instead of the hyperplane σ, the unbounded hypersurface $\sigma^{(0)} \subset \sigma$, which coincides with σ except for a "hole around the EWL" defined by the intersection between the light cone with apex at τ' ($\tau' < \tau$) and σ. We do this to avoid the infinites of the fields at the EWL (which can be dealt with later by the usual renormalization procedure) and to emphasize that the existence of the integrals that we are discussing is related only to the asymptotic behavior of the field at spatial infinity.

Since the integrals defining P_{mix}^μ and $M_{\text{mix}}^{\lambda\mu}$ are absolutely convergent (see Appendix A) we shall consider only the following integrals:

$$P^\mu[\sigma^{(0)}] \equiv \int_{\sigma^{(0)}} \Theta_{\text{ret}}^{\mu\nu}\, d\sigma_\nu, \tag{2.1}$$

$$M^{\lambda\mu}[\sigma^{(0)}] \equiv \int_{\sigma^{(0)}} M_{\text{ret}}^{\lambda\mu\nu}\, d\sigma_\nu. \tag{2.2}$$

First, we want to show that $M^{\lambda\mu}[\sigma^{(0)}]$ does not exist. To do this we split our proof into the following parts (i) and (ii).

(i) *Case of a free particle:* Here we shall show directly that the integral

$$\int_{\sigma^{(0)}} |M_{\text{ret}}^{\lambda\mu\nu} v_\nu|\, d^3\sigma, \tag{2.3}$$

diverges. Without losing generality, we suppose that the particle is at rest. In this case,

$$M_{\text{ret}}^{klv} v_\nu = 0, \quad k,l = 1,2,3, \tag{2.4}$$

$$M_{\text{ret}}^{0kv} v_\nu = (e^2/8\pi)[z^k(\tau) + \rho u^k]/\rho^4, \quad k = 1,2,3. \tag{2.5}$$

We obtain that, for $k = 1,2,3$,

$$\int_{\sigma^{(0)}} |M_{\text{ret}}^{0kv} v_\nu|\, d^3\sigma = \frac{e^2}{8\pi} \lim_{\rho'' \to \infty} \int_{\rho'}^{\rho''} d\rho \int d\Omega \left| \frac{u^k}{\rho} + \frac{z^k(\tau)}{\rho^2} \right|, \tag{2.6}$$

where $\rho' = \tau - \tau'$, is the value of the ρ coordinate corresponding to the intersection between the light cone with apex at τ' and σ. Hence from (2.6),

$$\int_{\sigma^{(0)}} |M_{\text{ret}}^{0kv} v_\nu|\, d^3\sigma \geqslant \frac{e^2}{8\pi} \lim_{\rho'' \to \infty} \left\{ \int_{\rho'}^{\rho''} \frac{d\rho}{\rho} \int d\Omega |u^k| - \int_{\rho'}^{\rho''} \frac{d\rho}{\rho^2} \int d\Omega |z^k(\tau)| \right\} = +\infty. \tag{2.7}$$

(ii) *Case of a particle in an external field* $F^{\mu\nu}_{\text{ext}}$: Here we shall show that different choices for the sequence of bounded regions $\{\sigma_n\}$, such that $\sigma_n \to \sigma^{(0)}$, throw different results, and hence the integral defining $M^{\lambda\mu}[\sigma^{(0)}]$ does not exist. In fact we shall show that

$$\lim_{E\to\infty} M^{\lambda\mu}(\sigma') \equiv \lim_{E\to\infty} \int_{\sigma'} M^{\lambda\mu\nu}_{\text{ret}}\, d\sigma_\nu, \tag{2.8}$$

and

$$\lim_{\tau''\to-\infty} M^{\lambda\mu}(\bar\sigma) \equiv \lim_{\tau''\to-\infty} \int_{\bar\sigma} M^{\lambda\mu\nu}_{\text{ret}}\, d\sigma_\nu, \tag{2.9}$$

where σ' and $\bar\sigma$ are defined in Appendix B, give different results (notice that under the respective limits $\sigma' \to \sigma^{(0)}$ and $\bar\sigma \to \sigma^{(0)}$).

Using (B14), (B16), and the fact that the particle is free for $\tau \leqslant \tau_0$, we obtain

$$\lim_{E\to\infty} M^{\lambda\mu}(\sigma') = [z^\lambda(\tau'), f^\mu_\tau(\tau')] + \frac{2}{3}e^2\int_{\tau_0}^{\tau'} [v^\lambda(\tau), a^\mu(\tau)]\,d\tau + \frac{2}{3}e^2\int_{\tau_0}^{\tau'} [z^\lambda(\tau), v^\mu(\tau)]\,a^2(\tau)\,d\tau. \tag{2.10}$$

Also, using (B6) and (B7), we obtain that

$$\lim_{\tau''\to-\infty} M^{\lambda\mu}(\bar\sigma) = \lim_{E\to\infty} M^{\lambda\mu}(\sigma') + (e^2/6)\,\frac{[v^\lambda(\tau_0), v^\mu(\tau)]}{[v(\tau_0)\cdot v(\tau)]}. \tag{2.11}$$

We see that the last term of the right-hand side of (2.11) being a function of τ is not identically zero $\forall \tau > \tau_0$ (except for the case of the free particle).

Now we want to show that the integral defining $P^\mu[\sigma^{(0)}]$ in (2.1) is absolutely convergent, i.e., $P^\mu[\sigma^{(0)}]$ exists. For that, it is sufficient to consider $\sigma^{(0)}$ for $\tau' \leqslant \tau_0$. In this case, we have that

$$\Theta^{\mu\nu}_{\text{ret}} v_\nu(\tau) = (e^2/4\pi\rho^4)\,[v^\mu(\tau)/2 + v(\tau_0)\cdot v(\tau)\, v^\mu(\tau_0) - u(\tau_0)\cdot v(\tau)\, u^\mu(\tau_0)], \tag{2.12}$$

and that the invariant area element on $\sigma^{(0)}$ in retarded coordinates is given by

$$d^3\sigma = v^\nu(\tau)\,d\sigma_\nu = -\rho^2\, d\rho\, d\Omega/[v(\tau_0)\cdot v(\tau)], \tag{2.13}$$

where $d\Omega$ is the solid-angle element for the inertial frame with time axis $v^\nu(\tau_0)$; see, e.g., (A16) in Ref. 8. The expression (2.13) is easily evaluated following the standard procedure (cf. Refs. 8, 10, and 13) and the fact that the equation of the hyperplane σ in retarded coordinates is given by (B12) for $\tau_R \leqslant \tau_0$, with E substituted by ρ. Hence it follows that

$$\int_{\sigma^{(0)}} |\Theta^{\mu\nu}_{\text{ret}}\, v_\nu|\, d^3\sigma = \lim_{\tau''\to-\infty} \int d\Omega\, |A^\mu| \int_{\rho_0(\tau_0)}^{\rho_0(\tau'')} \frac{d\rho}{\rho^2}, \tag{2.14}$$

where $\rho_0(\tau')$ is defined in (A1) and $|A^\mu| \equiv \rho^4 |\Theta^{\mu\nu}_{\text{ret}}\, v_\nu| / |v(\tau_0)\cdot v(\tau)|$, is independent of ρ. It follows directly that (2.14) is *finite*, hence the integral defining $P^\mu[\sigma^{(0)}]$ exists. Then, from (B5), (B7), (B13), and (B15) we obtain

$$P^\mu[\sigma^{(0)}] = \lim_{E\to\infty} P^\mu(\sigma') = \lim_{\tau''\to-\infty} P^\mu(\bar\sigma) = f^\mu_\tau(\tau') + \frac{2}{3}e^2\int_{\tau_0}^{\tau} a^2(\tau) v^\mu(\tau)\,d\tau. \tag{2.15}$$

It should be mentioned that using (2.13) and a procedure similar to the one used in (i), it can be shown directly that the integral $\int_{\sigma^{(0)}} |M^{\lambda\mu\nu} v_\nu|\, d^3\sigma$ for the general case of a particle in an external field diverges [this will constitute an alternative proof from the one given in (ii)].

Finally, using the standard mass renormalization procedure, we discuss the evaluation of the quantities

$$P^\mu[\sigma^{(0)}], \quad \lim_{E\to\infty} M^{\lambda\mu}(\sigma'), \quad \lim_{\tau''\to-\infty} M^{\lambda\mu}(\bar\sigma),$$

defined on $\sigma^{(0)}$, when $\sigma^{(0)} \to \sigma$.

Since

$$\lim_{\tau'\to\tau} \{f^\mu_\tau(\tau') - e^2 v^\mu(\tau)/2(\tau-\tau')\} = -\tfrac{2}{3}e^2 a^\mu(\tau), \tag{2.16}$$

we perform the renormalization with the bare four-momentum $P^\mu_0 = m_0 v^\mu(\tau)$, obtaining the usual result[2,5,8]:

$$\lim_{\tau'\to\tau} \{P^\mu_0(\tau) + P^\mu[\sigma^{(0)}]\} = m v^\mu(\tau) - \frac{2}{3}e^2 a^\mu(\tau) + \frac{2}{3}e^2\int_{\tau_0}^{\tau} a^2(\tau') v^\mu(\tau')\,d\tau', \tag{2.17}$$

where m is the renormalized mass. Then we get for the total four-momentum P^μ, the usual result[2,5,8]:

$$P^\mu = m v^\mu(\tau) - \frac{2}{3}e^2 a^\mu(\tau) + \frac{2}{3}e^2\int_{\tau_0}^{\tau} a^2(\tau') v^\mu(t')\,d\tau' - e\int_{\tau_0}^{\tau} F^{\mu\nu}_{\text{ext}}(\tau') v_\nu(\tau')\,d\tau'. \tag{2.18}$$

Similarly,

$$\lim_{\tau'\to\tau} \{[z^\lambda(\tau'), f^\mu_\tau(\tau')] - e^2[z^\lambda(\tau), v^\mu(\tau)]/2(\tau-\tau')\} = -\tfrac{2}{3}e^2[z^\lambda(\tau), a^\mu(\tau)]. \tag{2.19}$$

Then, renormalizing the quantities $\lim_{E\to\infty} M^{\lambda\mu}(\sigma')$ and $\lim_{\tau''\to-\infty} M^{\lambda\mu}(\bar\sigma)$ [see (2.10) and (2.11)], with the bare angular momentum $M^{\lambda\mu}_0(\tau) = [z^\lambda(\tau), P^\mu_0(\tau)]$, we obtain

$$\lim_{\tau'\to\tau}\left[M^{\lambda\mu}_0(\tau) + \lim_{E\to\infty} M^{\lambda\mu}(\sigma')\right] \equiv M^{\lambda\mu}_{\text{ren}} = [z^\lambda(\tau), m v^\mu(\tau)] - \frac{2}{3}e^2[z^\lambda(\tau), a^\mu(\tau)] + \frac{2}{3}e^2\int_{\tau_0}^{\tau} [v^\lambda(\tau'), a^\mu(\tau')]\,d\tau' + \frac{2}{3}e^2\int_{\tau_0}^{\tau} [z^\lambda(\tau'), v^\mu(\tau')]\,a^2(\tau')\,d\tau', \tag{2.20}$$

and

$$\lim_{\tau'\to\tau}\left[M_0^{\lambda\mu}(\tau)+\lim_{\tau''\to-\infty}M^{\lambda\mu}(\bar{\sigma})\right]$$
$$=M_{\rm ren}^{\lambda\mu}+\frac{e^2}{6}\frac{[v^\lambda(\tau_0),v^\mu(\tau)]}{[v(\tau_0)\cdot v(\tau)]}. \tag{2.21}$$

We have shown that the integral (1.3), defining the total angular momentum $M^{\lambda\mu}$ does not exist [see (i) and (ii)]. However, we can define the following quantities:

$$M^{\lambda\mu}(\sigma'\to\sigma)\equiv M_{\rm ren}^{\lambda\mu}+M_{\rm mix}^{\lambda\mu}, \tag{2.22}$$

$$M^{\lambda\mu}(\bar{\sigma}\to\sigma)\equiv M_{\rm ren}^{\lambda\mu}+M_{\rm mix}^{\lambda\mu}$$
$$+(e^2/6)[v^\lambda(\tau_0),v^\mu(\tau)]/[v(\tau_0)\cdot v(\tau)], \tag{2.23}$$

where $M_{\rm ren}^{\lambda\mu}$ is defined in (2.20). The quantity $M^{\lambda\mu}(\sigma'\to\sigma)$ is the one found in the literature[4,5,8] (the term $M_{\rm mix}^{\lambda\mu}$ being understood) under the name of the total angular momentum.

III. CONSERVATION LAWS

Let us define the tensors

$$\Theta^{\mu\nu}\equiv K_0^{\mu\nu}+\Theta_{\rm ret}^{\mu\nu}+\Theta_{\rm mix}^{\mu\nu}, \tag{3.1}$$

$$M^{\lambda\mu\nu}\equiv M_0^{\lambda\mu\nu}+M_{\rm ret}^{\lambda\mu\nu}+M_{\rm mix}^{\lambda\mu\nu}. \tag{3.2}$$

The conservation laws for the energy and the angular momentum, expressed in differential form, are

$$\partial_\nu\Theta^{\mu\nu}=0,\quad \partial_\nu M^{\lambda\mu\nu}=0, \tag{3.3}$$

everywhere in space-time [at the EWL, (3.3) is an assumption about the cancellation of the infinities].

As is well-known,[3,14,15] from (3.3), the Gauss theorem, and the vanishing of certain integrals at spatial infinity, the conservation laws in integral form for the total four-momentum P^μ and the total angular momentum $M^{\lambda\mu}$ are obtained whenever the quantities P^μ and $M^{\lambda\mu}$ defined by (1.1) and (1.2) exist. As we can see, there are, in principle, two problems that must be clarified: first, the existence of P^μ and $M^{\lambda\mu}$, and, second, if they exist, their conservation.

In the following, we shall need the two bounded space-time regions Ω' and $\bar{\Omega}$, seen in Fig. 1. Let σ_1 and σ_2 be two hyperplanes that cut the EWL orthogonally at $z(\tau_1)$ and $z(\tau_2)$, respectively. Let Σ be the bounded part of the Bhabha tube of radius E given by the intersection between this Bhabha tube and the hyperplanes σ_1 and σ_2. Let C be the bounded part of the light cone with the apex at τ' ($\tau'<\tau_0$) given by the intersection between this light cone and the hyperplanes σ_1 and σ_2. Then, Ω' is determined by the hypersurfaces Σ, σ_1', and σ_2', where σ_i', $i=1,2$, is the bounded part of σ_i given by the intersection between Σ and σ_i. Finally, $\bar{\Omega}$ is determined by the hypersurfaces C, $\bar{\sigma}_1$, and $\bar{\sigma}_2$, where $\bar{\sigma}_i$, $i=1,2$, is the bounded part of σ_i given by the intersection between C and σ_i.

Now, knowing that P^μ exists (see, Sec. II), let us show that it is a conserved quantity. Applying the Gauss theorem and (3.3) in $\bar{\Omega}$, we obtain

$$\int_{\bar{\sigma}_1}\Theta^{\mu\nu}\,d\sigma_\nu-\int_{\bar{\sigma}_2}\Theta^{\mu\nu}\,d\sigma_\nu=f_{\tau_2}^\mu(\tau')-f_{\tau_1}^\mu(\tau'), \tag{3.4}$$

where $f_\tau^\mu(\tau')$ is defined in (B7). Hence, under the limit $\tau'\to-\infty$, we have

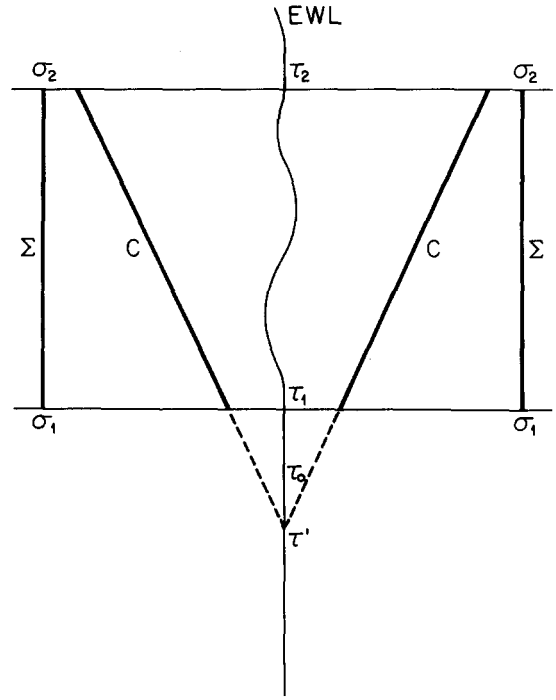


FIG. 1. Space-time regions used to test conservations laws in integral form.

$$P^\mu(\tau_1)=P^\mu(\tau_2), \tag{3.5}$$

i.e., the total four-momentum is conserved. Let us note that the result (3.5) is also obtained if the region Ω' is used (as it should be).

Since $M^{\lambda\mu}$ defined by (1.3) does not exist (see Sec. II), it is senseless to consider its conservation. However, it makes sense to ask if quantities (obtained through a prescription on the way that the spatial infinity is reached) like $M^{\lambda\mu}(\sigma'\to\sigma)$ and $M^{\lambda\mu}(\bar{\sigma}\to\sigma)$, defined in (2.22) and (2.23), are conserved.

Then, let us show that $M^{\lambda\mu}(\sigma'\to\sigma)$ is a conserved quantity. Applying the Gauss theorem and (3.3) on Ω', we obtain

$$\int_{\sigma_1'}M^{\lambda\mu\nu}\,d\sigma_\nu-\int_{\sigma_2'}M^{\lambda\mu\nu}\,d\sigma_\nu=\int_\Sigma M^{\lambda\mu\nu}\,d\Sigma_\nu. \tag{3.6}$$

Under the limit $E\to\infty$, we obtain from (3.6), (2.22), and (B16) that

$$M^{\lambda\mu}(\sigma_1'\to\sigma_1)=M^{\lambda\mu}(\sigma_2'\to\sigma_2). \tag{3.7}$$

That is, $M^{\lambda\mu}(\sigma'\to\sigma)$ is a conserved quantity.

From (2.18) and (2.22), it follows that

$$\dot{M}^{\lambda\mu}(\sigma'\to\sigma)=[z^\lambda,\dot{P}^\mu], \tag{3.8}$$

which shows that the conservations of P^μ and $M^{\lambda\mu}(\sigma'\to\sigma)$ are compatible.[5]

Finally, let us show that $M^{\lambda\mu}(\bar{\sigma}\to\sigma)$ is not a conserved quantity. Applying the Gauss theorem and (3.3) on $\bar{\Omega}$, we obtain

$$\int_{\bar{\sigma}_1}M^{\lambda\mu\nu}\sigma_\nu-\int_{\bar{\sigma}_2}M^{\lambda\mu\nu}\,d\sigma_\nu$$
$$=[z^\lambda(\tau'),f_{\tau_2}^\mu(\tau')]-[z^\lambda(\tau'),f_{\tau_1}^\mu(\tau')]. \tag{3.9}$$

Under the limit $\tau' \to -\infty$, we obtain from (3.9), (2.23), and (B7) that

$$M^{\lambda\mu}(\bar{\sigma}_1 \to \sigma_1) - M^{\lambda\mu}(\bar{\sigma}_2 \to \sigma_2) = \frac{e^2}{6}\left\{\frac{[v^\lambda(\tau_0), v^\mu(\tau_1)]}{v(\tau_0)\cdot v(\tau_1)} - \frac{[v^\lambda(\tau_0), v^\mu(\tau_2)]}{v(\tau_0)\cdot v(\tau_2)}\right\}. \tag{3.10}$$

From (3.10) we see that except for a free particle, $M^{\lambda\mu}(\bar{\sigma} \to \sigma)$ is not a conserved quantity [if the particle is free, $M^{\lambda\mu}(\sigma' \to \sigma)$ and $M^{\lambda\mu}(\bar{\sigma} \to \sigma)$ are equal because it happens that in this case the sequence of bounded regions σ' and $\bar{\sigma}$ coincide].

IV. DISCUSSION

In Sec. III we showed that the integral (1.3) defining the angular momentum tensor for a point particle does not converge absolutely, i.e., that *the angular momentum tensor does not exist*. One can ask what is the situation for the case of an extended charged body. This question has been considered in Ref. 16 (p. 398), where it is mentioned that the integral defining the angular momentum tensor is not, generally, absolutely convergent, expecting, however, that in important special cases it will converge. Nevertheless, our result shows that what it is reasonable to expect, is that the integral defining the angular momentum tensor for *any* extended body is not absolutely convergent, because asymptotically the behavior of the field of a body is like the corresponding one to a point particle.[17]

We believe that we have made it clear enough that the nonexistence of $M^{\lambda\mu}$ that we have discussed in this paper has nothing to do with the infinities at the EWL (as dealt with by the renormalization procedure). As is obvious from our calculations, the nonexistence of $M^{\lambda\mu}$ is related to the asymptotic behavior of the *Coulomb* part of $\Theta^{\mu\nu}_{\text{ret}}$ (in fact, the result holds even for a free particle). We want to point out although the Coulomb part of $\Theta^{\mu\nu}_{\text{ret}}$ is the most singular in ρ, and hence the one from which we expect the best asymptotic behavior for the integral, this is not the case, because the other parts depend also on the acceleration of the particle, which has the asymptotic behavior

$$a(\tau) \underset{\tau \to -\infty}{\to} 0.$$

In order to emphasize the independence of our result of the singularity at the EWL, let us note that the nonexistence of the integral defining $M^{\lambda\mu}$ is also obtained (the calculations are straightforward) for the model characterized by the four-current (30) in Ref. 18 (essentially a free-spherical charged shell). This model does not present any infinity in its corresponding $\Theta^{\mu\nu}$.

We have seen that even though the definition (1.2) is perfectly consistent, because P^μ exists and is a conserved quantity (see, Secs. II and III), the invariance of the theory under the Poincaré group cannot be carried on through the definitions (1.2) and (1.3) since the definition (1.3) is not appropriate to discuss the consequence of Lorentz invariance of the electromagnetic theory. We know that a given choice to reach the spatial infinity on σ leads to a definite result, e.g., $M^{\lambda\mu}(\sigma' \to \sigma)$ and $M^{\lambda\mu}(\bar{\sigma} \to \sigma)$ [see (2.22) and (2.23)]. The point of view of taking a definite choice to reach the spatial infinity on σ in order to define a unique $M^{\lambda\mu}$ is not physically consistent since *there are no physical reasons to choose a particular way to reach the spatial infinity* (two arbitrary points on σ are spatially related).

Summarizing, the definition (1.2) is quite adequate to discuss the translational invariance of the electromagnetic theory since P^μ exists and is a conserved quantity. On the other hand, the definition (1.3) is inadequate to discuss the Lorentz invariance of the electromagnetic theory. This invariance can be considered, however, in the form of a local conservation law through $M^{\lambda\mu\nu}$, as is discussed in Sec. 9 of Ref. 8, or in Ref. 19, for a point particle.

The evaluation of $M^{\lambda\mu}(\sigma' \to \sigma)$ through the region defined in Appendix B has been done in the literature,[5] and our result is the same as the one found in Ref. 5. Here we present an alternative procedure to calculate $M^{\lambda\mu}(\sigma' \to \sigma)$ in which there is no need to take the limit $E \to \infty$ inside the integrals. As is well-known, this procedure can be invalid. In fact, this procedure has been taken in the literature[20] (for a different purpose) and criticized by the same author.[21]

It is worthwhile to mention that in the formalism set by Van Dam and Wigner[22] for classical relativistic mechanics of interacting point particles (which includes Wheeler–Feynman electrodynamics[23]), an asymptotic interaction angular momentum is present.[24] Although their framework and ours are different, the root of their unexpected result, like ours (i.e., the nonexistence of $M^{\lambda\mu}$,) is the same, that is, the *long range* of the electromagnetic interaction.[24] We notice that in their formalism[22,24] the total angular momentum tensor exists and is a conserved quantity.

Finally, we want to mention that in the *standard* formulation of classical field theory, the fields are assumed to vanish fast enough, in order for the integrals (1.2) and (1.3) (with σ being, in general, an arbitrary spacelike hypersurface) to exist and define globally conserved quantities (see, e.g., Refs. 14, 15, 25, and 26). Our results show that for the very simple model of an electromagnetic field with a source, this is not the case for $M^{\lambda\mu}$ (even for a free point charged particle).

APPENDIX A: EVALUATION OF $M^{\lambda\mu}_{\text{mix}}$

In this appendix we want to evaluate $M^{\lambda\mu}_{\text{mix}}$. For this purpose (and for some calculations in Appendix B) let us introduce the following notation.

We let $\mathscr{C}(\tau')$ stand for the bounded segment of the future light cone with apex at $z(\tau')$, determined by the intersection between this light cone and the hypersurfaces σ and the Bhabha tube of radius E.

We have that[5]

$$\rho_0(\tau') \equiv v(\tau)\cdot[z(\tau) - z(\tau')]/v(\tau)\cdot[v(\tau') + u], \tag{A1}$$

is the set of values of the coordinate ρ given by the intersection between $\mathscr{C}(\tau')$ and σ (for $\tau' < \tau$).

In order to evaluate $M^{\lambda\mu}_{\text{mix}}$, we choose the bounded space-time region $\Omega(\tau'', \tau')$ shown in Fig. 2, where $\bar{\Sigma}$ represents a segment of the Bhabha tube of radius E [$E < \inf \rho_0(\tau')$], and $\bar{\sigma}$ is the part of the hyperplane σ deter-

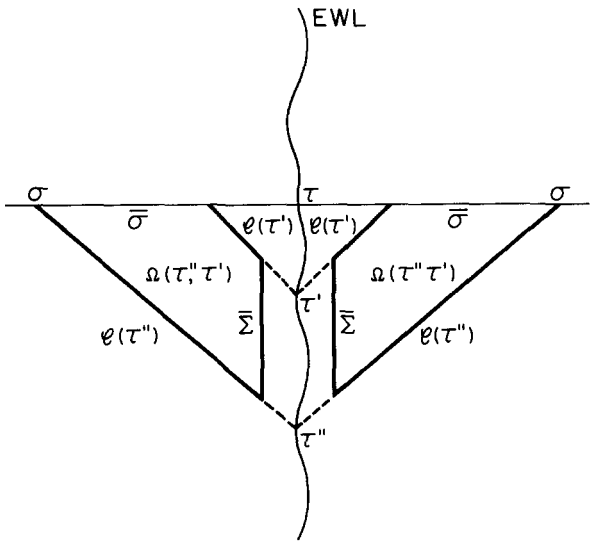


FIG. 2. Space-time region used in the evaluation of $M^{\lambda\mu}_{\rm mix}$, $P^{\mu}(\bar\sigma)$, and $M^{\lambda\mu}(\bar\sigma)$.

mined by the intersection between σ and the hypersurfaces $\mathscr{C}(\tau')$ and $\mathscr{C}(\tau'')$, with $\tau''<\tau'<\tau$. Since $\partial_\nu M^{\lambda\mu\nu}_{\rm mix}=0$ in that region, by the Gauss theorem we obtain

$$\int_{\bar\sigma} M^{\lambda\mu\nu}_{\rm mix}\,d\sigma_\nu=\int_{\bar\Sigma} M^{\lambda\mu\nu}_{\rm mix}\,d\Sigma_\nu+\int_{\mathscr{C}(\tau'')} M^{\lambda\mu\nu}_{\rm mix}\,dC_\nu - \int_{\mathscr{C}(\tau')} M^{\lambda\mu\nu}_{\rm mix}\,dC_\nu, \tag{A2}$$

where $d\Sigma_\nu=E^2\,d\tau\,d\Omega(u_\nu+(a\cdot u)R_\nu)$, and $dC_\nu=-\rho\,d\rho\,d\Omega R_\nu$.

Performing the required integrations, we obtain that

$$\int_{\bar\Sigma} M^{\lambda\mu\nu}_{\rm mix}\,d\Sigma_\nu=-\frac{e}{4\pi}\int_{\tau''}^{\tau'}d\tau\int d\Omega \times\left[z^\lambda(\tau),F^{\mu\alpha}_{\rm ext}(E,\tau,\Omega)\right]v_\alpha(\tau)+EG^{\lambda\mu}, \tag{A3}$$

where $G^{\lambda\mu}$ is a function of good behavior that satisfies $\lim_{E\to0}|G^{\lambda\mu}|<\infty$ because of the properties assumed for $F^{\mu\nu}_{\rm ext}$. If we use again the properties imposed on $F^{\mu\nu}_{\rm ext}$, we have that

$$\lim_{E\to0}\frac{1}{4\pi}\int d\Omega\,F^{\mu\alpha}_{\rm ext}(E,\tau,\Omega)=F^{\mu\alpha}_{\rm ext}[z(\tau)]\equiv F^{\mu\alpha}_{\rm ext}(\tau). \tag{A4}$$

Then,

$$\lim_{E\to0}\int_{\bar\Sigma} M^{\lambda\mu\nu}_{\rm mix}\,d\Sigma_\nu=-e\int_{\tau''}^{\tau'}d\tau\left[z^\lambda(\tau),F^{\mu\alpha}_{\rm ext}(\tau)\right]v_\alpha(\tau). \tag{A5}$$

Also, we find that

$$\int_{\mathscr{C}(\tau')} M^{\lambda\mu\nu}_{\rm mix}\,dC_\nu =\frac{e}{4\pi}\int d\Omega\int_E^{\rho_0(\tau')}d\rho\left[x^\lambda,F^{\mu\alpha}_{\rm ext}(\rho,\tau',\Omega)\right]\frac{R_\alpha}{\rho}, \tag{A6}$$

where $\rho_0(\tau')$ is defined in (A1).

Hence from (A6) it is clear that

$$\lim_{\tau''\to-\infty}\int_{\mathscr{C}(\tau'')} M^{\lambda\mu\nu}_{\rm mix}\,dC_\nu=0, \tag{A7}$$

because of the asymptotic properties of $F^{\mu\nu}_{\rm ext}$.

Since $F^{\mu\nu}_{\rm ext}$ is well defined and bounded on the EWL and $\lim_{\tau'\to\tau}\rho_0(\tau')=0$, from (A6), we obtain that

$$\lim_{\tau'\to\tau}\lim_{E\to0}\int_{\mathscr{C}(\tau')} M^{\lambda\mu\nu}_{\rm mix}\,dC_\nu=0. \tag{A8}$$

The final result is that

$$\int_\sigma M^{\lambda\mu\nu}_{\rm mix}\,d\sigma_\nu=-e\int_{-\infty}^{\tau}d\tau'\left[z^\lambda(\tau'),F^{\mu\alpha}_{\rm ext}(\tau')\right]v_\alpha(\tau'). \tag{A9}$$

Because of the assumed properties of $F^{\mu\nu}_{\rm ext}$, the integral of $M^{\lambda\mu\nu}_{\rm mix}$, pointed out in (A9), is absolutely convergent since it is not an improper integral [the same comment is valid for $P^{\mu}_{\rm mix}$ given by (1.11)]. For the same reason, the limit of integration $-\infty$ in (A9) can be substituted by τ_0.

APPENDIX B: EVALUATION OF $P^{\mu}(\bar\sigma)$, $M^{\lambda\mu}(\bar\sigma)$, $P^{\mu}(\sigma')$, AND $M^{\lambda\mu}(\sigma')$

In this appendix we want to evaluate some useful integrals.

First, we shall evaluate the integrals

$$P^\mu(\bar\sigma)\equiv\int_{\bar\sigma}\Theta^{\mu\nu}_{\rm ret}\,d\sigma_\nu,\quad M^{\lambda\mu}(\bar\sigma)\equiv\int_{\bar\sigma}M^{\lambda\mu\nu}_{\rm ret}\,d\sigma_\nu, \tag{B1}$$

where $\bar\sigma$ is the bounded part of the hyperplane σ defined in Appendix A and already shown in Fig. 2.

Since, in the region $\Omega(\tau'',\tau')$ shown in Fig. 2,

$$\partial_\nu\Theta^{\mu\nu}_{\rm ret}=0,\quad \partial_\nu M^{\lambda\mu\nu}_{\rm ret}=0, \tag{B2}$$

we obtain, using the Gauss theorem, that

$$\int_{\bar\sigma}\Theta^{\mu\nu}_{\rm ret}\,d\sigma_\nu=\int_{\bar\Sigma}\Theta^{\mu\nu}_{\rm ret}\,d\Sigma_\nu+\int_{\mathscr{C}(\tau'')}\Theta^{\mu\nu}_{\rm ret}\,dC_\nu - \int_{\mathscr{C}(\tau')}\Theta^{\mu\nu}_{\rm ret}\,dC_\nu, \tag{B3}$$

$$\int_{\bar\sigma}M^{\lambda\mu\nu}_{\rm ret}\,d\sigma_\nu=\int_{\bar\Sigma}M^{\lambda\mu\nu}_{\rm ret}\,d\Sigma_\nu+\int_{\mathscr{C}(\tau'')}M^{\lambda\mu\nu}_{\rm ret}\,dC_\nu - \int_{\mathscr{C}(\tau')}M^{\lambda\mu\nu}_{\rm ret}\,dC_\nu. \tag{B4}$$

The evaluation of the integrals in (B3) and (B4) are very simple and only the results will be stated. One obtains

$$P^\mu(\bar\sigma)=f^\mu_\tau(\tau')+\frac{2}{3}e^2\int_{\tau''}^{\tau'}a^2(\tau)v^\mu(\tau)d\tau-f^\mu_\tau(\tau''), \tag{B5}$$

$$M^{\lambda\mu}(\bar\sigma)=\left[z^\lambda(\tau'),f^\mu_\tau(\tau')\right]+\frac{2}{3}e^2\int_{\tau''}^{\tau'}\left[v^\lambda(\tau),a^\mu(\tau)\right]d\tau +\frac{2}{3}e^2\int_{\tau''}^{\tau'}\left[z^\lambda(\tau),v^\mu(\tau)\right]a^2(\tau)d\tau - \left[z^\lambda(\tau''),f^\mu_\tau(\tau'')\right], \tag{B6}$$

where

$$f^\mu_\tau(\tau')\equiv\frac{e^2[4v(\tau)\cdot v(\tau')v^\mu(\tau')+v^\mu(\tau)]}{6v(\tau)\cdot[z(\tau)-z(\tau')]}. \tag{B7}$$

We notice that $P^\mu(\bar\sigma)$ and $M^{\lambda\mu}(\bar\sigma)$ in Eqs. (B5) and (B6) do not depend on E (as it should be).

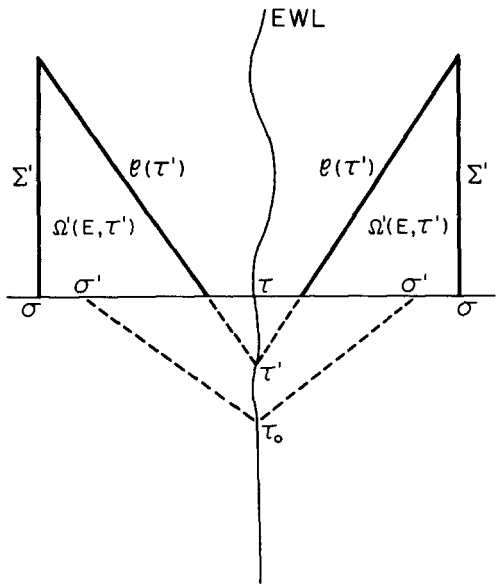


FIG. 3. Space-time region used in the evaluation of $P^{\mu}(\sigma')$ and $M^{\lambda\mu}(\sigma')$.

Second, we shall evaluate the integrals

$$P^{\mu}(\sigma') \equiv \int_{\sigma'} \Theta^{\mu\nu}_{\rm ret}\, d\sigma_{\nu}, \quad M^{\lambda\mu}(\sigma') \equiv \int_{\sigma'} M^{\lambda\mu\nu}_{\rm ret}\, d\sigma_{\nu}, \tag{B8}$$

where σ' is the bounded part of the hyperplane σ shown in Fig. 3 determined by the intersections between σ and the hypersurfaces $\mathscr{C}(\tau')$ and Σ' (see Refs. 5 and 8). The hypersurface Σ' is a segment of the Bhabha tube of radius E $[E > \sup \rho_0(\tau_0)]$ determined by the intersection of this tube and the hypersurfaces $\mathscr{C}(\tau')$ and σ.

Since, in the space-time region $\Omega'(E,\tau')$, shown in Fig. 3,

$$\partial_{\nu}\Theta^{\mu\nu}_{\rm ret} = 0, \quad \partial_{\nu} M^{\lambda\mu\nu}_{\rm ret} = 0, \tag{B9}$$

we obtain, using the Gauss theorem, that

$$\int_{\sigma'} \Theta^{\mu\nu}_{\rm ret}\, d\sigma_{\nu} = \int_{\mathscr{C}(\tau')} \Theta^{\mu\nu}_{\rm ret}\, dC_{\nu} + \int_{\Sigma'} \Theta^{\mu\nu}_{\rm ret}\, d\Sigma_{\nu}, \tag{B10}$$

$$\int_{\sigma'} M^{\lambda\mu\nu}_{\rm ret}\, d\sigma_{\nu} = \int_{\mathscr{C}(\tau')} M^{\lambda\mu\nu}_{\rm ret}\, dC_{\nu} + \int_{\Sigma'} M^{\lambda\mu\nu}_{\rm ret}\, d\Sigma_{\nu}. \tag{B11}$$

In order to calculate the integral on Σ' in Eqs. (B10) and (B11), we have to evaluate $\tau_R(E,\theta,\phi)$, that is, the set of values of the retarded coordinate τ determined by the intersection between σ and Σ'. The imposed condition, $E > \sup \rho_0(\tau_0)$, implies that the retarded coordinates at the intersection are the ones corresponding to a *free* particle. Then, using the equation of σ in retarded coordinates and the fact that the particle is free for proper times smaller than τ_0 [and hence $z(\tau') = z(\tau_0) + v(\tau_0)(\tau' - \tau_0)$, $\forall \tau' < \tau_0$], we obtain that

$$\tau_R(E,\theta,\phi) \equiv \tau_R = \frac{v(\tau)\cdot[z(\tau) - z(\tau_0) - Eu(\tau_0,\theta,\phi)]}{v(\tau)\cdot v(\tau_0)} - E + \tau_0. \tag{B12}$$

The evaluation of the integrals in (B10) and (B11) is very simple and only the results will be stated. One obtains

$$P^{\mu}(\sigma') = f^{\mu}_{\tau}(\tau') + \frac{2}{3}e^2\int_{\tau_0}^{\tau'} a^2(\tau)v^{\mu}(\tau)d\tau - \frac{e^2}{2E}v^{\mu}(\tau_0) + E^2\int d\Omega \int_{\tau_R}^{\tau_0} \Theta^{\mu\nu}_{\rm ret} u_{\nu}\, d\tau, \tag{B13}$$

$$M^{\lambda\mu}(\sigma') = [z^{\lambda}(\tau'), f^{\mu}_{\tau}(\tau')] + \frac{2}{3}e^2\int_{\tau_0}^{\tau'} [v^{\lambda}(\tau), a^{\mu}(\tau)]d\tau + \frac{2}{3}e^2\int_{\tau_0}^{\tau'} [z^{\lambda}(\tau), v^{\mu}(\tau)]a^2(\tau)d\tau - \frac{e^2}{2E}[z^{\lambda}(\tau), v^{\mu}(\tau_0)] + E^2\int d\Omega \int_{\tau_R}^{\tau_0} M^{\lambda\mu\nu}_{\rm ret} u_{\nu}\, d\tau, \tag{B14}$$

where f^{μ}_{τ} and τ_R are given by (B7) and (B12). The double integrals appearing in Eqs. (B13) and (B14) are easy to evaluate using the fact that $a^{\mu}(\tau) = 0$, $\forall \tau \in [\tau_R, \tau_0]$. It is obtained that

$$\int d\Omega \int_{\tau_R}^{\tau_0} \Theta^{\mu\nu}_{\rm ret} u_{\nu}\, d\tau = \frac{e^2}{6E^3}\left[\frac{v^{\mu}(\tau)}{v(\tau)\cdot v(\tau_0)} + v^{\mu}(\tau_0)\right], \tag{B15}$$

$$\int d\Omega \int_{\tau_R}^{\tau_0} M^{\lambda\mu\nu}_{\rm ret} u_{\nu}\, d\tau = -\frac{e^2}{6E^3}\left\{\left[z^{\lambda}(\tau_0), \frac{v^{\mu}(\tau)}{v(\tau)\cdot v(\tau_0)} + v^{\mu}(\tau_0)\right] + \frac{v(\tau)\cdot[z(\tau) - z(\tau_0)]}{[v(\tau)\cdot v(\tau_0)]^2}[v^{\lambda}(\tau_0), v^{\mu}(\tau)]\right\}. \tag{B16}$$

[1] F. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, MA, 1965).
[2] C. Teitelboim, Phys. Rev. D **1**, 1572 (1970).
[3] Ch. G. Van Weert, Physica **76**, 345 (1974).
[4] C. A. López and D. Villarroel, Phys. Rev. D **11**, 2724 (1975).
[5] R. Tabensky and D. Villarroel, J. Math. Phys. **16**, 1380 (1975).
[6] C. A. López and D. Villarroel, Phys. Rev. D **13**, 1802 (1976).
[7] C. A. López, Phys. Rev. D **17**, 2004 (1978).
[8] C. Teitelboim, D. Villarroel, and Ch. G. Van Weert, Riv. Nuovo Cimento **3**, 1 (1980).
[9] P. A. M. Dirac, Proc. R. Soc. London Ser. A **167**, 148 (1938).
[10] R. Tabensky, Phys. Rev. D **13**, 267 (1976).
[11] W. H. Fleming, *Functions of Several Variables* (Springer, New York, 1977).
[12] V. A. Ilyin and E. G. Poznyak, *Fundamentals of Mathematicals Analysis* (Mir, Moscow, 1982).
[13] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto, Toronto, 1949).
[14] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).
[15] N. N. Bogoliubov and D. V. Shirkov, *Introduction to Theory of Quantized Fields* (Wiley-Interscience, New York, 1980).
[16] E. Arnous, J. Madore, and A. Papapetrou, Nuovo Cimento A **53**, 393 (1968).
[17] J. N. Goldberg and R. P. Kerr, J. Math. Phys. **5**, 172 (1964).
[18] Julian Schwinger, Found. Phys. **13**, 373 (1983).
[19] E. G. P. Rowe, Phys. Rev. D **20**, 580 (1979).
[20] P. A. Hogan, Nuovo Cimento B **15**, 136 (1973).
[21] P. A. Hogan, Nuovo Cimento B **19**, 122 (1974).
[22] H. Van Dam and E. P. Wigner, Phys. Rev. **138**, B1576 (1965).
[23] J. A. Wheeler and R. P. Feynmann, Rev. Mod. Phys. **21**, 425 (1949).
[24] H. Van Dam and E. P. Wigner, Phys. Rev. **142**, 838 (1966).
[25] C. Moller, *The Theory of Relativity* (Oxford U.P., Oxford, 1969).
[26] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).

Coefficients of fractional parentage in the L–S coupling scheme

Akiva Novoselsky, Jacob Katriel,[a] and Robert Gilmore
Department of Physics and Atmospheric Science, Drexel University, Philadelphia, Pennsylvania 19104

(Received 2 October 1987; accepted for publication 27 January 1988)

An efficient procedure for the evaluation of the coefficients of fractional parentage (cfp's) for L–S coupled wave functions is presented. The cfp's are calculated separately for N particles, each with angular momentum l (s), coupled into a total angular momentum L (S). The N-particle states formed can belong to any permutational symmetry. The procedure for the evaluation of the L and the S cfp's for arbitrary permutational symmetry is a generalization of the procedure proposed by Bayman and Lande [Nucl. Phys. 77, 1 (1966)] for symmetric and antisymmetric states. It involves the construction and diagonalization of the matrices representing the quadratic Casimir operators for the appropriate special unitary and symplectic (or orthogonal) groups. The cfp's of the antisymmetric L–S coupled states are obtained in terms of products of cfp's for L and S corresponding to conjugate representations of the symmetric group. This method is demonstrated to provide cfp's for L–S states for systems with a considerably larger number of particles than is feasible using the procedures heretofore available.

I. INTRODUCTION

The theoretical study of many-particle systems possessing spherical symmetry has been a problem of central interest of both atomic and nuclear physics since the advent of quantum mechanics. One of the most efficient methods for the construction of many-particle wave functions with well-defined permutational symmetry and total angular momentum is the iterative procedure, originally proposed by Bacher and Goudsmit[1] and extensively developed by Racah.[2] In this procedure the N-particle wave function is expressed in terms of states formed from appropriate $(N-1)$-particle wave functions by the coupling of one more particle. The coefficients in the expansion of the N-particle state in terms of the $(N-1)$-particle states are known as the coefficients of fractional parentage (cfp's).

The most common procedure for the evaluation of the cfp's is the Redmond iteration procedure.[3] This procedure was extensively applied in the calculation of j–j coupled states in nuclear physics.[4] However, this procedure is limited to a relatively small number of particles, because of inherent numerical difficulties.[5] These difficulties become more severe in the context of L–S coupling and restrict the feasible applications to a rather small number of particles. An improved version of this procedure, which enlarges its domain of applicability, was recently introduced by Ji and Vallieres.[5]

A well-known and very powerful procedure for the evaluation of the cfp's for the symmetric and antisymmetric states in j–j coupling was introduced by Ginocchio[6] and, independently, by Bayman and Lande.[7] This procedure uses a group theoretical labeling of the states of interest. It involves the diagonalization of the matrices for the quadratic Casimir operators of the appropriate special unitary and symplectic (or orthogonal) groups in terms of a set of N-particle states. These N-particle states are obtained from symmetric or antisymmetric $(N-1)$-particle states specified by their total angular momentum and seniority, to which an Nth particle is coupled to form a state of well-defined total angular momentum. Consequently, the common eigenvectors are the desired cfp's and the eigenvalues specify the permutational symmetry and the seniority. This method was recently used by one of us[8] for the rapid and accurate computation of cfp's for identical particles with integral spin.

The Bayman and Lande procedure[7] was developed for j–j coupled states, involving Young frames with either one row or one column. In order to construct the L–S coupled wave function one has to couple an L function specified by an arbitrary Young frame with a corresponding S function. As commonly applied in atomic structure theory the construction of L–S coupled wave functions involves spin-$\frac{1}{2}$ particles. Consequently, the Young frames specifying the permutational symmetry have at most two rows for spin states. In the recently developed fermion dynamical symmetry model (FDSM)[9] the total angular momentum of a nucleon is decomposed into a pseudo-orbital angular momentum $\mathbf{k}$ and a pseudospin $\mathbf{i}$. This $\mathbf{i}$ can assume values up to $\frac{15}{2}$, and $\mathbf{k}$ can be 0, 1, or 2. The construction of the K–I coupled wave function in the case of the actinides requires the coupling of up to 22 particles. The calculation of the cfp's for such cases is certainly beyond the scope of the presently available codes.

Our objective in this work is to provide a fast and efficient mechanism for computing the cfp's for a relatively large number of particles, each with a spin j, possessing an arbitrary permutational symmetry. This procedure would be applicable to the construction of an L–S coupled state in a many-particle system in which each individual particle's l and s are not restricted to very small values.

Figures 1 and 2 serve as a guide through the calculation of symmetrized states and associated cfp's, carried out in this work. The principal characters are identified in the legend: states, cfp's, and quantum numbers for the important groups—J for SU(2), Γ_N for SU_{2j+1}, and $\mathbf{v}$ (seniority) for the symplectic group Sp_{2j+1} (j half-integral) or the orthogonal group SO_{2j+1} (j integral). The sections in this work

[a] Permanent address: Department of Chemistry, Technion—Israel Institute of Technology, 32000 Haifa, Israel.

 0022-2488/88/061368-21$02.50

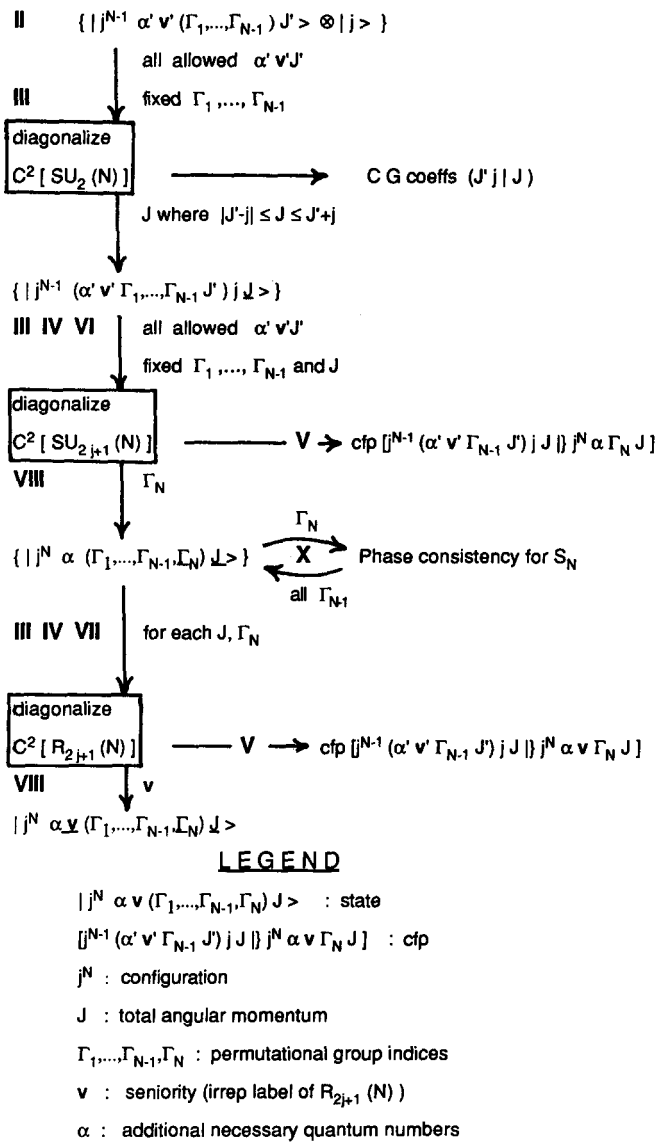


FIG. 1. Schematic presentation of the procedure for the evaluation of the cfp's for arbitrary permutational symmetry.

where the particular calculations are carried out are indicated by a Roman number. For example, the maximal set of quantum numbers generated by these symmetries, and used to label the states are identified in Sec. II.

The procedure for computing symmetrized states and cfp's is recursive. Beginning with symmetrized states for $N-1$ particles, the symmetrized N-particle states are constructed by following essentially the same procedure three successive times, once for each of the groups $SU_2(N)$, $SU_{2j+1}(N)$, and $R_{2j+1}(N)$ [i.e., $Sp_{2j+1}(N)$ or $SO_{2j+1}(N)$]. These three steps are displayed in Fig. 1, each identifying one new N-particle quantum number (which is underlined in the figure). To initate this succession of calculations, unsymmetrized N-particle states are constructed by taking direct products of symmetrized $(N-1)$-particle states with a single-particle state. All the $(N-1)$-particle states considered have the same value of Γ_{N-1} (which remains a good quantum number of the symmetrized N-particle states) and all the values of J' and $\mathbf{v}'$ consistent with it. The N-particle states so constructed span a space with the multiplicity of N quantum numbers: J, Γ_N, and $\mathbf{v}$.

States with a good quantum number J can in principle be constructed by diagonalizing the total angular momentum operator J_N^2, which is the Casimir operator for $SU_2(N)$ in this basis. The corresponding eigenvectors are the standard Clebsch–Gordan (CG) coefficients. In fact, the widespread knowledge of these cfp's for SU_2 makes this step unnecessary in practice. The quantum number J generated in this step is indicated below the appropriate box in Fig. 1, and underlined in the symbol of the resulting N-particle state.

The second stage in this procedure involves diagonalization of the quadratic Casimir operator of $SU_{2j+1}(N)$ in the basis with fixed J. The subspace of states corresponding to each distinct eigenvalue is associated with an irrep of SU_{2j+1}; these irreps may be uniquely identified through the eigenvalues of the quadratic Casimir operator. This is a somewhat surprising result, which is a consequence of the recursive buildup employed, as discussed in Sec. VIII. The corresponding eigenvectors are the seniority-free cfp's. The states so constructed have Γ_N, in addition to J, as a good quantum number (underlined for emphasis). The new quantum number Γ_N, constructed using the group SU_{2j+1}, specifies in fact an irrep of the permutational group S_N. This comes about because of the duality between $SU_{2j+1}(N)$ and S_N. The diagonalization destroys two of the $(N-1)$-particle quantum numbers (J' and $\mathbf{v}'$), since the corresponding operators do not commute.

The cfp's corresponding to the same Γ_N but distinct Γ_{N-1} result from "distinct" diagonalizations. In consequence, the phase information between the cfp's is an artifact of the computational procedure. These cfp's are made phase consistent following a procedure elaborated in Sec. X.

The third stage in this process involves diagonalization of the quadratic Casimir operator for $R_{2j+1}(N)$ in the basis of states with good J and Γ_N. The subspace of states corresponding to each distinct eigenvalue belongs to a distinct irrep of R_{2j+1}. The eigenvalue is sufficient to uniquely label the irrep of R_{2j+1}. The irrep label, which is a Young partition $\mathbf{v}$, is the state's seniority. The states that result have good J, Γ_N, and $\mathbf{v}$ (underlined for emphasis). The eigenvectors of this matrix diagonalization are the cfp's sought.

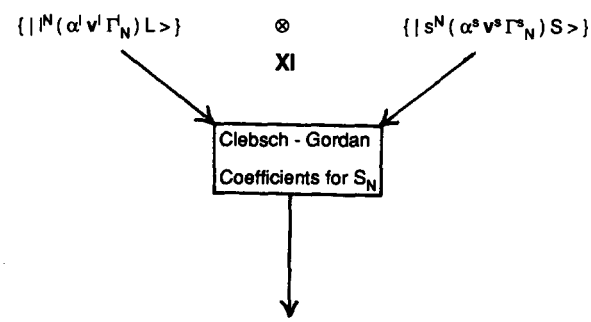


FIG. 2. Schematic presentation of the procedure for the evaluation of the L–S coupled cfp's.

The illustration of the procedure continues in Fig. 2, displaying the L–S coupling step. To obtain the L–S coupled states and cfp's, the procedure presented in Fig. 1 is carried out separately for the l coupling and the s coupling. These cfp's are then combined by means of S_N (permutation group) CG coefficients (actually, inner product isoscalar factors) to obtain the L–S coupled cfp's exhibiting appropriate permutational symmetry. In the usual case of Bose and Fermi symmetries, and the S_N CG coefficients vanish except when the L and S permutational symmetries involve equivalent and conjugate representations, respectively.

The procedure described above is presented in this paper as follows: In Sec. II we discuss the state labeling scheme adopted in this work and compare it with the Weyl complete labeling scheme. In Sec. III we construct the form of the N-particle quadratic Casimir operators for the groups SU_{2j+1}, Sp_{2j+1}, SO_{2j+1}, and SU_2. A result concerning the matrix element of a one-body operator between two states of arbitrary permutational symmetries, which is needed for the evaluation of the matrix elements of the quadratic Casimir operators, is developed in Sec. IV. The cfp's themselves are defined in Sec. V. The matrix elements for the quadratic Casimir operators are constructed explicitly in Sec. VI for SU_{2j+1} and in Sec. VII for Sp_{2j+1} and SO_{2j+1}. These are the matrices actually diagonalized in the recursive procedure outlined above and indicated in Fig. 1. The identification of the irreps of the special unitary and symplectic (or orthogonal) groups by means of the eigenvalues of the quadratic Casimir operators is discussed in Sec. VIII. Section IX is devoted to a set of illustrative examples. The procedure used to construct a phase-consistent set of cfp's for degenerate irreps is discussed in Sec. X. The L–S coupled cfp's are constructed in Sec. XI. The overall computational procedure and some detail concerning the performance of the computer code implementing it (a copy of which is available upon request), are presented in Sec. XII. In Sec. XIII we present some concluding remarks.

II. STATE LABELING SCHEMES

The purpose of the present section is to describe the set of quantum numbers which specify the appropriately symmetrized states. As a first step, we determine the symmetries present. A single particle of angular momentum j carries a representation of the group SU_{2j+1}. A system of N identical particles carries a representation of the group $[SU_{2j+1}]^{\otimes N}$ (N times direct product of the group SU_{2j+1}), which can be reduced[10] to the direct product group $SU_{2j+1} \otimes S_N$, where S_N is the symmetric (or permutation) group on N particles. It is useful to introduce a subgroup, $R_{2j+1} \subset SU_{2j+1}$, to further refine the state classification procedure. This subgroup is the symplectic group, Sp_{2j+1}, for half-integral angular momentum, or the orthogonal group SO_{2j+1}, for integral angular momentum.[11] Finally, states are classified according to their total angular momentum J by introducing the subgroup $SU_2 \subset R_{2j+1}$. The group–subgroup chain is

$$SU_{2j+1} \xrightarrow{N \text{ particles}} [SU_{2j+1}]^{\otimes N} \supset SU_{2j+1} \otimes S_N \supset R_{2j+1} \otimes S_N \supset SU_2 \otimes S_N. \tag{2.1}$$

The state labels are related to the chain of subgroups given in (2.1). For an N-particle system described by the subgroup $SU_{2j+1} \otimes S_N$ they are $\{\lambda\}_N$ and Γ_N, where $\{\lambda\}_N$ is an N-box proper Young frame (i.e., a Young frame with at most $2j+1$ rows) labeling an Nth-order irreducible tensor representation of SU_{2j+1}, and Γ_N is a Young partition labeling an irreducible representation (irrep) of S_N. The partitions Γ_N and $\{\lambda\}_N$ must be identical, according to a result by Weyl.[10]

The representation labels $\{\lambda\}_N$ and Γ_N generate internal indices that describe the individual states within these representations. The internal state labels of the symmetric group S_N are usually denoted by μ_N. A more informative labeling of the internal states is obtained by constructing the chain of subgroups

$$S_N \supset S_{N-1} \supset \cdots S_3 \supset S_2 \supset S_1. \tag{2.2}$$

The corresponding sequence of irreps, $\Gamma_N \Gamma_{N-1} \cdots \Gamma_1$, is equivalent to a Yamanouchi symbol $Y(S_N)$.[11]

The irrep $\{\lambda\}_N$ of SU_{2j+1} decomposes into a direct sum of irreps under the reduction of R_{2j+1} (cf. Ref. 11):

$$\{\lambda\}_N \xrightarrow{SU_{2j+1} \downarrow Sp_{2j+1}} \sum \langle \lambda \rangle_N, \tag{2.3}$$

$$\{\lambda\}_N \xrightarrow{SU_{2j+1} \downarrow SO_{2j+1}} \sum [\lambda]_N. \tag{2.4}$$

The branching rules for the reduction are known.[12] The partition $\{\lambda\}_N$ for SU_{2j+1} has N boxes, but the partition $\langle \lambda \rangle_N$ ($[\lambda]_N$) for an N-particle representation of Sp_{2j+1} (SO_{2j+1}) may have fewer than N boxes. In the case of the fully antisymmetric (symmetric) representation of SU_{2j+1} the number of boxes in $\langle \lambda \rangle_N$ ($[\lambda]_N$) has been called the *seniority* of the state. In general, seniority is the representation label for the subgroup R_{2j+1}, and it is a good quantum number. We shall use the generic symbol $\mathbf{v}$ for either $\langle \lambda \rangle_N$ or $[\lambda]_N$, depending on whether $2j+1$ is even or odd. The seniority $\mathbf{v}$ is the permutational symmetry type with the smallest number of particles in which the symplectic (orthogonal) symmetry is first encountered. The total number of boxes in $\mathbf{v}$, $v = \Sigma_i \lambda_i$, will be referred to as the seniority index. For symmetric and antisymmetric states, for which $\mathbf{v}$ has only one row or column, the seniority index specifies the irrep of R_{2j+1} (i.e., SO_{2j+1} and Sp_{2j+1}, respectively) uniquely.

Under the reduction $R_{2j+1} \supset SU_2$ the representation $\langle \lambda \rangle_N$ or $[\lambda]_N$ decomposes into a direct sum of irreps of SU_2. Their angular momentum J is a good quantum number. This reduction is generally not simple: A given value of angular momentum may occur more than once.

Besides the good numbers described above, there are additional good quantum numbers. We show in Sec. III that the SU_{2j+1} representation labels for $(N-1, N-2, \ldots, 2, 1)$-particle states are all good quantum numbers. The N-particle states with the appropriate symmetry may thus be written

$$|j^N; \alpha \mathbf{v} \{\lambda\}_N, \{\lambda'\}_{N-1}, \ldots, \{\lambda''\}_1, \Gamma_N \mu_N; J, M\rangle. \tag{2.5}$$

Here j is the single-particle angular momentum, and j^N denotes the N-particle configuration. The sequence $\{\lambda\}_N$, $\{\lambda'\}_{N-1}, \ldots, \{\lambda''\}_1$ represents SU_{2j+1} representation labels

for states with $N,N-1,\ldots,1$ particles. The label $\{\lambda'\}_{i+1}$ is obtained from $\{\lambda\}_i$ by proper addition of a single box. In this sense, the entire sequence may be summarized by a Yamanouchi symbol $Y(\mathrm{SU}_{2j+1})$.[11] The representation label $\mathbf{v}$ determines the seniority of the states, and J,M are the angular momentum quantum numbers. The symbol α represents all other quantum numbers "required" for a complete state specification.

The dual roles of the special unity group SU_{2j+1} and the symmetric group S_N may be emphasized by writing the state (2.5) in the form

$$|j^N;\alpha\mathbf{v}Y(\mathrm{SU}_{2j+1}),Y(\mathrm{S}_N);JM\rangle. \tag{2.6}$$

The Yamanouchi symbols for the special unitary and symmetric groups are based on the same Young frame. Since the duality between the special unitary and symmetric groups holds for each pair of labels $\{\lambda\}_{N'}$ and $\Gamma_{N'}$ $(N'=1,2,\ldots,N)$, the Yamanouchi symbols are actually identical. This property is central to the procedure we employ, allowing the use of the special unitary Casimir operators to specify the permutational symmetry characteristics.

While it is usually assumed that the additional labels represented by α can be found "in principle," this is rarely demonstrated. It has been established by Weyl[10] that all the states in the Hilbert space for $[\mathrm{SU}_{2j+1}]^{\otimes N}\supset \mathrm{SU}_{2j+1}\otimes \mathrm{S}_N$ are uniquely labeled. The labeling proceeds via the group–subgroup reduction (2.2) for the permutation group S_N as well as the dual group–subgroup reduction for the special unitary group

$$\mathrm{SU}_{2j+1}\supset \mathrm{SU}_{2j}\supset\cdots\supset \mathrm{SU}_3\supset \mathrm{SU}_2\supset U_1. \tag{2.7}$$

Every state for an irreducible representation of $\mathrm{SU}_{2j+1}\otimes \mathrm{S}_N$ is uniquely labeled as follows:

$$|\Gamma_N\mu_N\{\lambda\}_N\nu_N\rangle, \tag{2.8}$$

where Γ_N is an N-box frame describing an irreducible representation of S_N and $\{\lambda\}_N$ is an N-box frame describing an irreducible representation of SU_{2j+1}. As for (2.5), $\Gamma_N=\{\lambda\}_N$ and the internal label μ_N is a Yamanouchi symbol for S_N. The label ν_N in (2.8) describes a sequence of irreps for the reduction (2.7), which differs from the reduction $\{\lambda\}_N,\{\lambda'\}_{N-1},\ldots,\{\lambda''\}_1$ used to describe the state in (2.5). The internal labels μ_N and ν_N, which are not related to one another in any rigid way, produce a unique classification for all states in the Hilbert space for $\mathrm{SU}_{2j+1}\otimes \mathrm{S}_N$.

The group–subgroup reduction (2.7) is not useful for our general purpose, as the physical angular momentum group is not contained in any of the subgroups of SU_{2j+1}:

$$\mathrm{SU}_{2(\mathrm{phys})}\subset \mathrm{SU}_{2j+1}\supset \mathrm{SU}_{2j}\not\supset \mathrm{SU}_{2(\mathrm{phys})}. \tag{2.9}$$

The failure of this last inclusion is responsible for the complexity in constructing appropriately symmetrized N-particle states of good angular momentum.

We have only referred to the Weyl labeling scheme (2.8) in order to point out that each state in each component Hilbert space can be uniquely identified. The labeling scheme which we actually adopt is that specified by (2.5). Comparison to the complete labeling achieved in the Weyl scheme establishes that the "additional quantum numbers," α, introduced in (2.5), exist in principle.

III. THE SET OF COMMUTING OPERATORS CHARACTERIZING THE STATES

The various labels specifying the N-particle states were introduced in the previous section [Eq. (2.5)]. In the present section we discuss some further properties of the groups involved and present the operators of interest in a form suitable for our further development.

A. The quadratic Casimir operator of the special unitary group SU_{2j+1}

The special unitary group SU_{2j+1} is a compact continuous group and has, therefore, a countable set of finite-dimensional inequivalent unitary irreducible representations, characterized by Young frames with an arbitrary number of boxes. What we are really interested in is a realization of this group over the space of N-particle states, denoted by $\mathrm{SU}_{2j+1}(N)$. In terms of Racah's unit tensor operators,[2,4] which satisfy

$$\langle\psi^{j'}_{m'}|u^k_q|\psi^j_m\rangle=\delta_{jj'}\,(kjqm|j'm'), \tag{3.1}$$

the generators of the group are given by

$$U^k_q(N)=\sum_{i=1}^{N}u^k_q(i). \tag{3.2}$$

The quadratic Casimir operator is given by the expression[7]

$$C^2[\mathrm{SU}_{2j+1}(N)]=\frac{1}{2j+1}\sum_{k=1}^{2j}(-1)^k(2k+1)^{3/2}\times[U^k(N)U^k(N)]^0_0\,. \tag{3.3}$$

The most important property of the realization $\mathrm{SU}_{2j+1}(N)$ is that its operation within the N-particle space is in very close correspondence with that of the symmetric group S_N. In particular, as emphasized in the previous section, the irreducible basis corresponding to a specific Young frame of S_N corresponds to the same Young frame of $\mathrm{SU}_{2j+1}(N)$.

In order to discuss some properties of the Casimir operators corresponding to different values of N we note that

$$U^k_q(N)=U^k_q(N-1)+u^k_q(N) \tag{3.4}$$

and obtain

$$[U^k(N)U^k(N)]^0_0=[U^k(N-1)U^k(N-1)]^0_0+2[U^k(N-1)u^k(N)]^0_0+[u^k(N)u^k(N)]^0_0. \tag{3.5}$$

From Eq. (6) of Ref. 7, the last term in Eq. (3.5) is

$$[u^k(N)u^k(N)]^0_0=(-1)^{2j}\sqrt{2j+1}\begin{Bmatrix}k & k & 0\\ j & j & j\end{Bmatrix}u^0_0(N), \tag{3.6}$$

and from p. 521 of Ref. 4, the middle term in Eq. (3.5) is

$$[U^k(N-1)u^k(N)]^0_0=[(-1)^k/\sqrt{2k+1}\,](U^k(N-1)\cdot u^k(N)). \tag{3.7}$$

Using Eqs. (3.4)–(3.7) we obtain the following expression for the quadratic Casimir operator of the special unitary

group:

$$C^2[\mathrm{SU}_{2j+1}(N)] = C^2[\mathrm{SU}_{2j+1}(N-1)] + \frac{2}{2j+1}\sum_{k=1}^{2j}(2k+1) \times (U^k(N-1)\cdot u^k(N)) + u_0^0(N)\frac{4j(j+1)}{2j+1}. \quad (3.8)$$

From this expression and from the fact that $C^2[\mathrm{SU}_{2j+1}(N-1)]$ commutes with $U_q^k(N-1)$ (Ref. 7), it follows that $C^2[\mathrm{SU}_{2j+1}(N-1)]$ and $C^2[\mathrm{SU}_{2j+1}(N)]$ commute with one another. Using the above relation one can express $C^2[\mathrm{SU}_{2j+1}(N)]$ in terms of $C^2[\mathrm{SU}_{2j+1}(N')]$ for any $N'<N$ and use this expression to show that these two operators commute.

B. The total angular momentum J_N^2

The total angular momentum operator J_N^2 can be expressed in terms of its Cartesian components

$$J_a = \sum_{i=1}^{N} J_a(i), \quad a = x,y,z,$$

where the $J_a(i)$ are the single-particle angular momentum operators. This operator commutes with S_N as well as with each one of its subgroups $\mathrm{S}_{N'}$, $N' \leqslant N$. It follows immediately that it commutes with each one of the realizations $C^2[\mathrm{SU}_{2j+1}(N')]$. Note, however, that $J_{N'}^2$ $(N'<N)$, which trivially commutes with J_N^2, does not commute with $C^2[\mathrm{SU}_{2j+1}(N)]$ (or, for that matter, with any $C^2[\mathrm{SU}_{2j+1}(N'')]$ with $N'<N''$). An immediate consequence, which is well-known in the context of antisymmetric wave functions, is that when an N-particle state with an angular momentum quantum number J is formed from an antisymmetric $(N-1)$-particle state with an angular momentum quantum number J' by means of angular momentum coupling, the coupled state is a common eigenstate of J_N^2 and J_{N-1}^2, but it is, in general, not antisymmetric. In order to antisymmetrize the N-particle state one has to express it in terms of $(N-1)$-particle states with several different values of J'. Thus the common eigenstate of J_N^2 and $C^2[\mathrm{SU}_{2j+1}(N)]$ is not, in general, an eigenstate of J_{N-1}^2, while it is an eigenstate of all $C^2[\mathrm{SU}_{2j+1}(N')]$ with $N'<N$.

C. The quadratic Casimir operators of the symplectic and orthogonal groups

When $2j+1$ is even (j is a half-integer), the symplectic group Sp_{2j+1} is capable of providing a further classification when a degeneracy remains after classification with respect to $\{\mathrm{SU}_{2j+1}(N'); N'=1,2,...,N\}$ and J_N^2 has been performed. When $2j+1$ is odd (j is an integer), the same role is assumed by SO_{2j+1}. As stated in Sec. II, we denote these two groups by the generic symbol R_{2j+1}, the value of $2j+1$ (even or odd) identifying R as either Sp or SO.

The quadratic Casimir operator for R_{2j+1} is[7]

$$C^2[R_{2j+1}(N)] = \frac{1}{2j+1}\sum_{k=1,3,...}^{2[j-1/2]+1}(-1)^k(2k+1)^{3/2} \times [U^k(N)U^k(N)]_0^0. \quad (3.9)$$

Using Eq. (3.5) it can be written in the form

$$C^2[R_{2j+1}(N)] = C^2[R_{2j+1}(N-1)] + \frac{2}{2j+1}\sum_{k=1,3,...}^{2[j-1/2]+1}(2k+1) \times (U^k(N-1)\cdot u^k(N)) - \frac{u_0^0(N)}{2j+1}\sum_{k=1,3,...}^{2[j-1/2]+1}(2k+1)^{3/2} \times (-1)^{2j}\sqrt{2j+1}\begin{Bmatrix} k & k & 0 \\ j & j & j \end{Bmatrix}. \quad (3.10)$$

The sum in the third term can be shown to be equal to $-(j+1)(2j+1)$ for half-integral j and to $-j(2j+1)$ for integral j.

Since R_{2j+1} is a subgroup of SU_{2j+1}, their Casimir operators commute. Moreover, since $C^2[R_{2j+1}(N)]$ is symmetric with respect to all the permutations in S_N, it is symmetric with respect to any $S_{N'}$, $N' \leqslant N$. However, $C^2[R_{2j+1}(N')]$ is not symmetric with respect to S_N, for $N'<N$, and, consequently, does not commute with $C^2[\mathrm{SU}_{2j+1}(N)]$. This is completely analogous to the behavior of the angular momentum operators, as presented in the previous section.

D. Maximal set of commuting operators

The foregoing discussion suggests that the set of quadratic Casimir operators corresponding to $\mathrm{SU}_{2j+1}(1)$, $\mathrm{SU}_{2j+1}(2),...,\mathrm{SU}_{2j+1}(N)$, as well as to $R_{2j+1}(N)$ constitute, together with J_N^2 and J_z, a set of commuting operators. Since the sequence of Casimir operators for SU_{2j+1} specifies a sequence of Young frames with one box added at a time, the state they characterize can equivalently be designated by the corresponding Yamanouchi symbol.

We stress that the sequence $\{\mathrm{SU}_{2j+1}(N'); N'=1,2,...,N\}$, which we use to label the states, differs from the Weyl sequence $\{\mathrm{SU}_r(N); r=2j+1,2j,...,1\}$.[10] In our labeling the sequence of Young frames corresponding to the $\mathrm{SU}_{2j+1}(N')$ irreps is identical with the sequence of Young frames corresponding to the symmetric groups $\{\mathrm{S}_{N'}; N'=1,2,...,N\}$, so that a double Young frame notation is not needed.

IV. MATRIX ELEMENTS OF ONE-BODY OPERATORS

In this section we evaluate the matrix elements of the one-body operator

$$V = \sum_{i=1}^{N} v_i \quad (4.1)$$

between two states having internal labels μ_N and μ'_N, respectively, of the irreps Γ_N and Γ'_N of the symmetric group S_N [or of $\mathrm{SU}_{2j+1}(N)$]. Since there are $(N-1)!$ permutations P in S_N that transform any specific index m into some other index i, it follows that

$$\sum_{P\in \mathrm{S}_N} P^{-1}v_m P = (N-1)!\sum_{i=1}^{N} v_i. \quad (4.2)$$

Note that m could be any of the indices 1,2,...,N. For the applications we have in mind it will be convenient to set

$m = N$. The corresponding matrix elements satisfy

$$\sum_{P\in S_N} \langle \Gamma_N \mu_N | P^{-1} v_N P | \Gamma'_N \mu'_N \rangle = (N-1)! \left\langle \Gamma_N \mu_N \left| \sum_{i=1}^{N} v_i \right| \Gamma'_N \mu'_N \right\rangle . \tag{4.3}$$

The operation of the permutation P on a state $|\Gamma'_N \mu'_N\rangle$ is given by

$$P|\Gamma'_N \mu'_N\rangle = \sum_{\mu''_N=1}^{n_{\Gamma'_N}} |\Gamma'_N \mu''_N\rangle D^{(\Gamma'_N)}_{\mu''_N \mu'_N}(P), \tag{4.4}$$

where $\{D^{(\Gamma'_N)}_{\mu''_N \mu'_N}(P); \mu'_N, \mu''_N = 1,2,...,n_{\Gamma'_N}\}$ are the elements of the representation matrix of P in the irrep Γ'_N. Hence, the lhs of Eq. (4.3) is equal to

$$\frac{N!}{n_{\Gamma_N}} \delta_{\Gamma_N \Gamma'_N} \delta_{\mu_N \mu'_N} \sum_{\mu''_N=1}^{n_{\Gamma_N}} \langle \Gamma_N \mu''_N | v_N | \Gamma_N \mu''_N \rangle, \tag{4.5}$$

where use has been made of the representation orthogonality theorem.[11] Substituting Eq. (4.5) in Eq. (4.3) we obtain the relation

$$\left\langle \Gamma_N \mu_N \left| \sum_{i=1}^{N} v_i \right| \Gamma'_N \mu'_N \right\rangle = \frac{N}{n_{\Gamma_N}} \delta_{\Gamma_N \Gamma'_N} \delta_{\mu_N \mu'_N} \sum_{\mu''_N=1}^{n_{\Gamma_N}} \langle \Gamma_N \mu''_N | v_N | \Gamma_N \mu''_N \rangle. \tag{4.6}$$

Note that the only dependence on the internal state labels μ_N, μ'_N is via the Kronecker delta factor. For antisymmetric (or symmetric) N-body states, this equation reduces to the standard result

$$\left\langle (\lambda)_N \left| \sum_{i=1}^{N} v_i \right| (\lambda)_N \right\rangle = N \langle (\lambda)_N | v_N | (\lambda)_N \rangle, \tag{4.7}$$

where $(\lambda)_N = (1^N)$ or (N).

If the states of interest are labeled by angular momentum quantum numbers J and J', respectively, in addition to their characterization with respect to the symmetric group, we obtain the reduced matrix elements

$$\left\langle \Gamma_N \mu_N J \left| \left| \sum_{i=1}^{N} v_i \right| \right| \Gamma'_N \mu'_N J' \right\rangle = \frac{N}{n_{\Gamma_N}} \delta_{\Gamma_N \Gamma'_N} \delta_{\mu_N \mu'_N} \sum_{\mu''_N=1}^{n_{\Gamma_N}} \langle \Gamma_N \mu''_N J || v_N || \Gamma_N \mu''_N J' \rangle. \tag{4.8}$$

V. COEFFICIENTS OF FRACTIONAL PARENTAGE

The coefficients of fractional parentage (cfp's) are a central element in the buildup of N-particle states of well-defined total angular momentum and permutational symmetry from $(N-1)$-particle states similarly characterized.

For N particles, each having spin j, the state with total angular momentum J and additional labels α can be expressed in the form

$$|j^N \alpha J\rangle = \sum_{J',\beta} |j^{N-1}(\beta J')jJ\rangle [j^{N-1}(\beta J')jJ | \} j^N \alpha J]. \tag{5.1}$$

Here, $|j^{N-1}(\beta J')jJ\rangle$ is a state of total angular momentum J, obtained by coupling of a single particle to a state of $N-1$ particles with total angular momentum J' and additional labels β. The coefficient $[j^{N-1}(\beta J')jJ | \} j^N \alpha j J]$ multiplying this state in Eq. (5.1) is the corresponding cfp. If both the $(N-1)$-particle state $|j^{N-1}\beta J'\rangle$ and the N-particle state $|j^N \alpha J\rangle$ are assumed to be antisymmetric or symmetric, the above relation is the standard defining relation of the cfp's. However, if the additional quantum numbers α and β contain state labels $\Gamma_N \mu_N$ and $\Gamma_{N-1}\mu_{N-1}$, respectively, specifying their permutational (or unitary) symmetry as well as group representation labels $\mathbf{v}$ and $\mathbf{v}'$ specifying their symmetry under the symplectic (or orthogonal) group, we obtain

$$|j^N \alpha \mathbf{v} \Gamma_N \mu_N J\rangle = \sum_{\beta \mathbf{v}' J'} |j^{N-1}(\beta \mathbf{v}' \Gamma_{N-1} \mu_{N-1} J')jJ\rangle \times [j^{N-1}(\beta \mathbf{v}' \Gamma_{N-1}\mu_{N-1} J')jJ | \} j^N \alpha \mathbf{v} \Gamma_N \mu_N J]. \tag{5.2}$$

In the last expression we have redefined α and β to contain the remaining quantum numbers needed to uniquely specify the states.

The ranges of the quantum numbers J', $\mathbf{v}'$ are obtained directly from the properties of the Clebsch–Gordan series of the respective groups. For SU(2), $J' \otimes j = J' \otimes \square$ gives the familiar result $|J-j| \leqslant J' \leqslant J+j$. For Sp_{2j+1} and SO_{2j+1}, $\mathbf{v}' \otimes \square = \Sigma\, \mathbf{v}$, where the partition $\mathbf{v}$ is obtained by proper addition of a single box to or deletion of a single box from the partition $\mathbf{v}'$. Proper addition of a single box to $\mathbf{v}'$ may result in a partition $\mathbf{v}$ with more than the proper number of rows ($[j+\frac{1}{2}]$). In this case the appropriate modification rules must be applied.[13–16] These reduce, in the present circumstances, to

Sp_{2j+1}: if $\mathbf{v}$ has $(j+\frac{1}{2})+1$ rows, the *partition* must be deleted;

SO_{2j+1}: if $\mathbf{v}$ has $j+1$ rows, the *last row*, of length 1, must be removed.

From the modification rules it follows that for j half-integral the N-particle states have even seniority index if N is even and odd seniority index if N is odd. For integral spin particles, the N-particle state may have both even and odd seniority indices, when $N > j$. A well-known example of the SO_{2j+1} modification rule, the $j = 1$ case, is briefly discussed in Sec. IX B.

The representation and internal labels $\Gamma_{N-1}\mu_{N-1}$ are uniquely determined in terms of $\Gamma_N \mu_N$, so there is no summation over these indices in Eq. (5.2). To specify the relation between $\Gamma_N \mu_N$ and $\Gamma_{N-1}\mu_{N-1}$, we recall that the labeling of the N-particle states by the sequence of eigenvalues of

$$C^2[SU_{2j+1}(N')], \quad N' = 1,2,...,N,$$

corresponds to a sequence of Young frames with one box added at a time. The representation label Γ_N corresponds to the last frame in the sequence, and the internal label μ_N consists of a sequence of Young frames leading to Γ_N. Here Γ_{N-1} is the last [$(N-1)$-particle] frame in the sequence μ_N, and μ_{N-1} is the sequence obtained from μ_N by omitting the $(N-1)$-particle frame. It will be shown in Sec. VI that the cfp's depend only on the N- and $(N-1)$-particle frames Γ_N and Γ_{N-1}, and not on the preceding frames specifying μ_N and μ_{N-1}.

VI. MATRIX ELEMENTS OF $C^2[SU_{2j+1}(M)]$

In the present section we derive the expressions for the matrix elements of the quadratic Casimir operator $C^2[SU_{2j+1}(N)]$. These matrix elements are evaluated between N-particle states obtained by coupling one particle to appropriate $(N-1)$-particle states. The N-particle states, which have a total angular momentum J, are denoted by

$$|j^{N-1}(\alpha' Y' J')jJ\rangle. \tag{6.1}$$

Here Y' is the Yamanouchi symbol characterizing the $(N-1)$-particle state with respect to the sequence of commuting operators

$$\{C^2[SU_{2j+1}(N')], N' = 1,2,...,N-1\}, \tag{6.2}$$

J' is the $(N-1)$-particle angular momentum, and α' is the collection of all additional quantum numbers, including the seniority. No further labels are necessary since from a given $(N-1)$-particle state we obtain a unique N-particle state with a definite total angular momentum.

In Sec. III it was shown that $C^2[SU_{2j+1}(N)]$ commutes with both the N-particle angular momentum operator and with the sequence (6.2). Consequently, its eigenstates must be linear combinations of all the N-particle states of the form (6.2) with a given J and Y'. The coefficients in this linear combinations are the desired cfp's.

To evaluate the matrix elements of $C^2[SU_{2j+1}(N)]$ we start from Eq. (3.8) and treat each one of the three terms on the right-hand side of that equation separately.

(a) For the basis set chosen above, Eq. (6.1), $C^2[SU_{2j+1}(N-1)]$ is diagonal. Hence

$$\begin{aligned}&\langle j^{N-1}(\alpha' Y' J')jJ|C^2[SU_{2j+1}(N-1)]\\ &\quad\times|j^{N-1}(\alpha'' Y'' J'')jJ\rangle\\ &\quad=\delta_{Y'Y''}\delta_{\alpha'\alpha''}\delta_{J'J''}\\ &\qquad\times\langle\Gamma'_{N-1}|C^2[SU_{2j+1}(N-1)]|\Gamma'_{N-1}\rangle,\end{aligned} \tag{6.3}$$

where the eigenvalue of the quadratic Casimir operator for a well-defined eigenstate, appearing in the rhs, is given in Eq. (8.1). This eigenvalue depends only on Γ'_{N-1}, the $(N-1)$-particle Young shape, and not on the whole Yamanouchi symbol Y'.

(b) The matrix element of the scalar product $(U^k(N-1)\cdot u^k(N))$ is evaluated using (Ref. 4, p. 522)

$$\begin{aligned}&\langle j^{N-1}(\alpha' Y' J')jJ|(U^k(N-1)\cdot u^k(N))\\ &\quad|j^{N-1}(\alpha'' Y'' J'')jJ\rangle\\ &\quad=(-1)^{j+J+J''}\begin{Bmatrix}J' & j & J\\ j & J'' & k\end{Bmatrix}\\ &\qquad\times(j^{N-1}\alpha' Y' J'||U^k(N-1)||j^{N-1}\alpha'' Y'' J'')\\ &\qquad\times(j||u^k(N)||j).\end{aligned} \tag{6.4}$$

The one-particle reduced matrix element satisfies $(j||u^k||j) = (-1)^k\sqrt{2j+1}$. To evaluate the $(N-1)$-particle reduced matrix element we first note that $U^k(N-1)$, which is one of the generators of $SU_{2j+1}(N)$, commutes with $C^2[SU_{2j+1}(N-1)]$. Therefore, this matrix element vanishes if Y' and Y'' correspond to different irreps. The expression for this matrix element follows from the general result for a one-body reduced matrix element, Eq. (4.8), from which

$$\begin{aligned}&(j^{N-1}\alpha' Y' J'||U^k(N-1)||j^{N-1}\alpha'' Y'' J'')\\ &\quad=\delta_{Y'Y''}\frac{N-1}{n_{\Gamma'_{N-1}}}\sum_{\tilde{Y}\in\Gamma'_{N-1}}(j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J'\\ &\qquad||u^k(N-1)||j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J'').\end{aligned} \tag{6.5}$$

Here $\tilde{Y}$, which is an $(N-2)$-particle Yamanouchi symbol equivalent to a sequence of Young frames of the form $\Gamma_1\Gamma_2\cdots\Gamma_{N-2}$, ranges over the basis states for the the irrep Γ'_{N-1}.

The reduced matrix element in the sum in Eq. (6.5) is expressed in terms of cfp's from $N-2$ to $N-1$ as follows:

$$\begin{aligned}&(j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J'||u^k(N-1)||j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J'')\\ &\quad=\sum_{\beta'\Lambda'\beta''\Lambda''}[j^{N-2}(\beta'\tilde{Y}\Lambda')jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J']\\ &\qquad\times[j^{N-2}(\beta''\tilde{Y}\Lambda'')jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J'']\\ &\qquad\times(j^{N-2}(\beta'\tilde{Y}\Lambda')jJ'||u^k(N-1)\\ &\qquad||j^{N-2}(\beta''\tilde{Y}\Lambda'')jJ''),\end{aligned} \tag{6.6}$$

and by using Ref. 4, p. 552, we obtain

$$\begin{aligned}&(j^{N-2}(\beta'\tilde{Y}\Lambda')jJ'||u^k(N-1)||j^{N-2}(\beta''\tilde{Y}\Lambda'')jJ'')\\ &\quad=(-1)^{\Lambda'+j+J'+k}\sqrt{(2J'+1)(2J''+1)}\\ &\qquad\times(j||u^k(N-1)||j)\begin{Bmatrix}j & J' & \Lambda'\\ J'' & j & k\end{Bmatrix}\delta_{\beta'\beta''}\delta_{\Lambda'\Lambda''},\end{aligned} \tag{6.7}$$

where Λ' and Λ'' are the $(N-2)$-particle angular momenta.

By subtituting Eqs. (6.5)–(6.7) in Eq. (6.4), one obtains

$$\begin{aligned}&\langle j^{N-1}(\alpha' Y' J')jJ|(U^k(N-1)\cdot u^k(N))\\ &\quad|j^{N-1}(\alpha'' Y'' J'')jJ\rangle\\ &\quad=(-1)^{2j+J+J'+J''+k}\begin{Bmatrix}J' & j & J\\ j & J'' & k\end{Bmatrix}(2j+1)\delta_{Y'Y''}\\ &\qquad\times\frac{N-1}{n_{\Gamma'_{N-1}}}\sqrt{(2J'+1)(2J''+1)}\\ &\qquad\times\sum_{\tilde{Y}\in\Gamma'_{N-1}}\sum_{\beta\Lambda}(-1)^{\Lambda}[j^{N-2}(\beta\tilde{Y}\Lambda)jJ'|\}\\ &\qquad j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J']\\ &\qquad\times[j^{N-2}(\beta\tilde{Y}\Lambda)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J'']\\ &\qquad\times\begin{Bmatrix}j & J' & \Lambda\\ J'' & j & k\end{Bmatrix}.\end{aligned} \tag{6.8}$$

Using Eq. (6.8) we find that the sum over k in the second term in Eq. (3.8) reduces to

$$\begin{aligned}&\sum_{k=1}^{2j}(2k+1)(-1)^k\begin{Bmatrix}J' & j & J\\ j & J'' & k\end{Bmatrix}\begin{Bmatrix}j & J' & \Lambda\\ J'' & j & k\end{Bmatrix}\\ &\quad=(-1)^{J+\Lambda}\begin{Bmatrix}J' & j & J\\ J'' & j & \Lambda\end{Bmatrix}\\ &\qquad-\delta_{J'J''}\frac{(-1)^{2J'+2j+J+\Lambda}}{(2J'+1)(2j+1)}.\end{aligned} \tag{6.9}$$

Using Eq. (6.9), the total contribution of the second term in

Eq. (3.8) is

$$2(-1)^{2j+J+J'+J''}\sqrt{(2J'+1)(2J''+1)}\,\delta_{Y'Y''}\frac{N-1}{n_{\Gamma'_{N-1}}}$$
$$\times\sum_{\widetilde{Y}\in\Gamma'_{N-1}}\sum_{\beta\Lambda}(-1)^{\Lambda}[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ'|\}$$
$$j^{N-1}\alpha'\Gamma'_{N-1}\widetilde{Y}J']$$
$$\times[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\widetilde{Y}J'']$$
$$\times\left\{(-1)^{J+\Lambda}\begin{Bmatrix}J' & j & J\\ J'' & j & \Lambda\end{Bmatrix}-\delta_{J'J''}\frac{(-1)^{2J'+2j+J+\Lambda}}{(2J'+1)(2j+1)}\right\}. \tag{6.10}$$

The contribution involving the $\delta_{J'J''}$ factor in Eq. (6.10) can be simplified by the use of the orthogonality property of the cfp's (Ref. 4, p. 522):

$$\delta_{J'J''}\delta_{Y'Y''}2(-1)^{4j+4J'+2J}\frac{1}{2j+1}\frac{N-1}{n_{\Gamma'_{N-1}}}$$
$$\times\sum_{\widetilde{Y}}\sum_{\beta\Lambda}(-1)^{2\Lambda}[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\widetilde{Y}J']$$
$$\times[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ'|\}j^{N-1}\alpha''\Gamma'_{N-1}\widetilde{Y}J']$$
$$=\delta_{J'J''}\delta_{Y'Y''}\delta_{\alpha'\alpha''}[2(N-1)/(2j+1)]. \tag{6.11}$$

Finally, the contribution of the second term in Eq. (3.8) to the matrix element of $C^2[\mathrm{SU}_{2j+1}(N)]$ is

$$2(-1)^{2j+2J+J'+J''}\sqrt{(2J'+1)(2J''+1)}\,\delta_{Y'Y''}$$
$$\times\left[(N-1)/n_{\Gamma'_{N-1}}\right]\sum_{\widetilde{Y}\in\Gamma'_{N-1}}\sum_{\beta\Lambda}(-1)^{2\Lambda}$$
$$\times[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\widetilde{Y}J']$$
$$\times[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\widetilde{Y}J'']\begin{Bmatrix}J' & j & J\\ J'' & j & \Lambda\end{Bmatrix}$$
$$-\delta_{J'J''}\delta_{Y'Y''}\delta_{\alpha'\alpha''}[2(N-1)/(2j+1)]. \tag{6.12}$$

(c) The matrix element of the third term in Eq. (3.8) is diagonal in the quantum numbers of the $(N-1)$-particle state. Using Ref. 4, p. 522, we obtain

$$\langle j^{N-1}(\alpha'Y'J')jJ|u_0^0(N)|j^{N-1}(\alpha''Y''J'')jJ\rangle$$
$$=\delta_{Y'Y''}\delta_{\alpha'\alpha''}\delta_{J'J''}. \tag{6.13}$$

Adding the contributions of the three terms, Eqs. (6.3), (6.12), and (6.13), we obtain the matrix element of $C^2[\mathrm{SU}_{2j+1}(N)]$ in the form

$$\langle j^{N-1}(\alpha'Y'J')jJ|C^2[\mathrm{SU}_{2j+1}(N)]|j^{N-1}(\alpha''Y''J'')jJ\rangle$$
$$=\delta_{Y'Y''}\delta_{\alpha'\alpha''}\delta_{J'J''}\left[\langle j^{N-1}\alpha'Y'J'|C^2[\mathrm{SU}_{2j+1}(N-1)]|j^{N-1}\alpha''Y''J''\rangle+\frac{4j(j+1)-2(N-1)}{2j+1}\right]$$
$$+\delta_{Y'Y''}\frac{2N-1}{n_{\Gamma'_N}}(-1)^{2j+J'+J''}$$
$$\times\sqrt{(2J'+1)(2J''+1)}\sum_{\widetilde{Y}\in\Gamma'_{N-1}}\sum_{\beta\Lambda}\begin{Bmatrix}J & j & J'\\ \Lambda & j & J''\end{Bmatrix}$$
$$\times[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\widetilde{Y}J']$$
$$\times[j^{N-2}(\beta\widetilde{Y}\Lambda)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\widetilde{Y}J']. \tag{6.14}$$

For the antisymmetric representation this expression reduces to Eq. (19) in Bayman and Lande.[7]

The following observations can now be made concerning the matrix of $C^2[\mathrm{SU}_{2j+1}(N)]$. First, as expected, this matrix is diagonal in the eigenvalues of $C^2[\mathrm{SU}_{2j+1}(N-1)]$, which commutes with $C^2[\mathrm{SU}_{2j+1}(N)]$. Second, the matrix elements depend on the representation Γ'_{N-1} of $C^2[\mathrm{SU}_{2j+1}(N-1)]$, but they do not depend on the particular internal state $\widetilde{Y}$ within this representation. This means that the eigenvectors, i.e., the $N-1$ to N cfp's, depend on the Young frames Γ_N and Γ_{N-1} but not on the complete sequence of frames $\Gamma_1\Gamma_2\Gamma_3\cdots\Gamma_{N-2}$. This will turn out to result in a considerable simplification in the determination of the phase consistent sets of cfp's for degenerate irreps, discussed in Sec. X. Even more significantly, this property is crucial in the construction leading to the definition of cfp's for L–S (or L–S–T, etc.) coupled states, as discussed in Sec. XI. In view of this property we denote the cfp's by $[j^{N-1}(\alpha'\Gamma_{N-1}J')jJ|\}j^N\alpha\Gamma_NJ]$ and define the renormalized cfp's:

$$[j^{N-1}(\alpha'\Gamma_{N-1}J')jJ|\}j^N\alpha\Gamma_NJ]_r$$
$$=\sqrt{n_{\Gamma_{N-1}}/n_{\Gamma_N}}\,[j^{N-1}(\alpha'\Gamma_{N-1}J')jJ|\}j^N\alpha\Gamma_NJ]. \tag{6.15}$$

These renormalized cfp's are slightly less convenient for the representation of N-particle states, but slightly more convenient for the representation of matrix elements of operators. Replacing the cfp's by the renormalized cfp's in Eq. (6.14), it is modified in two ways: First, the sum over the internal labels $\widetilde{Y}$ of Γ'_{N-1} should be replaced by a sum over the Young frames Γ'_{N-2} obtained from Γ'_{N-1} by deleting one box; and second, division of the sum by the degeneracy factor has been absorbed into the renormalization. As a consequence of these modifications the expression for the matrix element involves unique contributions only and looks more similar to the standard expressions for matrix elements of antisymmetric states in terms of corresponding cfp's.

VII. MATRIX ELEMENTS OF $C^2[\mathrm{Sp}_{2j+1}(N)]$ AND $C^2[\mathrm{SO}_{2j+1}(N)]$

The matrix elements of the operators referred to in this section's heading, which we also denote collectively by $C^2[R_{2j+1}(N)]$, are easily obtained using expressions derived in the previous section. Referring to Eq. (3.10) we note that the first and third terms in the expression for the matrix element of $C^2[R_{2j+1}(N)]$ are easily written down using the corresponding results for the $C^2[\mathrm{SU}_{2j+1}(N)]$ [Eqs. (6.3) and (6.13)]. Their sum is

$$\delta_{Y'Y''}\delta_{\alpha'\alpha''}\delta_{J'J''}[\langle j^{N-1}\alpha'Y'J'|C^2[R_{2j+1}]|j^{N-1}\alpha''Y''J''\rangle+(3j-2[j])], \tag{7.1}$$

where $[j]$ is the integral part of j. Note that the last term is equal to j for integral j (SO_{2j+1}) and to $j+1$ for half-integral j (Sp_{2j+1}).

To obtain the contribution of the second term in Eq. (3.10) we use the results in Eqs. (6.4)–(6.8). The sum over k, which contains odd terms only, can be expressed in the

form

$$\sum_{k=1,3,\ldots}^{2j}(2k+1)(-1)^k\begin{Bmatrix}J' & j & J\\ j & J'' & k\end{Bmatrix}\begin{Bmatrix}j & J' & \Lambda\\ J'' & j & k\end{Bmatrix}$$
$$=\frac{1}{2}\left[-\frac{\delta_{J,\Lambda}}{2J+1}+(-1)^{J+\Lambda}\begin{Bmatrix}J' & j & J\\ J'' & j & \Lambda\end{Bmatrix}\right]. \quad (7.2)$$

Using this result the total contribution of the second term in Eq. (3.10) can be separated into that of the Kronecker delta term

$$-(-1)^{2j+2J+J'+J''}\frac{\sqrt{(2J'+1)(2J''+1)}}{2J+1}\delta_{Y'Y''}$$
$$\times\frac{N-1}{n_{\Gamma'_{N-1}}}\sum_{\tilde{Y}\in\Gamma'_{N-1}}\sum_{\beta}[j^{N-2}(\beta\tilde{Y}J)jJ'|\}$$
$$j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J']$$
$$\times[j^{N-2}(\beta\tilde{Y}J)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J''], \quad (7.3)$$

and that of the term containing the $6j$ symbol:

$$(-1)^{2j+2J+J'+J''}\sqrt{(2J'+1)(2J''+1)}\,\delta_{Y'Y''}$$
$$\times\frac{N-1}{n_{\Gamma'_{N-1}}}\sum_{\tilde{Y}\in\Gamma'_{N-1}}\sum_{\beta\Lambda}(-1)^{2\Lambda}$$
$$\times[j^{N-2}(\beta\tilde{Y}\Lambda)jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J']$$
$$\times[j^{N-2}(\beta\tilde{Y}\Lambda)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J'']$$
$$\times\begin{Bmatrix}J' & j & J\\ J'' & j & \Lambda\end{Bmatrix}. \quad (7.4)$$

The phase factors can be simplified by noting that $2(J+\Lambda)$ and $2(j+J+J'')$ are always even. Adding the contributions of Eqs. (7.1), (7.3), and (7.4), we obtain

$$\langle j^{N-1}(\alpha'Y'J')jJ|C^2[R_{2j+1}(N)]|j^{N-1}(\alpha''Y''J'')jJ\rangle$$
$$=\delta_{Y'Y''}\{\delta_{\alpha'\alpha''}\delta_{J'J''}[\langle j^{N-1}\alpha'Y'J'|C^2[R_{2j+1}(N-1)]|j^{N-1}\alpha''Y''J''\rangle+(3j-2[j])]$$
$$+(-1)^{2j+J'+J''}\sqrt{(2J'+1)(2J''+1)}\frac{N-1}{n_{\Gamma'_{N-1}}}\sum_{\tilde{Y}\in\Gamma'_{N-1}}\sum_{\beta\Lambda}\begin{Bmatrix}J' & j & J\\ J'' & j & \Lambda\end{Bmatrix}[j^{N-2}(\beta\tilde{Y}\Lambda)jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J']$$
$$\times[j^{N-2}(\beta\tilde{Y}\Lambda)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J'']-(-1)^{J'-J''}\frac{\sqrt{(2J'+1)(2J''+1)}}{2J+1}\frac{N-1}{n_{\Gamma'_{N-1}}}$$
$$\times\sum_{\tilde{Y}\in\Gamma'_{N-1}}\sum_{\beta}[j^{N-2}(\beta\tilde{Y}J)jJ'|\}j^{N-1}\alpha'\Gamma'_{N-1}\tilde{Y}J'][j^{N-2}(\beta\tilde{Y}J)jJ''|\}j^{N-1}\alpha''\Gamma'_{N-1}\tilde{Y}J''\}. \quad (7.5)$$

For antisymmetric states this result reduces to Eq. (13b) of Bayman and Lande.[7]

VIII. IDENTIFICATION OF THE IRREPS VIA THE EIGENVALUES OF THE CASIMIR OPERATORS

A Lie algebra of rank l has l Casimir operators that collectively label its irreps. However, in our procedure we are only using the quadratic Casimir operators of the SU_{2j+1} and R_{2j+1} (i.e., Sp_{2j+1} or SO_{2j+1}) groups. The special circumstances that enable the identification of the irreps using only these operators are closely associated with the recursive buildup of states being employed.

An irrep of SU_{2j+1} is labeled by the Young frame $\{\lambda_1,\lambda_2,\ldots,\lambda_{2j+1}\}$, where $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{2j+1}\geqslant 0$ and $\Sigma_{i=1}^{2j+1}\lambda_i=N$. The corresponding eigenvalue of $C^2[SU_{2j+1}(N)]$ is

$$\sum_{i=1}^{2j+1}\lambda_i(\lambda_i-2i+1)+\frac{N(2j+1)^2-N^2}{2j+1}. \quad (8.1)$$

Similarly, for an irrep of Sp_{2j+1}, labeled by $\langle\lambda_1,\lambda_2,\ldots,\lambda_{j+1/2}\rangle$, the eigenvalue of $C^2[Sp_{2j+1}(N)]$ is

$$\frac{1}{2}\sum_{i=1}^{j+1/2}\lambda_i(\lambda_i+2j-2i+3), \quad (8.2)$$

and for an irrep of SO_{2j+1}, labeled by $[\lambda_1,\lambda_2,\ldots,\lambda_j]$, the eigenvalue of $C^2[SO_{2j+1}(N)]$ is

$$\frac{1}{2}\sum_{i=1}^{j}\lambda_i(\lambda_i+2j-2i+1). \quad (8.3)$$

For the SU_{2j+1} group the situation is particularly straightforward because at each stage we add one box in an allowed position to a well-defined Young frame. Starting from the frame specified by $\{\lambda_1,\lambda_2,\ldots,\lambda_{2j+1}\}$, we have to show that any two Young frames generated from it by an allowed addition of one box have different values of the quadratic Casimir operator.

Compare the frame obtained by increasing λ_k into $\tilde{\lambda}_k=\lambda_k+1$ to the one obtained by increasing λ_l into $\tilde{\lambda}_l=\lambda_l+1$. It is assumed, of course, that $\lambda_k<\lambda_{k-1}$ and $\lambda_l<\lambda_{l-1}$. Using Eq. (8.1) we equate the eigenvalues of the quadratic Casimir operator of the two $(N+1)$-particle Young frames formed above and obtain $\lambda_k-\lambda_l=k-l$. This equality cannot hold, because if $k>l$ then $\lambda_k<\lambda_l$. This establishes the uniqueness of the characterization of the relevant irreps of SU_{2j+1} in terms of the eigenvalues of the quadratic Casimir operator.

For the symplectic group the situation is somewhat more involved because the number of boxes can either increase or decrease by 1, on going from the N- to the $(N+1)$-particle system. Moreover, unlike the SU_{2j+1} case, the symplectic irrep labels for the N-particle system are not good quantum numbers for the $(N+1)$-particle system. How-

ever, the $(N+1)$-particle symplectic irrep can be uniquely identified by inspecting the symplectic irrep of any N-particle state contributing to it.

In analogy with the argument presented above for the special unitary group, two shapes, which are both obtained by subtraction or both by addition from a common N-particle shape, correspond to different eigenvalues of the symplectic Casimir operator. Let us consider the frame obtained by addition of a box to λ_k and the frame obtained by subtraction of a box from λ_l. Clearly, this assumes that $\lambda_k < \lambda_{k-1}$ and that $\lambda_l > \lambda_{l+1}$. Equating the expressions for the eigenvalues of the quadratic Casimir operator, Eq. (8.2), for these two frames we obtain

$$\lambda_k + \lambda_l + 2j + 3 = k + l.$$

This equality cannot hold because the maximum number of rows in a symplectic Young frame is $j+\frac{1}{2}$ so that $k+l<2j+1$.

For the orthogonal group one can analogously establish that two Young frames obtained from an N-particle state by the addition or subtraction of one box in any allowed way have different eigenvalues of the quadratic Casimir operator. As mentioned in Sec. V, one of the $(N+1)$-particle irreps of the orthogonal group that can be generated from N-particle irrep with j rows has the same Young frame as the one it was generated from. The corresponding eigenvalue can also be shown to differ from the eigenvalues of all the other $N+1$ frames obtained from the given N-particle frame.

Stronger statements concerning the sets of irreps generated for the special unitary, symplectic, and orthogonal groups can be deduced from the expressions for the corresponding quadratic Casimir operators, Eqs. (8.1)–(8.3). Suppose $[\lambda]$ represents a partition identifying an irrep of SO_{2j+1} and $[\lambda_{+i}]$, $[\lambda_{-i}]$ represent partitions with one box added to the ith row or removed from it. If $C^2[\lambda]$ represents the quadratic Casimir operator SO_{2j+1}, we have the strict inequalities

$$\begin{aligned} C^2[\lambda_{-1}] &< C^2[\lambda_{-2}] < \cdots < C^2[\lambda_{-(j-1)}] \\ &< C^2[\lambda_{-j}] < C^2[\lambda] < C^2[\lambda_{+j}] < C^2[\lambda_{+(j-1)}] \\ &< \cdots < C^2[\lambda_{+2}] < C^2[\lambda_{+1}] . \end{aligned} \tag{8.4}$$

Similar inequalities hold for the Clebsch–Gordan series of $\langle\lambda\rangle \otimes \square$ and $\{\lambda\} \oplus \square$ corresponding to the symplectic and special unitary group, respectively. For the former Eq. (8.4) has to be modified by eliminating the term $C^2[\lambda]$. For the latter, only the terms with added boxes are present. These inequalities imply the uniqueness of the identification by means of the eigenvalue of the quadratic Casimir operator, and provide a convenient framework for carrying this identification out.

IX. ILLUSTRATIVE EXAMPLES

To illustrate the application of the results of the previous three sections we consider in some detail a sequence of simple cases. The examples presented illustrate the evaluation of the matrix elements of the quadratic Casimir operators of the special unitary, symplectic, and orthogonal groups, the calculation of the cfp's and the identification of the various irreps. In Secs. IX A and IX B we present general results for one- and two-particle states, respectively. In Secs. IX C and IX D we evaluate the matrix elements of the Casimir operator of the special unitary group for three- and four-particle states with $j=1$, identify the corresponding irreps, and obtain the cfp's. In Sec. IX E we illustrate the classification by means of the seniority quantum number for four-particle states with $j=\frac{3}{2}$.

A. One particle, arbitrary j

For a single particle, Eqs. (6.14) and (7.5) give

$$\langle j|C^2[SU_{2j+1}(1)]|j\rangle = 4j(j+1)/(2j+1), \tag{9.1}$$

$$\langle j|C^2[R_{2j+1}(1)]|j\rangle = j + [1-(-1)^{2j}]/2. \tag{9.2}$$

Equation (9.1) agrees with the case $\{1\}$ of Eq. (8.1). Equation (9.2) agrees with the case $\langle 1\rangle$ of Eq. (8.2) for half-integral j and with the case $[1]$ of Eq. (8.3) for integral j.

B. Two particles, arbitrary j

For two particles,

$$\begin{aligned} &\langle j^2 J|C^2[SU_{2j+1}(2)]|j^2 J\rangle \\ &\quad = \frac{8j(j+1)-2}{2j+1} + 2(2j+1)\begin{Bmatrix} J & j & j \\ 0 & j & j \end{Bmatrix} \\ &\quad = \frac{8j(j+1)-2}{2j+1} + 2(-1)^{2j-J}, \end{aligned} \tag{9.3}$$

$$\begin{aligned} &\langle j^2 J|C^2[R_{2j+1}(2)]|j^2 J\rangle \\ &\quad = 2j+1-(-1)^{2j} + (2j+1)\begin{Bmatrix} j & j & J \\ j & j & 0 \end{Bmatrix} \\ &\qquad - (2j+1)\delta_{J,0} \\ &\quad = 2j+1-(-1)^{2j} + (-1)^{2j-J} - (2j+1)\delta_{J,0} . \end{aligned} \tag{9.4}$$

The values of $C^2[SU_{2j+1}]$ for the two-particle irreps $[2]$ and $[1^2]$, as determined from Eq. (8.1), are in agreement with the values obtained in Eq. (9.3) for $2j-J$ even and odd, respectively. This is a well-known elementary result. For half-integral j, the relevant irreps of the symplectic group are $\langle 2\rangle$, $\langle 1^2\rangle$, and $\langle 0\rangle$. The corresponding eigenvalues, obtained from Eq. (8.2), are $2j+3$, $2j+1$, and 0, respectively. These are the values obtained from Eq. (9.4) for J odd, even, and zero, respectively. For integral $j>1$ the relevant irreps of the orthogonal group are $[2]$, $[1^2]$, and $[0]$. The corresponding eigenvalues, from Eq. (8.3), are $2j+1$, $2j-1$, and 0, in agreement with the value obtained from Eq. (9.4) for J even, odd, and zero, respectively. As discussed in Sec. V, for $j=1$, the Young frames can consist of only one row; the frame $[1^2]$ is modified into $[1]$. The eigenvalue of $C^2[SO_{2j+1}(2)]$ [Eq. (8.3)] is 1, corresponding to the state $|1^2, J=1\rangle$ [cf. Eq. (9.4)]. Thus this state is a two-particle state whose seniority index is equal to 1.

C. Three particles, $j=1$

Coupling one more particle to the two-particle states $|1^2\{2\}0\rangle$, $|1^2\{1^2\}1\rangle$, and $|1^2\{2\}2\rangle$, we construct the following three-particle states: $|1^2(\{2\}0)1, J=1\rangle$, $|1^2(\{1^2\}1)1, J=0,1,2\rangle$, and $|1^2(\{2\}2)1, J=1,2,3\rangle$. The only value of J for which a two-dimensional subspace has to be diagonalized

is $J=1$, for which there are two states belonging to the Young frame $\{2\}$. The $C^2[SU_3(3)]$ matrix in this subspace is

$$\begin{pmatrix} \frac{28}{3} & \frac{4}{3}\sqrt{5} \\ \frac{4}{3}\sqrt{5} & \frac{26}{3} \end{pmatrix},$$

with the eigenvalues 12 and 6 corresponding, according to Eq. (8.1), to $\{3\}$ and $\{2,1\}$, respectively. The corresponding eigenvectors provide the cfp's as follows:

$$|1^3\{3\},\ J=1\rangle = (\sqrt{5}/3)|1^2(\{2\}0)1,\ J=1\rangle + \tfrac{2}{3}|1^2(\{2\}2)1,\ J=1\rangle, \tag{9.5a}$$

$$|1^3\{2,1\},\ J=1\rangle = -\tfrac{2}{3}|1^2(\{2\}0)1,\ J=1\rangle + (\sqrt{5}/3)|1^2(\{2\}2)1,\ J=1\rangle. \tag{9.5b}$$

The irrep $\{3\}$ is one dimensional, but $\{2,1\}$ is two dimensional. The other $J=1$ state belonging to $\{2,1\}$ is $|1^2(\{1^2\}1)1, J=1\rangle$. The relative phase of these two states has to be determined in a manner discussed in the following section.

Each one of the other three-particle states is obtained from a unique two-particle state; the corresponding cfp's are equal to 1. Evaluating the expectation values of $C^2[SU_3(3)]$ for these three-particle states we identify the corresponding irreps as follows:

$$\langle 1^2(\{2\}2)1,\ J=2|C^2[SU_3(3)]|1^2(\{2\}2)1,\ J=2\rangle = 6 \Rightarrow \{2,1\}, \tag{9.6a}$$

$$\langle 1^2(\{2\}2)1,\ J=3|C^2[SU_3(3)]|1^2(\{2\}2)1,\ J=3\rangle = 12 \Rightarrow \{3\}, \tag{9.6b}$$

$$\langle 1^2(\{1^2\}1)1,\ J=0|C^2[SU_3(3)]|1^2(\{1^2\}1)1,\ J=0\rangle = 0 \Rightarrow \{1^3\}, \tag{9.6c}$$

$$\langle 1^2(\{1^2\}1)1,\ J=2|C^2[SU_3(3)]|1^2(\{1^2\}1)1,\ J=2\rangle = 6 \Rightarrow \{2,1\}. \tag{9.6d}$$

D. Four particles, $J=1$

Some of the four-particle states are obtained uniquely by coupling of the fourth particle to appropriate three-particle states. The expectation values of the $SU_3(4)$ Casimir operator for these states can be used to identify the irreps. The results are

$$\langle 1^3(\{3\}3)1,\ J=4|C^2[SU_3(4)]|1^3(\{3\}3)1,\ J=4\rangle = \tfrac{56}{3} \Rightarrow \{4\}, \tag{9.7a}$$

$$\langle 1^3(\{3\}1)1,\ J=0|C^2[SU_3(4)]|1^3(\{3\}1)1,\ J=0\rangle = \tfrac{56}{3} \Rightarrow \{4\}, \tag{9.7b}$$

$$\langle 1^3(\{3\}1)1,\ J=1|C^2[SU_3(4)]|1^3(\{3\}1)1,\ J=1\rangle = \tfrac{32}{3} \Rightarrow \{3,1\}, \tag{9.7c}$$

$$\langle 1^3(\{3\}3)1,\ J=3|C^2[SU_3(4)]|1^3(\{3\}3)1,\ J=3\rangle = \tfrac{32}{3} \Rightarrow \{3,1\}, \tag{9.7d}$$

$$\langle 1^3(\{2,1\}2)1,\ J=3|C^2[SU_3(4)]|1^3(\{2,1\}2)1,\ J=3\rangle = \tfrac{32}{3} \Rightarrow \{3,1\}, \tag{9.7e}$$

$$\langle 1^3(\{2,1\}1)1,\ J=0|C^2[SU_3(4)]|1^3(\{2,1\}1)1,\ J=0\rangle = \tfrac{20}{3} \Rightarrow \{2^2\}, \tag{9.7f}$$

$$\langle 1^3(\{1^3\}0)1,\ J=1|C^2[SU_3(4)]|1^3(\{1^3\}0)1,\ J=1\rangle = \tfrac{8}{3} \Rightarrow \{2,1^2\}. \tag{9.7g}$$

The state $|1^4\{3,1\}3\rangle$ is triply degenerate. Two of the three members of this degenerate state are obtained from $|1^3\{2,1\}2\rangle$, and the third is obtained from $|1^3\{3\}3\rangle$. While the first two states have identical cfp's, the phase of the third state relative to the first two has yet to be determined (cf. Sec. X). A complete labeling of these three states would involve the sequence of Young frames $\{1\}\{2\}\{3\}\{3,1\}$, $\{1\}\{2\}\{2,1\}\{3,1\}$, and $\{1\}\{1^2\}\{2,1\}\{3,1\}$. Similarly, $|1^4\{2^2\}0\rangle$ is doubly degenerate, the complete labels for the two components being $\{1\}\{2\}\{2,1\}\{2^2\}$ and $\{1\}\{1^2\}\{2,1\}\{2^2\}$. The states $|1^4\{3,1\}0\rangle$, $|1^4\{3,1\}1\rangle$, and $|1^4\{2,1\}1\rangle$ are also triply degenerate but the above list contains only one member of each of these basis sets. The construction of the other two members of each basis requires the diagonalization of appropriate 2×2 matrices of the $SU_3(4)$ Casimir operator.

The simplest case involving the diagonalization of a 2×2 matrix results in

$$|1^4\{4\}2\rangle = \sqrt{\tfrac{3}{10}}\,|1^3(\{3\}3)1,\ J=2\rangle + \sqrt{\tfrac{7}{10}}\,|1^3(\{3\}1)1,\ J=2\rangle, \tag{9.8a}$$

$$|1^4\{3,1\}2\rangle = \sqrt{\tfrac{7}{10}}\,|1^3(\{3\}3)1,\ J=2\rangle - \sqrt{\tfrac{3}{10}}\,|1^3(\{3\}1)1,\ J=2\rangle. \tag{9.8b}$$

In this case the three-particle states involved belong to the nondegenerate symmetric representation $\{3\}$.

A more interesting case involves the pair of states $|1^3(\{2,1\}2)1,\ J=1\rangle$ and $|1^3(\{2,1\}1)1,\ J=1\rangle$. In this case the three-particle states are doubly degenerate and the evaluation of the off-diagonal matrix element of $C^2[SU_3(4)]$ requires that the relative phases of the two components of these two representations be determined consistently. This problem was mentioned in connection with the

construction of these three-particle states, but it is here that it becomes crucially relevant. The appropriate procedure is discussed in Sec. X.

E. An illustration of the seniority classification

In all the cases encountered so far the resultant angular momentum and the sequence of realizations of the quadratic Casimir operator for the SU_{2j+1} group provide a complete classification of the states. While the states discussed could have been (and some were) characterized with respect to their seniority, this was unnecessary for their classification. To encounter a simple example in which the seniority classification is needed, we consider a system of particles with $j=\frac{5}{2}$. The results for one and two particles were already presented in Secs. IX A and IX B. For three particles the states are still completely classified by $C^2[SU_4(3)]$ and the total angular momentum.

The four-particle states $|(\frac{5}{2})^3(\{3\}_{\frac{5}{2}})_{\frac{5}{2}}, J=1\rangle$ and $|(\frac{5}{2})^3(\{3\}_{\frac{3}{2}})_{\frac{3}{2}}, J=1\rangle$ present a case in which classification by means of the symplectic Casimir operator is necessary. These two states both belong to the $\{3,1\}$ irrep of $SU_4(4)$, as one finds by evaluating the matrix of the appropriate Casimir operator. However, upon diagonalizing $C^2[Sp_4(4)]$ we obtain the seniority 4 state $\langle 3,1\rangle$ whose Casimir operator eigenvalue is 12, as well as the seniority 2 state $\langle 2\rangle$ whose eigenvalue is 6. The corresponding eigenstates are

$$|(\tfrac{5}{2})^4\langle 3,1\rangle\{3\}\{3,1\},\ J=1\rangle = \sqrt{\tfrac{7}{15}}\,|(\tfrac{5}{2})^3(\{3\}_{\frac{5}{2}})_{\frac{5}{2}},\ J=1\rangle + \sqrt{\tfrac{8}{15}}\,|(\tfrac{5}{2})^3(\{3\}_{\frac{3}{2}})_{\frac{3}{2}},\ J=1\rangle, \tag{9.9a}$$

$$|(\tfrac{5}{2})^4\langle 2\rangle\{3\}\{3,1\},\ J=1\rangle = \sqrt{\tfrac{8}{15}}\,|(\tfrac{5}{2})^3(\{3\}_{\frac{5}{2}})_{\frac{5}{2}},\ J=1\rangle - \sqrt{\tfrac{7}{15}}\,|(\tfrac{5}{2})^3(\{3\}_{\frac{3}{2}})_{\frac{3}{2}},\ J=1\rangle \tag{9.9b}$$

X. GENERATION OF CONSISTENT BASES FOR S_N REPRESENTATIONS

The procedure for the generation of the N-particle wave functions and cfp's proposed in the preceding sections is recursive. Assuming that the states for $N-1$ particles have been constructed, we form the set of N-particle states $|j^{N-1}(\alpha' \mathbf{v}' Y' J')jJ\rangle$ for all α' $\mathbf{v}'$ and J' allowed for given Y' and J. While these states have well-defined angular momenta, they do not, in general, belong to irreps of the special unitary and symplectic (or orthogonal) groups. Diagonalizing the matrices for the Casimir operators $C^2[SU_{2j+1}(N)]$ and $C^2[R_{2j+1}(N)]$ within the space specified above, we obtain the set of states $|j^N \alpha \mathbf{v} Y J\rangle$. Recall that Y is a Yamanouchi symbol equivalent to the sequence of Young frames $\Gamma_1\Gamma_2\cdots\Gamma_N$ and Y' is equivalent to $\Gamma_1\Gamma_2\cdots\Gamma_{N-1}$.

Basis sets for irreps of S_N labeled by Yamanouchi symbols can be split into subsets, each one of which is specified by an irrep of S_{N-1}. It was shown in Sec. VII that the $N-1$ to N cfp's depend on Γ_{N-1} and Γ_N, which also specify the irreps of $SU_{2j+1}(N-1)$ and $SU_{2j+1}(N)$. These cfp's do not, however, depend on the internal labels of the Γ_{N-1} irrep. This is a source of considerable simplification, since it implies that we have to generate only one representative of each subset in order to obtain all the cfp's needed to span a particular N-particle irrep. On the other hand, the generation of states labeled by the same Γ_N but different Γ_{N-1} has to be performed in a consistent way. This guarantees that the corresponding representation matrices conform to some standard form.

To clarify the issue discussed above let us consider the irrep $\{3,1\}$ of $SU_{2j+1}(4)$. The three-dimensional irrep $(3,1)$ of S_N associated with this representation is spanned by the states

$$\begin{aligned}&(1)(2)(3)(3,1)\leftrightarrow(2111),\\ &(1)(2)(2,1)(3,1)\leftrightarrow(1211),\\ &(1)(1^2)(2,1)(3,1)\leftrightarrow(1121).\end{aligned} \tag{10.1}$$

The Yamanouchi symbols of the sequences of Young frames in (10.1) are denoted to the right of the corresponding sequences. In this example the first state belongs to the one-dimensional irrep (3) of S_3 and the other two belong to the two-dimensional irrep $(2,1)$. The states belonging to the second and third sequences have the same cfp's. However, these cfp's have to be generated with the correct phase relative to the cfp's of the state belonging to the first sequence.

Recalling that all the relevant $(N-1)$-particle Young frames are generated by the removal of one box from the N-particle frame, it will now be demonstrated that we only have to consider transpositions $(N-1,N)$ of the last two indices in order to construct a complete phase-consistent set of cfp's. Let $Y_1=\Gamma_1\Gamma_2\cdots\Gamma_N$ be a Yamanouchi symbol of the form $(ab\cdots)$ and Y_2 a Yamanouchi symbol of the form $(ba\cdots)$, related to Y_1 by the transposition $(N-1,N)$. Let us further assume that $N-1$ and N appear in different rows and columns. The state labeled by Y_2 is generated from the corresponding state labeled by Y_1 with the standard phase convention, by using the relation[11]

$$Y_2=(1/\sqrt{1-\sigma_{rs}^2})\,[(N,N-1)-\sigma_{rs}]\,Y_1, \tag{10.2}$$

where r and s are the indices of the rows in which N and $N-1$ are placed in Y_1, λ_r and λ_s are the lengths of these rows, and $\sigma_{rs}=1/(\lambda_r-\lambda_s+s-r)$. A Yamanouchi symbol having $N-1$ and N in either the same row or the same column is an eigenstate of the transposition $(N-1,N)$, and is consequently of no interest in the present context.

The state $|j^N\alpha Y_1 J\rangle$, where α contains the seniority label, is given in terms of the appropriate cfp's:

$$|j^N\alpha Y_1 J\rangle=\sum_{\alpha' J'}[j^{N-1}(\alpha'\Gamma_{N-1}J')jJ|\}|j^N\alpha\Gamma_N J]\times|j^{N-1}(\alpha' Y_1' J')jJ\rangle. \tag{10.3}$$

We would like to determine the cfp's for the state $|j^N\alpha Y_2 J\rangle$ by means of Eq. (10.2). This is achieved by writing the original state, Eq. (10.3), using the $N-2$ to $N-1$ cfp's and a recoupling transformation. As a consequence, the particles $N-1$ and N are coupled into well-defined angular momentum states, which, for two particles, have unique permutational symmetry characters. The resulting expression for the

state (10.3) is

$$
\begin{aligned}
&|j^N\alpha\Gamma_1\cdots\Gamma_{N-2}\Gamma_{N-1}\Gamma_N J\rangle \\
&\quad=\sum_{\alpha' J'}[j^{N-1}(\alpha'\Gamma_{N-1}J')jJ|\}|j^N\alpha\Gamma_N J] \\
&\quad\times\sum_{\alpha'' J''}[j^{N-2}(\alpha''\Gamma_{N-2}J'')jJ'|\}j^{N-1}\alpha'\Gamma_{N-1}J'] \\
&\quad\times(-1)^{2j+J+J''}\sqrt{2J'+1} \\
&\quad\times\sum_{\tilde{J}=0}^{2j}\sqrt{2\tilde{J}+1}\begin{Bmatrix} J'' & j & J' \\ j & J & \tilde{J}\end{Bmatrix} \\
&\quad\times|(j^{N-2}\alpha''\Gamma_1\cdots\Gamma_{N-2}J'')(j^2\Gamma_2(\tilde{J})\tilde{J})J\rangle,
\end{aligned}
\tag{10.4}
$$

where

$$
\Gamma_2(\tilde{J})=\begin{cases}(2), & \tilde{J}=2j,2j-2,\ldots,0 \text{ or } 1\\ (1^2), & \tilde{J}=2j-1,2j-3\ldots,1 \text{ or } 0.\end{cases}
$$

Using the form of the N-particle state obtained in this way, it is a simple matter to apply the operator specified in Eq. (10.2), since $(N-1,N)(2)=(2)$ and $(N-1,N)(1^2)=-(1^2)$.

Application of the operator (10.2) to the state in Eq. (10.4) generates the state $|j^N\alpha Y_2 J\rangle$, expressed in terms of the already available cfp's of the state $|j^N\alpha Y_1 J\rangle$. To extract the cfp's of $|j^N\alpha Y_2 J\rangle$, we note that the sequence of Young frames corresponding to Y_2 differs from that for Y_1 only by the Young frame for $N-1$ particles. Thus, an expression for this state can be obtained by replacing Γ_{N-1} by Γ'_{N-1} in Eq. (10.4). This expression involves the still unknown $N-1$ to N cfp's of the state $|j^N\alpha Y_2 J\rangle$. Equating the coefficients of

$$
|(j^{N-2}\Gamma_1\cdots\Gamma_{N-2}J'')(j^2\Gamma_2(\tilde{J})(\tilde{J}))J\rangle
$$

in the two equivalent expressions for the state $|j^N\alpha Y_2 J\rangle$ whose derivation was described above, we obtain a set of linear equations for the cfp's desired.

To establish that the number of linear equations is sufficient to determine all the cfp's, we note that the number of terms of the form

$$
|(j^{N-2}\alpha''\Gamma_{N-2}J'')(j^2\Gamma_2(\tilde{J})\tilde{J})J\rangle
$$

is equal to the number of ways of getting J from J'' by coupling two more particles, disregarding permutational symmetry. The number of cfp's

$$
[j^{N-1}(\alpha'\Gamma'_{N-1}J')jJ|\}j^N\alpha\Gamma_N J]
$$

is equal to the number of $(N-1)$-particle states from which a particular symmetry N-particle state can be obtained. The $(N-1)$-particle states were obtained from the set of $(N-2)$-particle states $\{|j^{N-2}\alpha''\Gamma_{N-2}J''\rangle\}$ by coupling of one particle and selecting according to the permutational symmetry. Therefore, there are at least as many linear equations as there are unknown cfp's.

As a simple illustration we note that starting from Eq. (9.6a) and using Eq. (10.4), we obtain

$$
\begin{aligned}
|1^3\{2\}\{2,1\},\ J=2\rangle &= |1^2\{2\}2)1,\ J=2\rangle \\
&= \tfrac{1}{2}|(1^1\{1\}1)(1^2\{2\}2),\ J=2\rangle + (\sqrt{3}/2)|(1^1\{1\}1)(1^2\{1^2\}1),\ J=2\rangle.
\end{aligned}
\tag{10.5}
$$

Application of Eq. (10.2) yields

$$
\begin{aligned}
|1^3\{1^2\}\{2,1\},\ J=2\rangle &= (2/\sqrt{3})[(2,3)+\tfrac{1}{2}]|1^3\{2\}\{2,1\},\ J=2\rangle \\
&= (\sqrt{3}/2)|(1^1\{1\}1)(1^2\{2\}2),\ J=2\rangle - \tfrac{1}{2}|(1^1\{1\}1)(1^2\{1^2\}1),\ J=2\rangle.
\end{aligned}
\tag{10.6}
$$

To extract the cfp's for the latter state we use Eqs. (9.6d) and (10.4) to write

$$
\begin{aligned}
|1^3\{1^2\}\{2,1\},\ J=2\rangle &= [(1^2\{1^2\}1)1,\ J=2|\}1^3\{1^2\}\{2,1\},\ J=2]|1^2(\{1^2\}1)1,\ J=2\rangle \\
&= [(1^2\{1^2\}1)1,\ J=2|\}1^3\{1^2\}\{2,1\},\ J=2] \\
&\quad\times((\sqrt{3}/2)|(1^1\{1\}1)(1^2\{2\}2),\ J=2\rangle - \tfrac{1}{2}|(1^1\{1\}1)(1^2\{1^2\}1),\ J=2\rangle).
\end{aligned}
\tag{10.7}
$$

Comparison of the last two equations shows that $[(1^2\{1^2\}1)1, J=2|\}1^3\{1^2\}\{2,1\}, J=2]=1$. In this case it was only the phase of the cfp that had to be determined.

A somewhat more interesting case is obtained by using Eq. (10.4) to yield

$$
\begin{aligned}
|1^3\{1^2\}\{2,1\},\ J=1\rangle &= |(1^2\{1^2\}1)1,\ J=1\rangle \\
&= -(\sqrt{3}/3)|(1^1\{1\}1)(1^2\{2\}0),\ J=1\rangle + (\sqrt{15}/6)|(1^1\{1\}1)(1^2\{2\}2),\ J=1\rangle \\
&\quad+\tfrac{1}{2}|(1^1\{1\}1)(1^2\{1^2\}1),\ J=1\rangle.
\end{aligned}
\tag{10.8}
$$

Applying Eq. (10.2),

$$
\begin{aligned}
|1^3\{2\}\{2,1\},\ J=1\rangle &= (2/\sqrt{3})((2,3)-\tfrac{1}{2})|1^3\{1^2\}\{2,1\},\ J=1\rangle \\
&= -\tfrac{1}{3}|(1^1\{1\}1)(1^2\{2\}0),\ J=1\rangle + (\sqrt{5}/6)|(1^1\{1\}1)(1^2\{2\}2),\ J=1\rangle \\
&\quad-(\sqrt{3}/2)|(1^1\{1\}1)(1^2\{1^2\}1),\ J=1\rangle.
\end{aligned}
\tag{10.9}
$$

On the other hand, using Eq. (10.4),

$$
\begin{aligned}
|1^3\{2\}\{2,1\},\ J=1\rangle &= [1^2(\{2\}0)1,\ J=1|\}1^3\{2\}\{2,1\},\ J=1]|(1^2\{2\}0),\ J=1\rangle \\
&\quad + [1^2(\{2\}2)1,\ J=1|\}1^3\{2\}\{2,1\},\ J=1]|(1^2\{2\}2)1,\ J=1\rangle \\
&= [1^2(\{2\}0)1,\ J=1|\}1^3\{2\}\{2,1\},\ J=1](\tfrac{1}{3}|(1^1\{1\}1)(1^2\{2\}0),\ J=1\rangle \\
&\quad + (\sqrt{5}/3)|(1^1\{1\}1)(1^2\{2\}2),\ J=1\rangle - (1/\sqrt{3})|(1^1\{1\}1)(1^2\{1^2\}1),\ J=1\rangle) \\
&\quad + [1^2(\{2\}2)1,\ J=1|\}1^3\{2\}\{2,1\},\ J=1](\sqrt{5}/3)|(1^1\{1\}1)(1^2\{2\}0),\ J=1\rangle \\
&\quad + \tfrac{1}{6}|(1^1\{1\}1)(1^2\{2\}2),\ J=1\rangle + (\sqrt{15}/6)|(1^1\{1\}1)(1^2\{1^2\}1),\ J=1\rangle). \qquad (10.10)
\end{aligned}
$$

Equating coefficients of appropriate terms in Eqs. (10.9) and (10.10), we obtain the (redundent) set of linear equations

$$
\begin{pmatrix} \frac{1}{3} & \sqrt{5}/3 \\ \sqrt{5}/3 & \frac{1}{6} \\ -1/\sqrt{3} & \sqrt{15}/6 \end{pmatrix}
\begin{pmatrix} [1^2(\{2\}0)1,\ J=1|\}1^3\{2\}\{2,1\},\ J=1] \\ [1^2(\{2\}2)1,\ J=1|\}1^3\{2\}\{2,1\},\ J=1] \end{pmatrix}
= \begin{pmatrix} -\frac{1}{3} \\ \sqrt{5}/6 \\ -\sqrt{3}/2 \end{pmatrix}. \qquad (10.11)
$$

These equations have a consistent solution which yields the state

$$
|1^3\{2\}\{2,1\},\ J=1\rangle = \tfrac{2}{3}|1^2(\{2\}0)1,\ J=1\rangle - (\sqrt{5}/3)|1^2(\{2\}2)1,\ J=1\rangle. \qquad (10.12)
$$

Note that the same state, with reversed phase, is presented in Eq. (9.5b). The phase presently obtained is consistent with that of the state presented by Eq. (10.8). These two states span the (2,1) irrep.

XI. THE *L–S* COUPLED WAVE FUNCTION

To obtain the L–S coupled wave function with any desired permutational symmetry, we couple an L and an S state using the appropriate Clebsch–Gordan (CG) coefficients for the symmetric group. Consider a state consisting of N particles, each one having an orbital angular momentum l and a spin s. The orbital angular momenta are coupled into a resultant angular momentum L and into a permutational irrep Γ^l_N, and the spins are coupled into a resultant spin S and a permutational irrep Γ^s_N. The state with a total permutational symmetry Γ_N and internal state label μ_N can be written in the form

$$
|(ls)^N\alpha^l\alpha^s\Gamma^l_N\Gamma^s_N\Gamma_N\mu_N LS\rangle = \sum_{\mu^l_N\mu^s_N} (\Gamma^l_N\mu^l_N\Gamma^s_N\mu^s_N|\Gamma^l_N\Gamma^s_N\Gamma_N\mu_N)|l^N\alpha^l\Gamma^l_N\mu^l_N L;s^N\alpha^s\Gamma^s_N\mu^s_N S\rangle. \qquad (11.1)
$$

Since we are usually interested in totally antisymmetric (or totally symmetric) wave functions, we note that the corresponding CG coefficients have the very simple forms[11]

$$
\begin{aligned}
&(\Gamma^l_N\mu^l_N\Gamma^s_N\mu^s_N|\Gamma^l_N\Gamma^s_N(1^N)) = \delta_{\Gamma^l_N\mu^l_N,(\Gamma^s_N\mu^s_N)^\dagger}(1/\sqrt{n_{\Gamma^l_N}})\Lambda_{\Gamma^l_N\mu^l_N}, \\
&(\Gamma^l_N\mu^l_N\Gamma^s_N\mu^s_N|\Gamma^l_N\Gamma^s_N(N)) = \delta_{\Gamma^l_N\mu^l_N,\Gamma^s_N\mu^s_N}(1/\sqrt{n_{\Gamma^l_N}}).
\end{aligned} \qquad (11.2)
$$

Here, $\Lambda_{\Gamma^l_N\mu^l_N}$ is a phase factor. It is equal to 1 when the parity of the permutation from the highest Yamanouchi symbol (i.e., the Yamanouchi symbol represented by the largest "number") in the irrep Γ^l_N to the Yamanouchi symbol of the state with an internal label μ^l_N in that irrep is even, and to -1 otherwise. Here $(\Gamma_N\mu_N)^\dagger$ denotes the state conjugate to $\Gamma_N\mu_N$. Thus if $|\Gamma_N\mu_N\rangle = |\Gamma_1\Gamma_2\cdots\Gamma_N\rangle$ then $|(\Gamma_N\mu_N)^\dagger\rangle = |\Gamma^\dagger_1\Gamma^\dagger_2\cdots\Gamma^\dagger_N\rangle$, where $\Gamma^\dagger_i$ is the Young frame obtained from Γ_i by interchanging rows and columns. The integer $n_{\Gamma^l_N}$ is the dimension of the irrep Γ^l_N. The internal state label μ_N for the symmetric and antisymmetric irreps obtains a single value and was, therefore, suppressed in the CG coefficients in Eq. (11.2).

Using the L and S $N-1$ to N cfp's, we write the state in the rhs of Eq. (11.1) in the form

$$
\begin{aligned}
&|l^N\alpha^l\Gamma^l_N\mu^l_N L;s^N\alpha^s\Gamma^s_N\mu^s_N S\rangle \\
&\quad = \sum_{\alpha'^lL'\alpha'^sS'} [l^{N-1}(\alpha'^l\Gamma^l_{N-1}L')lL\,|\}l^N\alpha^l\Gamma^l_N L\,][s^{N-1}(\alpha'^s\Gamma^s_{N-1}S')sS\,|\}s^N\alpha^s\Gamma^s_N S\,] \\
&\quad\quad \times|l^{N-1}(\alpha'^l\Gamma^l_{N-1}\mu^l_{N-1}L')lL;s^{N-1}(\alpha'^s\Gamma^s_{N-1}\mu^s_{N-1}S')sS\rangle.
\end{aligned} \qquad (11.3)
$$

Substituting Eq. (11.3) in Eq. (11.1), setting $\Gamma^l_N\mu^l_N = (\Gamma^s_N\mu^s_N)^\dagger$, and using the CG coefficient for the totally antisymmetric state in (11.2), we obtain

$$
\begin{aligned}
&|(ls)^N\alpha^l\alpha^s\Gamma^l_N(\Gamma^l_N)^\dagger(1^N)LS\rangle \\
&\quad = \frac{1}{\sqrt{n_{\Gamma^l_N}}}\sum_{\mu^l_N}\Lambda_{\Gamma^l_N\mu^l_N}|l^N\alpha^l\Gamma^l_N\mu^l_N L;s^N\alpha^s(\Gamma^l_N\mu^l_N)^\dagger S\rangle \\
&\quad = \frac{1}{\sqrt{n_{\Gamma^l_N}}}\sum_{\mu^l_N\alpha'^lL'\alpha'^sS'}\Lambda_{\Gamma^l_N\mu^l_N}[l^{N-1}(\alpha'^l\Gamma^l_{N-1}L')lL\,|\}l^N\alpha^l\Gamma^l_N L\,][s^{N-1}(\alpha'^s(\Gamma^l_{N-1})^\dagger S')sS\,|\}s^N\alpha(\Gamma^l_N)^\dagger S\,] \\
&\quad\quad \times|l^{N-1}(\alpha'^l\Gamma^l_{N-1}\mu^l_{N-1}L')lL;s^{N-1}(\alpha'^s(\Gamma^l_{N-1}\mu^l_{N-1})^\dagger S')sS\rangle.
\end{aligned} \qquad (11.4)
$$

In order to introduce the L–S coupled cfp's we have to write the L–S coupled N-particle function in terms of L–S coupled $(N-1)$-particle functions. We recall that the L and the S cfp's for all the states spanning the same irrep of the $(N-1)$-particle system are equal. Using this property we split the sum over μ^l_N in Eq. (11.4) into a summation over the irreps Γ^l_{N-1} belonging to Γ^l_N, and over their internal labels μ^l_{N-1}. Each N-particle CG coefficient can be written as an $(N-1)$-particle CG coefficient multiplied by a factor which depends only on the N- and $(N-1)$-particle irreps and not on their internal labels. Using the factorization, we obtain

$$|(ls)^N\alpha^l\alpha^s\Gamma^l_N(\Gamma^l_N)^\dagger(1^N)LS\rangle = \sum_{\alpha'^l\alpha'^sL'S'\Gamma'^l_{N-1}} [(ls)^{N-1}(\alpha'^l\alpha'^s\Gamma'^l_{N-1}(\Gamma'^l_{N-1})^\dagger L'S'(1^{N-1}))lsLS|\}$$
$$(ls)^N\alpha^l\alpha^s\Gamma^l_N(\Gamma^l_N)^\dagger(1^N)LS]|(ls)^{N-1}(\alpha'^l\alpha'^s\Gamma'^l_{N-1}(\Gamma'^l_{N-1})^\dagger L'S'(1^{N-1}))lsLS\rangle. \quad (11.5)$$

In this expression the N-particle L–S state is written as a linear combination of $(N-1)$- to N-particle L–S coupled cfp's multiplying appropriate N-particle wave functions. The L–S coupled cfp's are related to the L and S cfp's by

$$[(ls)^{N-1}(\alpha'^l\alpha'^s\Gamma'^l_{N-1}(\Gamma'^l_{N-1})^\dagger L'S'(1^{N-1}))lsLS|\}(ls)^N\alpha^l\alpha^s\Gamma_N(\Gamma^l_N)^\dagger(1^N)LS]$$
$$= \sqrt{n_{\Gamma'^l_{N-1}}/n_{\Gamma^l_N}}\Lambda_{\Gamma^l_N\Gamma'^l_{N-1}}[l^{N-1}(\alpha'^l\Gamma'^l_{N-1}L')lL|\}l^N\alpha^l\Gamma^l_NL][s^{N-1}(\alpha'^s(\Gamma'^l_{N-1})^\dagger S')sS|\}s^N\alpha^s(\Gamma^l_N)^\dagger S]. \quad (11.6)$$

Here $\Lambda_{\Gamma^l_N\Gamma'^l_{N-1}}$ is the parity of the permutation from the highest Yamanouchi symbol in Γ^l_N to the highest Yamanouchi symbol in the subset of the Γ^l_N basis which spans the Γ'^l_{N-1} irrep. In obtaining the phase factor in Eq. (11.6) we use the straightforward relation

$$\Lambda_{\Gamma^l_N\mu^l_N} \equiv \Lambda_{\Gamma^l_N\Gamma^l_{N-1}\mu^l_{N-1}} = \Lambda_{\Gamma^l_N\Gamma^l_{N-1}}\cdot\Lambda_{\Gamma^l_{N-1}\mu^l_{N-1}}. \quad (11.7)$$

The phase factor $\Lambda_{\Gamma^l_N\Gamma^l_{N-1}}$ is easily evaluated as follows: Assume that the Young frame corresponding to Γ^l_{N-1} is obtained from that corresponding to Γ^l_N by deleting a box in row a. The set of N-particle Yamanouchi symbols with a as their leading index spans a basis set for Γ^l_{N-1}. The parity of the permutation from the highest symbol in that set to the highest symbol in Γ^l_N is equal to the number of indices larger than a, or, equivalently, to the total number of boxes in rows $a+1$ and down, in either Γ^l_N or Γ^l_{N-1}.

As a simple example of the determination of the phase factor $\Lambda_{\Gamma^l_N\Gamma^l_{N-1}}$, consider the two-dinensional irrep $\Gamma^l_3 = (2,1)$. The Yamanouchi symbols spanning the basis set for this irrep are (211) and (121). The first spans the $\Gamma^l_2 = (2)$ irrep and the second spans the $\Gamma^l_2 = (1^2)$ irrep upon deletion of a box from the row denoted by the leading (leftmost) index. Since for the symbol (121), the index of the box deleted first is $a=1$, there is one box with $a>1$, so that the phase of this state relative to the state (211) (whose Yamanouchi symbol is the largest) is -1. Similarly, the three-dimensional irrep (3,1) is spanned by (2111), (1211), and (1121). The first state, whose Yamanouchi symbol is the largest of the three, spans the irrep (3) upon deletion of the leading index, and the other two span the irrep (2,1). Of the two (2,1) states, (1211) is higher, and it can be obtained from (2111) by one transposition, i.e., $\Lambda_{(3,1)(2,1)} = -1$.

To illustrate the evaluation of the L–S coupled cfp's for a totally antisymmetric state, consider the three-particle state with $l=s=1$, $L=S=1$, and $\Gamma^l_3=(2,1)$. Substituting the appropriate L (S) cfp's [Eqs. (10.8) and (10.12)] in Eq. (11.6), we obtain the following L–S cfp's:

$$[(11)^2((2)0(1^2)1)(11)11|\}$$
$$(11)^3(2,1)(2,1)(1^3)11] = \sqrt{2}/3, \quad (11.8a)$$

$$[(11)^2((2)2(1^2)1)(11)11|\}$$
$$(11)^3(2,1)(2,1)(1^3)11] = -\sqrt{5}/3\sqrt{2}, \quad (11.8b)$$

$$[(11)^2((1^2)1(2)1)(11)11|\}$$
$$(11)^3(2,1)(2,1)(1^3)11] = -\sqrt{2}/3, \quad (11.8c)$$

$$[(11)^2((1^2)1(2)1)(11)11|\}$$
$$(11)^3(2,1)(2,1)(1^3)11] = \sqrt{5}/3\sqrt{2}. \quad (11.8d)$$

The factorization of the N-particle CG coefficient of the symmetric group into a product of an $(N-1)$-particle CG coefficient and a factor which only depends on the N- and $(N-1)$-particle irreps holds for arbitrary total symmetries [Ref. 11, Eq. (7-226)]. This enables the immediate generalization of Eqs. (11.5) and (11.6) to introduce L–S coupled cfp's for arbitrary total permutational symmetries. These, in turn, can be used to construct L–S–T coupled states and to calculate their cfp's.

XII. THE COMPUTATIONAL PROCEDURE

A computer code implementing the formalism described above for the evaluation of the $L(S)$ and L–S coupled cfp's was developed. The input parameters accepted by the code are the number of particles N and the individual l and s values. As a preliminary step the program constructs all Young frames with up to N boxes. The Young frames are ordered according to the number of particles. For a given number of particles, we follow a reversed lexicographic order, i.e., a Young frame precedes any other Young frame with a smaller number of boxes in the first row in which they are not equal. This ordering is illustrated in Fig. 3. For each Young frame the dimension of the corresponding irrep of S_N is evaluated. The eigenvalues of the quadratic Casimir operators of the special unitary and symplectic or orthogonal groups are also evaluated, using Eqs. (8.1)–(8.3).

The following steps are indicated in Fig. 1. First, the cfp's are calculated for L and S separately. The procedure employed in this step is recursive, starting from the one-particle cfp $[j^0(0)j, j|\}j^1j] = 1$, where $j=l$ or s. Assuming that all the cfp's up to $N-1$ particles were calculated, the

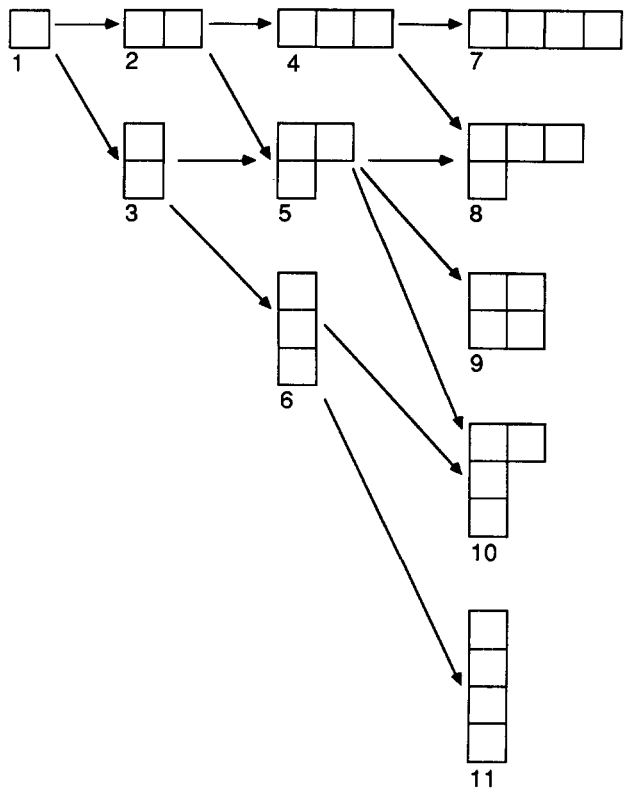


FIG. 3. Branching diagram for the Young frames, up to four particles. The arrows indicate Young frames connected by the addition of one box. The ordinal numbers of the Young frames, explained in the text, are written below each frame.

program calculates the $(N-1)$- to N-particle cfp's as follows: For each $(N-1)$-particle state, labeled by its angular momentum J', permutational Young frame, and seniority, it constructs sets of N-particle states with specific total angular momenta $J = L$ or S, obtained by coupling of one more particle. For each such set of states the code constructs and diagonalizes the matrix of the operator $C^2[\mathrm{SU}_{2j+1}(N)]$, using Eq. (6.14). The eigenstates obtained belong to well-defined irreps $\{\lambda\}_N$ of $\mathrm{SU}_{2j+1}(N)$. The $\mathrm{SU}_{2j+1}(N)$ irrep of each eigenstate is identified by comparison of its eigenvalue with the list of eigenvalues of $C^2[\mathrm{SU}_{2j+1}(N)]$ for all the frames which can be constructed from the original $(N-1)$-particle frame by adding one box. This is facilitated by the fact, demonstrated in Sec. VIII, that each of the eigenvalues in the above list is unique.

The step of the computation described above exploits the following two features.

(1) The diagonalization of the matrices for the quadratic Casimir operator of the special unitary group generates states that belong simultaneously to the irreps $\{\lambda\}_1,\ldots,\{\lambda\}_{N-1}$, $\{\lambda\}_N$ of the realizations

$$C^2[\mathrm{SU}_{2j+1}(1)],\ldots,C^2[\mathrm{SU}_{2j+1}(N-1)],$$
$$C^2[\mathrm{SU}_{2j+1}(N)].$$

(2) The Weyl duality between the special unitary and the symmetric group, discussed in Sec. II, enables the identification of these states with states which belong simultaneously to the irreps $\Gamma_1,\ldots,\Gamma_{N-1},\Gamma_N$ of the symmetric groups $S_1,\ldots,S_{N-1},S_N$, where $\Gamma_{N'} = \{\lambda\}_{N'}$, $N' = 1,\ldots,N-1,N$. We are primarily interested in the permutational symmetry properties of N-particle states—these are extracted from our knowledge of their properties under the special unitary group.

In the procedure as so far described we start from a set of states belonging to a particular irrep Γ_{N-1} and obtain a state belonging to the irrep Γ_N. While the original irrep Γ_{N-1} could have a dimension larger than 1, all the cfp's connecting the states spanning that irrep to the irrep Γ_N are equal (cf. Sec. VI). If, however, the dimension of Γ_N is larger than that of Γ_{N-1}, then the procedure described in Sec. X for the generation of a complete set of cfp's with mutually consistent phase relations is invoked.

The N-particle states obtained at this stage have definite total angular momenta and belong to definite irreps of $\mathrm{SU}_{2j+1}(N)$ [as well as of all the realizations $\mathrm{SU}_{2j+1}(N')$ with $N' = 1,2,\ldots,N-1$]. Each one of these states is a linear combination of $N-1$ to N cfp's multiplied by appropriate N-particle states, of the form of Eq. (5.1). These states are assembled into sets, each containing all the states with the same total angular momentum and the same irrep of SU_{2j+1}. The further classification by means of the seniority quantum number is accomplished by diagonalizing the matrix of $C^2[R_{2j+1}(N)]$ within the set of states with a particular Γ_{N-1} and using the transformation matrix generated for all the other sets of states with the same Γ_N, as well. The identification of the irrep of $R_{2j+1}(N)$ for the newly formed eigenstates is achieved by an analogous procedure to that described above for the SU_{2j+1} irreps. The group theoretical basis of this procedure is presented in Sec. VIII for the symplectic and the orthogonal groups. The states generated at this stage are of the form of Eq. (5.2).

In Appendices A and B we present the renormalized cfp's computed using the code described above, for up to three particles, for $j = 1$ and $j = \frac{3}{2}$, respectively. The states generated for each value of j are numbered consecutively, starting with the one-particle state. Each state is labeled by its total angular momentum, the Young diagram specifying its irrep with respect to SU_{2j+1} (and S_N), its seniority index and the Young diagram specifying its irrep with respect to R_{2j+1}. Each cfp is labeled by a set of quantum numbers for the $(N-1)$-particle state, analogous to the set of quantum numbers specifying the N-particle state. The appendices list renormalized cfp's as defined in Eq. (6.15). The cfp for each Γ_{N-1} is listed once.

The last step in the code is displayed in Fig. 2. It involves the enumeration of the L–S states and the evaluation of their L–S coupled cfp's. The L–S coupled cfp's are obtained by means of Eq. (11.6), using the L and S cfp's calculated before. For totally antisymmetric L–S states the limitation on the number of rows in the Young diagrams for l imposes a limitation in the number of columns in s, and vice versa. Our procedure allows the inclusion of these limitations, so as to calculate only the L and S cfp's required.

Appendix C lists the cfp's for the totally antisymmetric L–S states obtained from the states in Appendices B and A, interpreted as L and S states, respectively. Each L–S state is labeled by the sets of L and S quantum numbers mentioned above. So is each cfp.

The computer code (which is available upon request) has so far been tested on a VAX-750 computer and has been optimized. It has been demonstrated to be very rapid in comparison with existing state-of-the-art codes. A comparison with the improved GENESIS code[5] was carried out for $l = 3/2$ and $s = 1$, for four- and five-particle systems. For the four-particle case, our program yields all the L–S cfp's in less than 1 min CPU time on the VAX-750, compared to about 15 min with the GENESIS code. For five particles the respective times are 3 and 90 min. These results suggest that the relative improvement achieved with the present code increases with increasing number of particles, and also for higher values of l and s. It follows that the present code offers the possibility to considerably extend the size and complexity of the systems that can realistically be studied.

XIII. CONCLUSIONS

In the present article we present a procedure for the evaluation of the cfp's for systems consisting of N identical particles with arbitrary permutational symmetry. This procedure is a generalization of the very efficient and powerful method proposed by Bayman and Lande[7] more than 20 years ago for symmetric and antisymmetric N-particle states.

The procedure presented is recursive, generating the symmetrized N-particle states using symmetrized $(N-1)$-particle states. It involves a fortunate combination of a numerical method for the computation of the cfp's and an analytical method for the identification of the corresponding irreps. The numerical method consists of the diagonalization of the matrices of the special unitary and symplectic or orthogonal quadratic Casimir operators, and the analytical method provides an identification of the irreps of the relevant groups.

The identification of the irreps of both the special unitary and the symplectic or othogonal groups is achieved using the eigenvalues of the above matrices and a group theoretical result guaranteeing the uniqueness of these eigenvalaues for the set of N-particle irreps accessible from any particular $(N-1)$-particle state. This is a remarkable circumstance because in general the quadratic Casimir operator is not sufficient for the identification of the irreps. In addition, the state classification achieved is more detailed than is common in similar contexts, because the seniority label we use is the full irrep label of the symplectic or orthogonal group, rather than an index specifying only the total number of boxes in the corresponding irreps.

The computational efficiency of the procedure developed here has been demonstrated by means of a computer code implementing it for the evaluation of the L–S coupled cfp's for totally antisymmetric states. Further extensions, to L–S coupled states with arbitrary total permutational symmetry as well as to L–S–T type coupling states, etc., are feasible and relatively straightforward.

ACKNOWLEDGMENTS

We are grateful to professor Michel Vallieres and Professor Da Hsuan Feng and to Dr. Xiangdong Ji for their interest and very helpful comments. One of the authors (J.K.) wishes to express his gratitude for the kind hospitality extended to him at Drexel University during the academic year 1986/87.

This work is partly supported by the U.S. National Science Foundation and the Drexel University Louis and Bessie Stein Family Foundation.

APPENDIX A: cfp's FOR $j=1$ UP TO THREE PARTICLES

A slightly edited copy of the computer output for the states and cfp's for $j = 1$, up to three particles, is presented below. The states are numbered consecutively and each one of them is labeled by its total angular momentum quantum number J, its permutational Young frame Γ_N, which is denoted in the output by Y, its seniority index v denoted by SE, and the corresponding Young frame ν denoted by YSE. The Young frames are numbered consecutively (cf. Fig. 3). These numbers, along with the explicit symbols for the Young frames, appear for Y and YSE.

The list below presents the renormalized cfp's defined by Eq. (6.15). For each N-particle state, the cfp's from all the irreps Γ_{N-1} obained from Γ_N by subtraction of one box are listed. Each cfp is labeled by the ordinal number of the $(N-1)$-particle state, denoted by INIT, and the quantum numbers of that state: its angular momentum INITJ, the ordinal number of its permutational Young frame, YINIT, its seniority index SINIT, and the ordinal number of the corresponding Young frame YSINIT. In this edited version of the computer output, the squares of the cfp's are presented under the heading CFP**2 in terms of rational fractions. An asterisk denotes that the negative square root has to be taken.

```
  THE STATES FOR   1  PARTICLE
-----------------------------------

   1)   J=  1   Y=  1   (  1)   SE=  1   YSE=  1  [  1]

  CFP**2              INIT      INITJ      YINIT      SINIT      YSINIT
   1                     0          0          0          0           0

  THE STATES FOR   2  PARTICLES
-----------------------------------

   2)   J=  0   Y=  2   (  2)   SE=  0   YSE=  0  [    ]

  CFP**2              INIT      INITJ      YINIT      SINIT      YSINIT
   1                     1          1          1          1           1
```

```
   3)   J= 2   Y= 2   ( 2)   SE= 2   YSE= 2  [ 2]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
   1                  1         1         1         1         1

   4)   J= 1   Y= 3   ( 1 1)   SE= 1   YSE= 1  [ 1]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
   1                  1         1         1         1         1


  THE STATES FOR   3  PARTICLES
 ---------------------------------

   5)   J= 1   Y= 4   ( 3)   SE= 1   YSE= 1  [ 1]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  *5/9               2         0         2         0         0
  *4/9               3         2         2         2         2

   6)   J= 3   Y= 4   ( 3)   SE= 3   YSE= 4  [ 3]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
   1                 3         2         2         2         2

   7)   J= 1   Y= 5   ( 2 1)   SE= 1   YSE= 1  [ 1]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  *2/9               2         0         2         0         0
   5/18              3         2         2         2         2
  *1/2               4         1         3         1         1

   8)   J= 2   Y= 5   ( 2 1)   SE= 2   YSE= 2  [ 2]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
   1/2               3         2         2         2         2
   1/2               4         1         3         1         1

   9)   J= 0   Y= 6   ( 1 1 1)   SE= 0   YSE= 0  [  ]

  CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
   1                 4         1         3         1         1
```

APPENDIX B: cfp's FOR $j=\frac{3}{2}$ UP TO THREE PARTICLES

A slightly edited copy of the computer output for the states and cfp's for $j=\frac{3}{2}$, up to three particles, is presented below. For the notation used see Appendix A.

```
 THE STATES FOR   1  PARTICLE
---------------------------------

  1)   J= 3/2   Y= 1  ( 1)   SE= 1   YSE= 1  < 1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1                  0         0         0         0         0

 THE STATES FOR   2  PARTICLES
---------------------------------

  2)   J= 1   Y= 2  ( 2)   SE= 2   YSE= 2  < 2>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1                  1       3/2         1         1         1

  3)   J= 3   Y= 2  ( 2)   SE= 2   YSE= 2  < 2>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1                  1       3/2         1         1         1

  4)   J= 0   Y= 3  ( 1 1)   SE= 0   YSE= 0  <  >

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1                  1       3/2         1         1         1

  5)   J= 2   Y= 3  ( 1 1)   SE= 2   YSE= 3  < 1 1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1                  1       3/2         1         1         1
```

```
 THE STATES FOR   3  PARTICLES
 ---------------------------------

  6)   J= 3/2   Y=  4  (  3)   SE=  3   YSE=  4  <  3>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
 *7/10                2         1         2         2          2
  3/10                3         3         2         2          2

  7)   J= 5/2   Y=  4  (  3)   SE=  3   YSE=  4  <  3>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
 *8/15                2         1         2         2          2
 *7/15                3         3         2         2          2

  8)   J= 9/2   Y=  4  (  3)   SE=  3   YSE=  4  <  3>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1                   3         3         2         2          2

  9)   J= 1/2   Y=  5  (  2  1)   SE=  3   YSE=  5  <  2  1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1/2                 2         1         2         2          2
 *1/2                 5         2         3         2          3

 10)   J= 3/2   Y=  5  (  2  1)   SE=  1   YSE=  1  <  1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  3/20                2         1         2         2          2
  7/20                3         3         2         2          2
  5/12                4         0         3         0          0
  1/12                5         2         3         2          3

 11)   J= 5/2   Y=  5  (  2  1)   SE=  3   YSE=  5  <  2  1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
 *7/30                2         1         2         2          2
  8/30                3         3         2         2          2
 *1/2                 5         2         3         2          3

 12)   J= 7/2   Y=  5  (  2  1)   SE=  3   YSE=  5  <  2  1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
  1/2                 3         3         2         2          2
  1/2                 5         2         3         2          3

 13)   J= 3/2   Y=  6  (  1  1  1)   SE=  1   YSE=  1  <  1>

 CFP**2            INIT     INITJ     YINIT     SINIT     YSINIT
 *1/6                 4         0         3         0          0
  5/6                 5         2         3         2          3
```

APPENDIX C: *L–S* cfp's FOR $l=\frac{3}{2}$, $s=1$ UP TO THREE PARTICLES

A slightly edited copy of the computer output for the L–S coupled states and cfp's for $l=\frac{3}{2}$, $s=1$, up to three particles is presented below. The states are numbered consecutively, and each one is labeled by the ordinal number and list of quantum numbers for both the L and S factors, as defined in Appendix A. So is each cfp. The notation used is the logical extension of that described in Appendix A.

```
 THE STATES FOR 1 PARTICLE IN L-S
 ---------------------------------

   1)  I=1  L=3/2  Y=1  (  1)  SE=1  YSE=1  ;  I=1  S=1  Y=1  (  1)  SE=1  YSE=1

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               0      0     0     0     0      0      0     0     0     0      0

 THE STATES FOR 2 PARTICLES IN L-S
 ---------------------------------

   2)  I=2  L=1  Y=2  (  2)  SE=2  YSE=2  ;  I=4  S=1  Y=3  (  1  1)  SE=1  YSE=1
```

```
CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               1      1   3/2     1     1       1      1     1     1      1

   3)  I=3  L=3  Y=2  (  2)  SE=2  YSE=2  ;  I=4  S=1  Y=3  (  1  1)  SE=1  YSE=1

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               1      1   3/2     1     1       1      1     1     1      1

   4)  I=4  L=0  Y=3  (  1  1)  SE=0  YSE=0  ;  I=2  S=0  Y=2  (  2)  SE=0  YSE=0

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               1      1   3/2     1     1       1      1     1     1      1

   5)  I=4  L=0  Y=3  (  1  1)  SE=0  YSE=0  ;  I=3  S=2  Y=2  (  2)  SE=2  YSE=2

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               1      1   3/2     1     1       1      1     1     1      1

   6)  I=5  L=2  Y=3  (  1  1)  SE=2  YSE=3  ;  I=2  S=0  Y=2  (  2)  SE=0  YSE=0

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               1      1   3/2     1     1       1      1     1     1      1

   7)  I=5  L=2  Y=3  (  1  1)  SE=2  YSE=3  ;  I=3  L=2  Y=2  (  2)  SE=2  YSE=2

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               1      1   3/2     1     1       1      1     1     1      1


 THE STATES FOR 3 PARTICLES IN L-S
-----------------------------------

   8)  I=6  L=3/2  Y=4  (  3)  SE=3  YSE=4  ;  I=9  S=0  Y=6  (  1  1  1)  SE=0  YSE=0

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
*7/10            2      2     1     2     2       2      4     1     3      1      1
 3/10            3      3     3     2     2       2      4     1     3      1      1

   9)  I=7  L=5/2  Y=4  (  3)  SE=3  YSE=4  ;  I=9  S=0  Y=6  (  1  1  1)  SE=0  YSE=0

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
*8/15            2      2     1     2     2       2      4     1     3      1      1
*7/15            3      3     3     2     2       2      4     1     3      1      1

  10)  I=8  L=9/2  Y=4  (  3)  SE=3  YSE=4  ;  I=9  S=0  Y=6  (  1  1  1)  SE=0  YSE=0

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1               3      3     3     2     2       2      4     1     3      1      1

  11)  I=9  L=1/2  Y=5  (  2  1)  SE=3  YSE=5  ;  I=7  S=1  Y=5  (  2  1)  SE=1  YSE=1

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
*1/2             2      2     1     2     2       2      4     1     3      1      1
*2/9             6      5     2     3     2       3      2     0     2      0      0
 5/18            7      5     2     3     2       3      3     2     2      2      2

  12)  I=9  L=1/2  Y=5  (  2  1)  SE=3  YSE=5  ;  I=8  S=2  Y=5  (  2  1)  SE=2  YSE=2

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 1/2             2      2     1     2     2       2      4     1     3      1      1
 1/2             7      5     2     3     2       3      3     2     2      2      2

  13)  I= 10  L=3/2  Y=5  (  2  1)  SE=1  YSE=1  ;  I=7  S=1  Y=5  (  2  1)  SE=1  YSE=1

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
*3/20            2      2     1     2     2       2      4     1     3      1      1
*7/20            3      3     3     2     2       2      4     1     3      1      1
 5/27            4      4     0     3     0       0      2     0     2      0      0
*25/108          5      4     0     3     0       0      3     2     2      2      2
 1/27            6      5     2     3     2       3      2     0     2      0      0
*5/108           7      5     2     3     2       3      3     2     2      2      2

  14)  I= 10  L=3/2  Y=5  (  2  1)  SE=1  YSE=1  ;  I=8  S=2  Y=5  (  2  1)  SE=2  YSE=2

CFP**2       INITLS ; INIT INITL YINIT SINIT YSINIT ; INIT INITS YINIT SINIT YSINIT
 3/20            2      2     1     2     2       2      4     1     3      1      1
 7/20            3      3     3     2     2       2      4     1     3      1      1
*5/12            5      4     0     3     0       0      3     2     2      2      2
*1/12            7      5     2     3     2       3      3     2     2      2      2
```

15) I= 11 L=5/2 Y=5 (2 1) SE=3 YSE=5 ; I=7 S=1 Y=5 (2 1) SE=1 YSE=1

CFP**2	INITLS	INIT	INITL	YINIT	SINIT	YSINIT	INIT	INITS	YINIT	SINIT	YSINIT
7/30	2	2	1	2	2	2	4	1	3	1	1
*4/15	3	3	3	2	2	2	4	1	3	1	1
*2/9	6	5	2	3	2	3	2	0	2	0	0
5/18	7	5	2	3	2	3	3	2	2	2	2

16) I= 11 L=5/2 Y=5 (2 1) SE=3 YSE=5 ; I=8 S=2 Y=5 (2 1) SE=2 YSE=2

CFP**2	INITLS	INIT	INITL	YINIT	SINIT	YSINIT	INIT	INITS	YINIT	SINIT	YSINIT
*7/30	2	2	1	2	2	2	4	1	3	1	1
8/30	3	3	3	2	2	2	4	1	3	1	1
1/2	7	5	2	3	2	3	3	2	2	2	2

17) I= 12 L=7/2 Y=5 (2 1) SE=3 YSE=5 ; I=7 S=1 Y=5 (2 1) SE=1 YSE=1

CFP**2	INITLS	INIT	INITL	YINIT	SINIT	YSINIT	INIT	INITS	YINIT	SINIT	YSINIT
*1/2	3	3	3	2	2	2	4	1	3	1	1
2/9	6	5	2	3	2	3	2	0	2	0	0
*5/18	7	5	2	3	2	3	3	2	2	2	2

18) I= 12 L=7/2 Y=5 (2 1) SE=3 YSE=5 ; I=8 S=2 Y=5 (2 1) SE=2 YSE=2

CFP**2	INITLS	INIT	INITL	YINIT	SINIT	YSINIT	INIT	INITS	YINIT	SINIT	YSINIT
1/2	3	3	3	2	2	2	4	1	3	1	1
*1/2	7	5	2	3	2	3	3	2	2	2	2

19) I= 13 L=3/2 Y=6 (1 1 1) SE=1 YSE=1 ; I=5 S=1 Y=4 (3) SE=1 YSE=1

CFP**2	INITLS	INIT	INITL	YINIT	SINIT	YSINIT	INIT	INITS	YINIT	SINIT	YSINIT
5/54	4	4	0	3	0	0	2	0	2	0	0
2/27	5	4	0	3	0	0	3	2	2	2	2
*25/54	6	5	2	3	2	3	2	0	2	0	0
*10/27	7	5	2	3	2	3	3	2	2	2	2

20) I= 13 L=3/2 Y=6 (1 1 1) SE=1 YSE=1 ; I=6 S=3 Y=4 (3) SE=3 YSE=4

CFP**2	INITLS	INIT	INITL	YINIT	SINIT	YSINIT	INIT	INITS	YINIT	SINIT	YSINIT
*1/6	4	4	0	3	0	0	3	2	2	2	2
5/6	7	5	2	3	2	3	3	2	2	2	2

[1]R. F. Bacher and S. Goudsmit, Phys. Rev. **46**, 948 (1934).
[2]G. Racah, Phys. Rev. **63**, 367 (1943).
[3]P. J. Redmond, Proc. R. Soc. London Ser. A **222**, 84 (1954).
[4]A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).
[5]X. Ji and M. Vallieres, Phys. Rev. C **35**, 1583 (1987).
[6]J. Ginocchio, Ph.D. thesis, The University of Rochester, 1964.
[7]B. F. Bayman and A. Lande, Nucl. Phys. **77**, 1 (1966).
[8]A. Novoselsky, Ph.D. thesis, The Weizmann Institute of Science, 1986.
[9]C-L. Wu, D. H. Feng, X-G. Chen, J-Q. Chen, and M. W. Guidry, Phys. Lett. B **168**, 313 (1986); Phys. Rev. C **36**, 1157 (1987).
[10]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[11]M. Hammermesh, *Group Theory* (Addison–Wesley, Reading, MA, 1962).
[12]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, NJ, 1946).
[13]D. E. Littlewood, *Theory of Group Characters* (Clarendon, Oxford, 1950).
[14]F. Murnaghan, *The Theory of Group Representations* (Dover, New York, 1963).
[15]M. J. Newell, Proc. R. Irish Acad. **A54**, 153 (1951).
[16]R. Gilmore, J. Math. Phys. **11**, 513 (1970).

A relativistic quantum mechanical harmonic oscillator without space-time variables

F. A. Blood
Physics Department, Rutgers Unversity, Camden, New Jersey 08102

(Received 18 September 1987; accepted for publication 27 January 1988)

Solutions of a linear harmonic oscillatorlike second-order differential equation in a set of complex variables are investigated. This equation has an invariance group homomorphic to the inhomogeneous Lorentz group, IG_4^1. Because of this, it has solutions with particlelike properties—mass, energy, momentum, and spin. These results are interesting for an equation without space-time variables.

I. INTRODUCTION

We will investigate solutions of the linear second-order "harmonic oscillator" differential equation

$$\mathcal{O}\Psi = 0,$$

$$\mathcal{O} = -\left(\frac{\partial}{\partial u_{1i}}\frac{\partial}{\partial v_{2i}} - \frac{\partial}{\partial v_{1i}}\frac{\partial}{\partial u_{2i}} + \frac{\partial}{\partial \bar{u}_{1i}}\frac{\partial}{\partial \bar{v}_{2i}} - \frac{\partial}{\partial \bar{v}_{1i}}\frac{\partial}{\partial \bar{u}_{2i}}\right) + (u_{1i}v_{2i} - v_{1i}u_{2i} + \bar{u}_{1i}\bar{v}_{2i} - \bar{v}_{1i}\bar{u}_{2i}). \tag{1}$$

A summation is implied over repeated indices, with i running from 1 to n. The u_{bi}, v_{bi} $(b = 1,2)$ are complex variables, with the bar denoting complex conjugation. This equation has a large group of transformations, with a subgroup ISL(2) being homomorphic to the inhomogeneous Lorentz group, IG_4^1. The infinitesimal generators of the homogeneous Lorentz group SL(2) (homomorphic to G_4^1) are

$$\begin{aligned}
J_1 &= \frac{1}{2}\left(u_{bi}\frac{\partial}{\partial v_{bi}} + v_{bi}\frac{\partial}{\partial u_{bi}} - \bar{u}_{bi}\frac{\partial}{\partial \bar{v}_{bi}} - \bar{v}_{bi}\frac{\partial}{\partial \bar{u}_{bi}}\right),\\
J_2 &= \frac{i}{2}\left(v_{bi}\frac{\partial}{\partial u_{bi}} - u_{bi}\frac{\partial}{\partial v_{bi}} + \bar{v}_{bi}\frac{\partial}{\partial \bar{u}_{bi}} - \bar{u}_{bi}\frac{\partial}{\partial \bar{v}_{bi}}\right),\\
J_3 &= \frac{1}{2}\left(u_{bi}\frac{\partial}{\partial u_{bi}} - v_{bi}\frac{\partial}{\partial v_{bi}} - \bar{u}_{bi}\frac{\partial}{\partial \bar{u}_{bi}} + \bar{v}_{bi}\frac{\partial}{\partial \bar{v}_{bi}}\right),\\
K_1 &= \frac{i}{2}\left(u_{bi}\frac{\partial}{\partial v_{bi}} + v_{bi}\frac{\partial}{\partial u_{bi}} + \bar{u}_{bi}\frac{\partial}{\partial \bar{v}_{bi}} + \bar{v}_{bi}\frac{\partial}{\partial \bar{u}_{bi}}\right),\\
K_2 &= -\frac{1}{2}\left(v_{bi}\frac{\partial}{\partial u_{bi}} - u_{bi}\frac{\partial}{\partial v_{bi}} - \bar{v}_{bi}\frac{\partial}{\partial \bar{u}_{bi}} + \bar{u}_{bi}\frac{\partial}{\partial \bar{v}_{bi}}\right),\\
K_3 &= \frac{i}{2}\left(u_{bi}\frac{\partial}{\partial u_{bi}} - v_{bi}\frac{\partial}{\partial v_{bi}} + \bar{u}_{bi}\frac{\partial}{\partial \bar{u}_{bi}} - \bar{v}_{bi}\frac{\partial}{\partial \bar{v}_{bi}}\right),
\end{aligned} \tag{2}$$

and those of translations are

$$\begin{aligned}
P_0 &= -v_{1i}\frac{\partial}{\partial \bar{u}_{2i}} + u_{1i}\frac{\partial}{\partial \bar{v}_{2i}} + \bar{v}_{1i}\frac{\partial}{\partial u_{2i}} - \bar{u}_{1i}\frac{\partial}{\partial v_{2i}},\\
P_1 &= -u_{1i}\frac{\partial}{\partial \bar{u}_{2i}} + v_{1i}\frac{\partial}{\partial \bar{v}_{2i}} + \bar{u}_{1i}\frac{\partial}{\partial u_{2i}} - \bar{v}_{1i}\frac{\partial}{\partial v_{2i}},\\
P_2 &= i\left(u_{1i}\frac{\partial}{\partial \bar{u}_{2i}} + v_{1i}\frac{\partial}{\partial \bar{v}_{2i}} + \bar{u}_{1i}\frac{\partial}{\partial u_{2i}} + \bar{v}_{1i}\frac{\partial}{\partial v_{2i}}\right),\\
P_3 &= v_{1i}\frac{\partial}{\partial \bar{u}_{2i}} + u_{1i}\frac{\partial}{\partial \bar{v}_{2i}} - \bar{v}_{1i}\frac{\partial}{\partial u_{2i}} - \bar{u}_{1i}\frac{\partial}{\partial v_{2i}}.
\end{aligned} \tag{3}$$

These infinitesimal generators obey the usual commutation relations of IG_4^1:

$$\begin{aligned}
&[J_i,J_j] = i\epsilon_{ijk}J_k, \quad [J_i,K_j] = i\epsilon_{ijk}K_k,\\
&[K_i,K_j] = -i\epsilon_{ijk}J_k, \quad [J_i,P_j] = i\epsilon_{ijk}P_k,\\
&[K_i,P_j] = i\delta_{ij}P_0, \quad [P_i,P_j] = 0,\\
&[P_0,J_j] = 0, \quad [P_0,P_i] = 0, \quad [P_0,K_i] = -iP_i.
\end{aligned} \tag{4}$$

One reason we are interested in Eq. (1) is that it has two of the primary properties of the theory of quantum mechanics: linearity and Lorentz invariance. This has interesting consequences. We know that solutions of linear equations can be classified according to the irreducible representations of the invariance group of the equation. We also know that the irreducible representations of IG_4^1 are classified according to their mass and spin, and that basis vectors for the representations can be chosen as eigenfunctions of the P_μ. Thus we expect to, and indeed do, find solutions of Eq. (1) with particlelike properties—mass, spin, energy, and momentum.

A second reason for examining this particular operator is to emphasize that it is not necessary to have space and time as independent variables in order to have Lorentz invariance (and the resulting mass, momentum, etc.). This leads us to consider what roles space and time play, if they are not independent variables. We find that there are two possible ways in which they can be used. In one usage, the x_μ clearly serve as labels. In conventional quantum field theory, for example, the x_μ in the field operator $\psi^*(x_\mu)$ is simply a label telling us which operator we are looking at. The same is true for the functions $\psi_0(x_\mu)$ [see Eq. (16)] in Sec. III. The other usage is one in which the x_μ are functions of the independent variables which are conjugate to the P_μ. This role is discussed in Sec. VI.

But the most important reason for considering Eq. (1) is that it provides us with a simple example from which we hope to generalize. We find here that solutions of the equation possess a few of the characteristics (mass, etc.) of our quantum mechanically described physical world. This suggests we consider the hypothesis that there exists a more complex linear equation from which *all* of quantum mechanics follows. Such an equation, if it exists, would have the following properties.

(1) It would be invariant under ISL(2), so that it would have particlelike solutions.

 0022-2488/88/061389-07$02.50

(2) It would have a solution with the properties of the vacuum state.

(3) The vacuum and particlelike states, closed under the linear operator, could be used to form a vector space in which field operators would be defined.

(4) Because of closure the linear operator would be reexpressible in terms of the field operators and would apparently correspond to the action integral in conventional quantum field theory. The linear equation would force the particlelike solutions to interact in such a way that the field operators would have equations of motion identical to the usual equations of quantum field theory.

Another way to state the thesis is to say that the equations of quantum mechanics as we know them are *representations* of a particular linear equation. The vector space used for the representation consists of particlelike functions of the independent variables (analogous to the u, v of our example), so that the kets and bras, in terms of which quantum mechanics is given, are to be thought of as representing functions of the independent variables. Our motivation for suggesting this is the observation that in the momentum representation, all the "variables" of quantum mechanics can be viewed as *labels* on group representation basis vectors. Mass, spin, energy, momentum, and parity are labels associated with ISL(2); color, charge, hadron number, and lepton number with SU(5) (Ref. 1); and flavors with a flavor group. This suggests that one should at least entertain the hypothesis that the variables of quantum mechanics—the labels—are *labeling solutions of a linear equation* (which is given in terms of independent variables whose nature is not currently known).

There are two further general points to be made about this proposal. First, the underlying theory we are searching for is not to be viewed as being in competition with conventional quantum mechanics. It is a theory from which the usual quantum mechanics is to be derived. The usual probability interpretation is still to hold. That is, the average value in state $|\psi\rangle$ of the quantity associated with operator A is $\langle\psi|A\,|\psi\rangle$; and the average value of A, $\langle\psi|A\,|\psi\rangle = |\langle\psi|i\rangle|^2$, for the particular case when it is a projection operator, $A = |i\rangle\langle i|$, for state $|i\rangle$, is to be viewed as the probability that the system is observed in state $|i\rangle$. We also note that this theory is not a "hidden variable" theory in the sense that the phrase is normally used. That is, the independent variables are not variables that would determine the outcome of an experiment if their values were known. They are simply variables such that the vectors representing physical states are functions of them.

Second, this hypothesis forces one to consider the relationship between the physical world—or our mental picture of the physical world—and its mathematical description. This is not the proper place to fully consider this question, but we would like to point out what may be obvious. Our scheme is apparently not compatible with the metaphysical position that particles "exist independently" in space and time, with quantum mechanics being the best possible mathematical description of their behavior. We maintain, however, that the incompatibility is not troublesome because the "independently existing particle" mental picture does not appear to be a necessity; it does not add anything to the interpretation of quantum mechanics which is not already in the mathematics.[2]

The search for *the* underlying equation—assuming one exists—will almost certainly prove difficult. This paper is an early phase of that search in which we consider simple examples to see what the possibilities are. Several single particle equations (i.e., equations with no interactions) were described in a previous publication.[3] However, they all used space-time-like independent variables. The absence of such independent variables here—in what is also a single-particle equation—allows us to see more clearly how the group properties enter. In Sec. II, we give a solution to Eq. (1) which has just the Gaussian form we would expect from a harmonic oscillator problem. Then in Sec. III, we show how to obtain momentum eigenfunctions from the solution of Sec. II by the use of momentum projection operators. This allows us to build up the basis for a mass-m, spin-0 representation of ISL(2). In Sec. IV, we show how to construct a basis for spin-($\neq 0$) representations from the spin-0 basis functions by the use of raising and lowering operators. We look briefly at the construction of vacuumlike states and the attendent normalization problems in Sec. V, and show how to construct x_μ in Sec. VI. Finally, we summarize in Sec. VII.

We now start our investigation of solutions of Eq. (1) by considering a single simple solution.

II. A SIMPLE SOLUTION

There is one solution of Eq. (1) that is relatively easy to find. It is

$$\hat{\psi}_0 = c\exp\left(\sum_b\sum_i z_{0bi}\right), \quad c = \left(\frac{2}{\pi}\right)^{2n}, \qquad z_{0bi} = u_{bi}\bar{u}_{bi} + v_{bi}\bar{v}_{bi}, \tag{5}$$

where c is a normalizing constant. It can be verified that this is a solution by simply putting it into Eq. (1) and performing the derivatives. This solution was found by separation of variables. Let

$$u_i^+ = (u_{1i} + \bar{v}_{2i})/\sqrt{2}, \quad u_i^- = (u_{1i} - \bar{v}_{2i})/\sqrt{2}, \qquad v_i^+ = (v_{1i} - \bar{u}_{2i})/\sqrt{2}, \quad v_i^- = (v_{1i} + \bar{u}_{2i})/\sqrt{2}. \tag{6}$$

Then [with $\Re u = (u + \bar{u})/2$, $\Im u = (u - \bar{u})/2i$]

$$\begin{aligned}\mathcal{O} = \sum_i \Bigg\{ &-\left(\frac{\partial^2}{\partial(\Re u_i^+)^2} + \frac{\partial^2}{\partial(\Im u_i^+)^2} + \frac{\partial^2}{\partial(\Re v_i^+)^2} + \frac{\partial^2}{\partial(\Im v_i^+)^2}\right) \\ &+ \left((\Re u_i^+)^2 + (\Im u_i^+)^2 + (\Re v_i^+)^2 + (\Im v_i^+)^2\right) \\ &+ \left(\frac{\partial^2}{\partial(\Re u_i^-)^2} + \frac{\partial^2}{\partial(\Im u_i^-)^2} + \frac{\partial^2}{\partial(\Re v_i^-)^2} + \frac{\partial^2}{\partial(\Im v_i^-)^2}\right) \\ &- \left((\Re u_i^-)^2 + (\Im u_i^-)^2 + (\Re v_i^-)^2 + (\Im v_i^-)^2\right)\Bigg\}.\end{aligned} \tag{7}$$

Since $\mathcal{O}$ factors, it is easy to find product solutions. If we use

$\exp(-x^2/2)$ as the solution to the eigenvalue problem for each single-variable problem, then the total solution is a product which, when translated back to u_{bi}, v_{bi} variables, is just $\hat{\psi}_0$.

As an aside, we note that the form of Eq. (7) makes it easy to find the full invariance group of $\mathscr{O}$. It is the real group G_{8n}^{4n} (where, to illustrate the notation, G_4^1 is the set of all real transformations which leave $t^2-x^2-y^2-z^2$ invariant).

It is necessary to have a scalar product for functions of the u_{bi}, v_{bi}. We use

$$\langle f|g\rangle = \int_{-\infty}^{+\infty}\cdots\int \prod_{b=1}^{2}\prod_{i=1}^{n} d^4u_{bi}\, \bar{f}g \equiv \int du\, \bar{f}g, \tag{8}$$

where

$$d^4u = d(\Re u)d(\Im u)d(\Re v)d(\Im v). \tag{9}$$

The c of Eq. (5) is found by requiring that $\langle\hat{\psi}_0|\hat{\psi}_0\rangle = 1$ under this definition of the scalar product. If we make a transformation from SL(2) on the dummy variables of integration from u to u',

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}, \tag{10}$$

$$\det(A) = a_{11}a_{22} - a_{12}a_{21} = 1,$$

then

$$d^4u' = d^4u, \tag{11}$$

so that the scalar product is invariant. We find that it is also invariant under translations, $\exp(iP\cdot x)$, so that it is invariant under the full ISL(2).

In the next section, we need to know how z_{0bi} behaves under the Lorentz transformations of Eq. (10). We find that it is invariant under the unitary subgroup, SU(2), of SL(2), and is one of four vectors in a $(\frac{1}{2},\frac{1}{2})$ representation of SL(2). These four vectors transform in the same way as the P_μ of Eq. (3); their form, without subscripts, is

$$\begin{aligned} &z_0 = u\bar{u} + v\bar{v}, \quad z_3 = u\bar{u} - v\bar{v}, \\ &z_1 = u\bar{v} + v\bar{u}, \quad z_2 = i(v\bar{u} - u\bar{v}). \end{aligned} \tag{12}$$

We will show how to construct solutions of Eq. (1) which are basis vectors for irreducible representations of ISL(2) by modifying the $\hat{\psi}_0$ of Eq. (5).

III. SPIN-0 BASIS FUNCTIONS

The particlelike irreducible representations of ISL(2) are labeled by mass and spin. Basis vectors for these representations can be chosen as eigenfunctions of P_μ,

$$P_\mu\psi(p) = p_\mu\psi(p), \tag{13}$$

where the p_μ are numbers. If our functions form a basis for a representation of mass m, then

$$p_\mu p_\mu = p\cdot p = p_0^2 - \mathbf{p}\cdot\mathbf{p} = m^2. \tag{14}$$

The spin is defined using the $\mathbf{p}=0$ vector by

$$\mathbf{J}\cdot\mathbf{J}\psi_j(p_0 = m, \mathbf{p}=0) = j(j+1)\psi_j(m,0), \tag{15}$$

where the $\mathbf{J}$ are defined in Eq. (2). An alternative way of specifying zero spin is to say that $\psi(m,0)$ is invariant under the transformations of Eq. (10) when A belongs to SU(2).

The $\hat{\psi}_0$ of Eq. (5) does not obey Eq. (13), so it is not a particlelike solution. In order to construct such solutions from it, we first define "position" functions,

$$\psi_0(x) = e^{iP\cdot x}\hat{\psi}_0. \tag{16}$$

These functions obey the equations

$$\begin{aligned} P_0\psi_0(x) &= -i\frac{\partial\psi_0(x)}{\partial x_0}, \\ P_j\psi_0(x) &= +i\frac{\partial\psi_0(x)}{\partial x_j}, \end{aligned} \tag{17}$$

because $\partial[\exp(iP\cdot x)f(u,v)]/\partial x = iP\exp(iP\cdot x)f(u,v)$. Since P_μ and therefore $\exp(iP\cdot x)$ commute with $\mathscr{O}$, the $\psi_0(x)$ are solutions of Eq. (1).

Next, we form the set of functions

$$\psi_0(p) = \int e^{-ip\cdot x}\psi_0(x). \tag{18}$$

Then

$$\begin{aligned} P_\mu\psi_0(p) &= \int d^4x\, e^{-ip\cdot x}P_\mu\psi_0(x) \\ &= \pm i\int d^4x\, e^{-ip\cdot x}\frac{\partial\psi_0(x)}{\partial x_\mu} \\ &= \mp i\int d^4x\left(\frac{\partial e^{-ip\cdot x}}{\partial x_\mu}\right)\psi_0(x) \\ &= p_\mu\psi_0(p), \end{aligned} \tag{19}$$

provided that taking the derivatives with respect to u, v (in P_μ) within the integral sign and the integration by parts can be justified. Thus we have momentum eigenfunctions; and since they are linear combinations of solutions, they are also solutions of Eq. (1).

If we combine Eqs. (16) and (18), we see that what we have done is to define projection operators,

$$\mathscr{P}(p) = \int d^4x\, e^{i(P-p)\cdot x}, \tag{20}$$

which project out momentum eigenfunctions from an arbitrary function (provided their action on the function is well defined).

The conditions under which we can take the derivatives associated with P_μ are that $\psi_0(x,u,v)$ is an integrable function of x for each value of u, v, and that the partials with respect to u, v exist and are continuous functions of x, u, v (Leibniz's rule). To see that these properties hold, we note first that because P_μ is a first-order differential operator,

$$e^{-iP\cdot x}f(u) = f(e^{-iP\cdot x}u), \tag{21}$$

where u stands for all the variables u_{bi}, v_{bi}. As an example of the action of the translation operators,

$$e^{-iP\cdot x}u_{2i} = u_{2i} - i\bar{v}_{1i}x_0 + i\bar{u}_{1i}x_1 - \bar{u}_{1i}x_2 - i\bar{v}_{1i}x_3. \tag{22}$$

We see from this example that the exponent in $\psi_0(x)$ will continue to be quadratic in the u, v, and will contain terms linear and quadratic in the x_μ. This function is certainly integrable and derivatives with respect to u, v exist and give a continuous function.

The integration by parts is justified if $\psi_0(x)\to 0$ as $x\to\infty$ in any direction. In order to see that this holds, consider the part of the exponent quadratic in the x_μ. It has the form

$$2(z\cdot x)(c\cdot x) - (z\cdot c)(x\cdot x), \tag{23}$$

where $c_0 = 1$, $\mathbf{c} = 0$, $z_\mu = \Sigma_i z_{\mu 1 i}$. By means of a linear nonsingular change of variables from x to x', we can put this quadratic part in the diagonal form

$$-(u_{1i}\bar{u}_{1i})(x_0'^2 + x_1'^2 + x_2'^2) - (v_{1i}\bar{v}_{1i})(x_0'^2 + x_1'^2 + x_3'^2). \tag{24}$$

Now when $x_\mu \to \infty$ in any direction, at least one of the x'_μ must go to ∞. This implies $\psi_0(x) \to 0$ as $x \to \infty$ if the coefficients on the x are nonzero. Thus the functions $\psi_0(p,u,v)$ are eigenfunctions of P_μ at all points in (u, v) space except where $u_{1i}\bar{u}_{1i}$ or $v_{1i}\bar{v}_{1i}$ equals zero.

We have considered the action of translations on $\hat{\psi}_0$ and must now look at transformations from SL(2). Define an operator $U(A)$ such that

$$U(A)f(u,v) = f(u',v'), \tag{25}$$

where the matrix A and the u', v' are given in Eq. (10). We can show that

$$\begin{aligned} U(A)(P\cdot x) &= \Lambda^{-1}P\cdot x = P\cdot \Lambda x, \\ U(A)(c\cdot z) &= c\cdot\Lambda^{-1}z = \Lambda c\cdot z, \end{aligned} \tag{26}$$

so that

$$U(A)\hat{\psi}_0 = e^{-\Lambda c\cdot z} \equiv \hat{\psi}(\Lambda c), \tag{27}$$

again, with $c_0 = 1$, $\mathbf{c} = 0$. The Λ are real 4×4 matrices from the homogeneous Lorentz group G_4^1 (with a 2 to 1 mapping of the set of A's onto the set of Λ's). Note also that Λ belongs to $\mathcal{O}(3)$ when A belongs to SU(2).

We see from these formulas that

$$\begin{aligned} U(A)\psi_0(p) &= \int d^4x\, e^{-i\Lambda p\cdot y} U(A) e^{iP\cdot x}\hat{\psi}_0 \\ &= \int d^4x\, e^{-i\Lambda p\cdot y} e^{iP\cdot y}\hat{\psi}(\Lambda c), \end{aligned} \tag{28}$$

with $y = \Lambda x$ and $d^4y = d^4x$.

Now suppose that $p_0 = m$, $\mathbf{p} = 0$ and that A belongs to SU(2). Then $U(A)\psi(m,0) = \psi(m,0)$. Since spin is defined for the $\mathbf{p} = 0$ basis functions of an irreducible representation, we see that we are building a basis for a spin-0 representation.

Next, we define a standard $\Lambda_s(\mathbf{p})$ such that

$$\Lambda_s(\mathbf{p})[m,0] = [E,\mathbf{p}], \quad E = \sqrt{m^2 + \mathbf{p}\cdot\mathbf{p}}. \tag{29}$$

This Λ is defined as the product of two Lorentz transformations in the following way: First, there is a unique Λ, involving only the 0 and 3 components, such that $\Lambda[m,0] = [E,0,0,p]$. Then there is a unique rotation about an axis perpendicular to both the z axis and $\mathbf{p}$ that takes $(0,0,p)$ into (p_x,p_y,p_z). Here $\Lambda_s(\mathbf{p})$ is the product of these two transformations, with $A_s(\mathbf{p})$ being the A which gives $\Lambda_s(\mathbf{p})$. [The choice of $A_s(\mathbf{p})$ is not unique because of the 2 to 1 mapping. It can be made so by requiring that $A_s(\mathbf{p})$ be a continuous function of $\mathbf{p}$ which goes to the identity matrix as $\mathbf{p}$ goes to 0.]

Consider now the set of functions

$$U(A_s(\mathbf{p}))\psi_0(m,0) \equiv \psi(E,\mathbf{p}) \equiv \psi(p). \tag{30}$$

They are eigenfunctions of P_μ with eigenvalues E, $\mathbf{p}$, since

$$P_\mu U(A_s(\mathbf{p})) = U(A_s(\mathbf{p}))(\Lambda_s(\mathbf{p})P)_\mu. \tag{31}$$

This set of functions is also closed under SL(2) because

$$U(A)\psi(p) = \psi(\Lambda \approx p_\mu), \tag{32}$$

where Λ is the G_4^1 matrix associated with A. Thus these vectors form the basis for a mass-m, spin-0 irreducible representation of ISL(2). They are also solutions of Eq. (1), since $\exp(iP\cdot x)$ and $U(A)$ commute with $\mathcal{O}$.

Finally, we consider the scalar product of these functions, using Eq. (8), with the two masses not necessarily equal:

$$\begin{aligned} \langle p'|p\rangle &= \langle \psi(p')|\psi(p)\rangle \\ &= \langle U(A_s(\mathbf{p}))\psi_0(m,0)|U(A_s(\mathbf{p}))\psi_0(m,0)\rangle \\ &= \langle \psi_0(m',0)|U(A)\psi_0(m,0)\rangle, \end{aligned} \tag{33}$$

where $A = A_s^{-}(\mathbf{p}')A_s(\mathbf{p})$, and we used the Lorentz invariance of the scalar product. We can further reduce this result:

$$\begin{aligned} \langle p'|p\rangle &= \int du \int d^4y\, d^4y'\, \overline{e^{iM'\cdot y} e^{iP\cdot y}\hat{\psi}_0} e^{iM\cdot y'} e^{i\Lambda^{-1}P\cdot y'}\hat{\psi}(\Lambda c) \\ &= \int d^4y\, d^4y''\, e^{iM'\cdot y - \Lambda M\cdot y''} \int du\, e^{iP\cdot(y-y'')}\, \overline{\hat{\psi}_0}\hat{\psi}_0(\Lambda c), \end{aligned} \tag{34}$$

where $M' = [m',0]$, $M = [m,0]$, $y'' = \Lambda y'$. Now

$$M'\cdot y - \Lambda M\cdot y'' = (M' - \Lambda M)\cdot\frac{(y+y'')}{2} + \frac{(M' + \Lambda M)}{2}\cdot(y - y''), \tag{35}$$

so that if we change variables to $(y + y'')/2$, $w = y - y''$, then [remembering that $\delta^4(p) = \delta^4(\Lambda p)$]

$$\begin{aligned} \langle p'|p\rangle &= (2\pi)^4\delta^4(M' - \Lambda M)\hat{f}(m) \\ &= (2\pi)^4\delta^4(p' - p)\hat{f}(m), \end{aligned} \tag{36}$$

$$f(m) = \int d^4w\, e^{iM\cdot w}\langle\hat{\psi}_0|e^{-iP\cdot w}|\hat{\psi}_0\rangle. \tag{37}$$

In order to examine the behavior of $f(m)$, we note that $\overline{\hat{\psi}_0}e^{-iP\cdot w}\hat{\psi}_0$ is an exponential in u,v. Thus the function and the scalar product will factor, with an identical factor for each value of i. We therefore need to do the scalar product only for $n = 1$. Since the exponent is quadratic, we can change to a new set of u, v in which the exponent is diagonal, and then easily perform the integrals. The details are not interesting. The result is

$$\begin{aligned} &\langle\hat{\psi}_0|e^{-iP\cdot w}|\hat{\psi}\rangle = c/(4 + (w_0 - r)^2)(4 + (w_0 + r)^2), \\ &r = \sqrt{\mathbf{w}\cdot\mathbf{w}}, \end{aligned} \tag{38}$$

where c is a constant of no consequence. For $n > 1$, the $\hat{f}(m)$ is just that of Eq. (37) raised to the nth power.

For $n > 1$, $\hat{f}(m)$ is a well-behaved function of m. But for $n = 1$, the scalar product goes to zero sufficiently slowly in w_0, r that one has problems with the Fourier transform. We conclude from this that the functions in the $n = 1$ case cannot be considered a good set of basis vectors. The reason is that, in the $n = 1$ case—and only in that case—the four-momentum operators are not independent, but satisfy $z\cdot P = 0$. Since the P_μ are linearly dependent, we do not expect to be able to construct valid basis functions.

The result of Eq. (36), that $\langle\psi(p)|\psi(p')\rangle$ is proportion-

al to $\delta^4(p-p')$, is not what one expects. The scalar product of basis vectors for a given irreducible representation (fixed mass, spin) is normally taken to be proportional to $\delta^3(\mathbf{p}-\mathbf{p}')$. The reason we get the δ^4 here is that Eq. (1) has a continuum of possible mass values associated with its solutions, rather than having a discrete spectrum. The cause of this, in turn, is that there is a part of the total invariance group G_{8n}^{4n} which does not commute with the ISL(2) invariant operator $P\cdot P$.

We have seen in this section how to construct functions which serve as basis vectors for mass-m, spin-0 representations. We turn now to the construction of bases for representations with nonzero spin.

IV. NONZERO SPIN REPRESENTATIONS

We will expand the problem to be solved from that of Eq. (1) to the eigenvalue problem

$$\mathcal{O}\psi_\lambda = \lambda\psi_\lambda. \tag{39}$$

Solutions with $m=0$, spin$\neq 0$ exist, but there is not additional insight to be gained by treating this special case, so we confine ourselves to $m\neq 0$. It is easiest to obtain nonzero spin solutions by using raising and lowering operators. That is, we find operators $A_{\lambda'}$ such that

$$[\mathcal{O}, A_{\lambda'}] = \lambda' A_{\lambda'}. \tag{40}$$

Then if we have one solution, ψ_λ of Eq. (39), we can obtain another, $A_{\lambda'}\psi_\lambda$, which has eigenvalue $\lambda+\lambda'$;

$$\mathcal{O}(A_{\lambda'}\psi_\lambda) = (\lambda+\lambda')A_{\lambda'}\psi_\lambda. \tag{41}$$

The $8n$ raising operators, having $\lambda = +1$, for $\mathcal{O}$ can be chosen as

$$\begin{aligned}
&R\left(1,\frac{1}{2}\right) = u_{1i} - \frac{\partial}{\partial v_{2i}}, \quad R\left(1,-\frac{1}{2}\right) = v_{1i} + \frac{\partial}{\partial u_{2i}},\\
&R\left(2,\frac{1}{2}\right) = \bar{v}_{1i} + \frac{\partial}{\partial \bar{u}_{2i}}, \quad R\left(2,-\frac{1}{2}\right) = -\bar{u}_{1i} + \frac{\partial}{\partial \bar{v}_{2i}},\\
&R\left(3,\frac{1}{2}\right) = u_{2i} + \frac{\partial}{\partial v_{1i}}, \quad R\left(3,-\frac{1}{2}\right) = v_{2i} - \frac{\partial}{\partial u_{1i}},\\
&R\left(4,\frac{1}{2}\right) = \bar{v}_{2i} - \frac{\partial}{\partial \bar{u}_{1i}}, \quad R\left(4,-\frac{1}{2}\right) = -\bar{u}_{2i} - \frac{\partial}{\partial \bar{v}_{1i}}.
\end{aligned} \tag{42}$$

The lowering operators, having $\lambda = -1$, are obtained from

$$L(m,j_z) = R^*(m,-j_z), \tag{43}$$

where the asterisk indicates the Hermitian adjoint operator $[u^* = \bar{u},\ (\partial/\partial u)^* = -\partial/\partial\bar{u})]$.

Consider now the construction of basis vectors, $\psi_1(m,j;\ \mathbf{p},j_z)$ for a $j=\frac{1}{2}$ mass-m representation associated with $\lambda=1$. The zero momentum vectors can be chosen as

$$\psi_1(m,\tfrac{1}{2}{:}0,j_z) = R(i,j_z)\psi_0(m,0), \tag{44}$$

where $\psi_0(m,0)$ is the zero spin and momentum ψ of Eq. (18), and $i=1$ or 2 [$R(3), R(4)$ do not commute with P_μ]. The $\psi_1(m,\frac{1}{2}{:}\ 0,\ \pm\frac{1}{2})$ form a two-dimensional, spin-$\frac{1}{2}$ representation of SU(2). If we want ψ_{-1}, we use L instead of R. There are no $\lambda=0$ spin-$\frac{1}{2}$ basis vectors.

The $\mathbf{p}\neq 0$ basis vectors are obtained by using a Lorentz transformation in the same way as they were for spin 0;

$$\psi_\lambda(m,j{:}\mathbf{p},j_z) = U(A_s(\mathbf{p}))\psi_\lambda(m,j{:}0,j_z), \tag{45}$$

where here $j=\frac{1}{2}, j_z = \pm\frac{1}{2}, \lambda = \pm 1$, but the equation applies more generally for any (allowed) λ, j, j_z.

Consider next the construction of a spin-1 basis;

$$\begin{aligned}
&\psi_2(m,1{:}0,1) = R(i,\tfrac{1}{2})R(j,\tfrac{1}{2})\psi(m,0),\\
&\psi_0(m,1{:}0,1) = R(i,\tfrac{1}{2})L(j,\tfrac{1}{2})\psi(m,0),\\
&\psi_{-2}(m,1{:}0,1) = L(i,\tfrac{1}{2})L(j,\tfrac{1}{2})\psi(m,0)
\end{aligned} \tag{46}$$

$(i,j=1,2)$ with a very similar construction for $j_z=0,\ -1$ The $\mathbf{p}\neq 0$ vectors are obtained as in Eq. (45). Spin-0 basis vectors can also be constructed from quadratic products of raising and lowering operators.

If we use the product of three raising and/or lowering operators, we can, by taking suitable linear combinations, obtain both spin-$\frac{1}{2}$ and spin-$\frac{3}{2}$ basis vectors. Thus there are many different possible λ's for each spin. These examples are sufficient to indicate the possibilities for constructing $j\neq 0$ representations.

V. AN INVARIANT STATE

In addition to particlelike states, which are specified by Eqs. (13)–(15), there is one other type of state that occurs in elementary particle physics. It is the ground or vacuum state, Ψ_0, which is an ISL(2) invariant,

$$J_i\Psi_0 = K_i\Psi_0 = 0, \tag{47}$$

$$P_\mu\Psi_0 = 0. \tag{48}$$

It is possible to construct solutions to Eq. (1) with these properties. As an example, suppose we have eight sets of variables ($n=4$). We construct spin-0 $\psi_1(p_\mu)$ as in Eq. (30) from the first four sets of variables, and $\psi_2(p_\mu)$ from the second four. From these, we then construct

$$\Psi_0 = \int d^4p_1\, g(p_1\cdot p_1)\psi_1(p_1)\psi_2(-p_1), \tag{49}$$

where g is an arbitrary function of p_1^2. This function is an ISL(2) invariant; and because $\psi_1(p_1)\psi_2(-p_1)$ is a solution of Eq. (1), the linear combination of Eq. (49) is also a solution. Thus we see that we can construct solutions of Eq. (1) that have the properties of a vacuum state.

If we take the norm of this state, using Eq. (8), we find a difficulty inherent in the construction of vacuumlike states;

$$\begin{aligned}
\langle\Psi_0|\Psi_0\rangle &= \int d^4p_1'\, d^4p_1 |g|^2\\
&\quad\times\langle\psi_1(p_1')\psi_2(-p_1')|\psi_1(p_1)\psi_2(-p_1)\rangle\\
&= \int d^4p_1'\, d^4p_1 (\delta^4(p_1'-p_1))^2|g|^2|\hat{f}(m)|^2,
\end{aligned} \tag{50}$$

where $\hat{f}(m_1) = \hat{f}(\sqrt{p_1\cdot p_1})$ is from Eq. (37). This will be infinite for two reasons. One is that we have two δ functions with the same argument, resulting from the translational invariance. The other is that, even if we could ignore the δ^2, we would still get infinity, because $\int d^4p f(p) = \infty$ when the integrand is an SL(2) invariant. Thus we see that the norm of the vacuum state will always be infinite in a scheme like ours. This need not be viewed as a weakness of our theory, however, for when particle states are superimposed on the vacu-

um, and the ISL(2) invariance broken, the norm of the particle plus vacuum state need not be infinite.

It is of interest to build up particle states from this Ψ_0 instead of the $\hat{\psi}_0$ of Eq. (5), because it shows the beginning of the construction of a vector space where creation and annihilation operators can be defined. This construction is most easily done by using a "pseudocreation" operator in the bra–ket form,

$$\hat{a}^*(p{:}1) = \int d^4p' |p' + p{:}1\rangle h(\mathbf{p}'^2) \langle p'{:}1|, \tag{51}$$

where $|p{:}1\rangle$ represents the function of Eq. (30) with variable set 1 ($b = 1,2$, $i = 1,2$). Here $h(\mathbf{p}'^2)$ is an arbitrary function put in so the scalar product will converge. It is a degree of freedom which will probably not be present when we go to interacting systems. We note that, because $\mathscr{O}|p{:}1\rangle = \langle p'{:}|\mathscr{O} = 0$, we have

$$[\mathscr{O}, \hat{a}^*(p{:}1)] = 0. \tag{52}$$

Because of this and $\mathscr{O}\Psi_0 = 0$, the single-particle state

$$\begin{aligned}\Psi(p{:}1) &= \hat{a}^*(p{:}1)\Psi_0 \\ &= \int d^4p_1\, \psi_1(p_1 + p) g(p_1^2) \psi_2(-p_1) \\ &\quad \times h(\mathbf{p}_1^2) \hat{f}(m_1)\end{aligned} \tag{53}$$

is a solution of Eq. (1).

The scalar product of this single-particle state is now well-behaved provided h^2 converges sufficiently rapidly, for

$$\begin{aligned}&\langle \Psi(p'{:}1) | \Psi(p{:}1) \rangle \\ &= \int d^4p_1'\, d^4p_1 \hat{f}(p_1') \hat{f}(p_1) \bar{h}(\mathbf{p}_1'^2) h(\mathbf{p}_1^2) g^2(p_1^2) \\ &\quad \times \langle \psi_1(p_1' + p') \psi_2(p_1' | \psi_1(p_1 + p) \psi_2(p_1) \rangle \\ &= \delta^4(p' - p) \int d^4p_1 |\hat{f}(p_1)|^4 |h(\mathbf{p}_1^2)|^2 g^2(p_1^2),\end{aligned} \tag{54}$$

where we expect the δ^4 by analogy with Eq. (36). We see that $\Psi(E,\mathbf{p} = 0{:}1)$ is an SU(2) invariant, so that if $p \cdot p = m^2$, then the $\Psi(p{:}1)$ form the basis for a mass-m, spin-0 representation of ISL(2).

VI. SPACE-TIME VARIABLES

We are concerned here with how space and time enter in a theory where they are not independent variables. One way, as was mentioned in the Introduction, is as labels on solutions, like those of Eq. (16). A second way they can enter is as dependent variables, that is, as functions, $\hat{x}_\mu$, of the independent variables (we use the hat to differentiate functional x_μ from x_μ's used as labels) conjugate to the P_μ, so that they obey

$$[P_0, \hat{x}_0] = -[P_j, \hat{x}_j] = i, \quad [P_\mu, \hat{x}_\nu] = 0, \quad \mu \neq \nu. \tag{55}$$

We will give an example for the $n = 2$ (four sets of u, v) case. To have the correct SL(2) transformation properties, the $\hat{x}_\mu$ must have the form of the z_μ of Eq. (12), so we try

$$\hat{x}_0 = a_{ij}(u_{1i}\bar{u}_{2j} + v_{1i}\bar{v}_{2j}) + \text{c.c.} \tag{56}$$

The sums on i, j run from 1 to 2, we assume the a_{ij} are SL(2) invariants which do not depend on $u_{2j}, v_{2j}, \bar{u}_{2j}, \bar{v}_{2j}$, c.c. means complex conjugate, and we have made $\hat{x}_0$ real. To satisfy Eqs. (55), we must have

$$\begin{aligned}[P_3, \hat{x}_0] &= a_{ij}(u_{1i}v_{1j} + v_{1i}u_{ij}) - \text{c.c.} = 0, \\ [P_0, \hat{x}_0] &= a_{ij}(-u_{1i}v_{1j} + v_{1i}u_{1j}) - \text{c.c.} = i.\end{aligned} \tag{57}$$

These condtions do not uniquely determine the a's (and so the $\hat{x}_\mu$ are not unique). We choose

$$\begin{aligned}&a_{21} = i/[2(u_{12}v_{11} - u_{11}v_{12}) + 2(\bar{u}_{12}\bar{v}_{11} - \bar{u}_{11}\bar{v}_{12})], \\ &a_{12} = -a_{21}, \quad a_{11} = a_{22} = 0.\end{aligned} \tag{58}$$

If we define $\hat{\mathbf{x}}$ as in Eq. (56), using the form of the $\mathbf{z}$ of Eq. (12), and the a's of Eq. (58), we find that they obey all of Eqs. (55). So, at least in the sense that they have the correct transformation properties under ISL(2), these are acceptable $\hat{x}_\mu$.

One possible use for these $\hat{x}_\mu$ is in the construction of an operator $\mathscr{O}$ with interactions. Suppose our independent variables consist of sets of four-set variables, with each set labeled by η. We will have

$$P_\mu = \sum_\eta P_\mu^\eta. \tag{59}$$

We construct $\hat{x}_\mu^\eta$ conjugate to the P_μ^η, as in Eqs. (56) and (58). An ISL(2) invariant form for an interaction is then

$$\mathscr{O}_I = \sum_{\eta \neq \eta'} V(x^{\eta\eta'2}), \tag{60}$$

where $x^{\eta\eta'2} = (x_\mu^\eta - x_\mu^{\eta'})(x_\mu^\eta - x_\mu^{\eta'})$. We note that this form for the interaction is only one among many possibilities. We will simply have to try the different forms to see if they give the familiar equations of quantum mechanics.

Also note that the P_0 of Eqs. (3) and (59) is not the Hamiltonian. Rather, it is identically equal to $i(\partial/\partial t)$. The Hamiltonian is associated with a method of solving $\mathscr{O}\Psi = 0$ in which Ψ is given on one "constant t" surface and $\mathscr{O}\Psi = 0$ is used to integrate forward in time. This method is equivalent to using a Hamiltonian; the form of the Hamiltonian is determined by the form of $\mathscr{O}$, and it will contain the interaction. This method is analogous to using the variational principle in quantum field theory to obtain a Hamiltonian from the action integral.[4]

VII. SUMMARY

We have seen that the solutions of Eqs. (1) or (39) include functions which can serve as basis vectors for irreducible representations with any spin and (nonzero) mass. In constructing these representations, we have used methods that should be applicable to other linear, Lorentz invariant equations. These include the construction of $\psi(x_\mu)$ in Eq. (16); the construction of momentum eigenfunctions by the use of the projection operator of Eq. (20); the use of standard Lorentz transformations in Eqs. (29) and (45) to construct $\mathbf{p} \neq 0$ basis functions from $\mathbf{p} = 0$ functions; the construction of single-particle states from the vacuum using Eqs. (51) and (53); and the construction of functions $\hat{x}_\mu$ from Eq. (55).

The switch to the problem with interactions will be quite difficult. The most basic difficulty is that we do not know the form of the linear operator in terms of the independent vari-

ables (although we have suggested a possibility in Sec. VI); we will simply have to experiment. The next problem is that we do not expect an operator which produces interactions to separate, as does the one here. So there will almost certainly be no exact solutions. Because of this, the procedure that one will apparently be forced to follow, in order to carry out the program outlined in the introduction, is as follows.

(1) Construct an "appropriate" set of particlelike functions from each set of independent variables. These "bare particle" functions—the same for each set of variables—are to be closed under the action of the linear operator.

(2) Construct an approximate vacuum state solution to $\mathcal{O}\Psi = 0$ from sums of products of the bare particle functions, in analogy to the construction of Eq. (49).

(3) Introduce single- and multiple-particle states in a manner similar to that of Eqs. (51) and (53). From these states, construct a basis for occupation number space, and define creation and annihilation operators for bare particles. From the closure property, we should then be able to reexpress $\mathcal{O}$ in terms of these creation and annihilation operators.

(4) Finally, we look at the equations of motion of the field operators and see if they match those used in the current conventional quantum field theory. If they do, we have chosen the correct operator $\mathcal{O}$ and we have derived quantum mechanics from a single underlying equation.

[1]See. H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974), for an SU(5)-based theory which apparently unifies the weak, electromagnetic, and strong interactions. There are other candidate groups, but the idea of labels remains the same.

[2]See H. Everett, III, Rev. Mod. Phys. **29**, 454 (1957). He shows that the state of a system under observation and the states of the detectors of the results of the observation are, of *mathematical* necessity, correlated in such a way that one has the appearance of "independently existing" particles.

[3]F. A. Blood, J. Math. Phys. **22**, 67 (1981).

[4]S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1962), Sec. 7g.

On a time-dependent extension of the Morse potential

Anne E. Kondo and D. Rodney Truax
Department of Chemistry, The University of Calgary, Calgary, Alberta T2N 1N4, Canada

(Received 8 July 1987; accepted for publication 3 February 1988)

A time-dependent extension of a Morse potential is formulated. Both the bound and unbound state wave functions are obtained algebraically for the resulting time-dependent Schrödinger equations, based upon the representations of su(2) and su(1,1), respectively. The method of R-separation of variables is instrumental in the analysis.

I. INTRODUCTION

The time-independent Schrödinger equation for the bound states of the Morse potential[1] has been solved exactly by a number of authors using symmetry techniques.[2–5] The Morse potential has been applied to the analysis of anharmonic vibrations, molecular energy transfer, atom surface scattering, and inelastic collisions.[4,6,7] Recently, a group approach has also been applied to generate and solve for the unbound or scattering states of the Morse potential.[8–11] The relationship between the bound states of the Morse oscillator and the time-independent harmonic oscillator has been well established.[9,12,13] A presentation of the unbound states was generated by the analytic continuation of the compact group SU(2), describing the bound states, to the noncompact group SU(1,1).[8–11] Problems involving time-dependent potentials are useful in the study of reaction dynamics.[14–17] In this paper, an extension to a time-dependent Morse potential for the bound and unbound states is made.

There appear to be two distinct classes of evolution equations for the quantum mechanical Morse potential. The first type is

$$[\partial_{\rho\rho} + E_{\rm I} + (i/2)e^{-\rho}\partial_t - g_2(t)e^{-2\rho}]\Psi(\rho,t) = 0 . \tag{1.1}$$

We call this type of equation a class I Schrödinger equation for the Morse oscillator. Class II equations have the form

$$[\partial_{\rho\rho} + (i/2)\partial_t - \{g_2(t)e^{-2\rho} + g_0(t)e^{-\rho}\}]\Psi(\rho,t) = 0 . \tag{1.2}$$

The functions $g_2(t)$ and $g_0(t)$ in Eq. (1.1) are piecewise continuous and differentiable, but otherwise arbitrary functions of time. The authors know of no transformation of variables connecting the two classes of equations. When g_2 and g_0 are constant, both classes can be reduced to time-independent Schrödinger equations of similar form. Class I equations can be derived from a time-dependent harmonic oscillator equation; class II cannot. In class I equations, the energy spectrum is given by $E_{\rm I}$, where $E_{\rm I}$ is a separation constant obtained by reduction of the dimensionality of the harmonic oscillator equations by one or more variables. In class II equations, the energy $E_{\rm II}$ is tied to the quantum mechanical energy operator $i\partial_t$.

Although Eq. (1.1) appears to be somewhat more complicated, class I equations are those which extend the time-independent case and are solved below. Analytic solutions to class II equations are, as yet, unknown. Class I equations may prove useful in modeling problems with anharmonic time-dependent potentials.

Symmetries of a differential equation map solutions of the equation into solutions. The set of such symmetries forms a Lie algebra. There exist explicit methods for finding the symmetry operators of differential equations. For example, the symmetries of the Schrödinger equation with arbitrary time-independent potentials have been worked out.[18–20] Methods for time-dependent potentials are more complicated.[21–23] The symmetries and exact solutions for the Schrödinger equation for a time-dependent harmonic oscillator with a time-dependent field have been calculated by Truax[22,23] and we adopt that approach in this paper.

In Sec. II we calculate the symmetry operators of the two equations,

$$\begin{aligned} Q_{(i)}\Psi(\mathbf{x},t) &= \Bigg[(\epsilon_i h_i)^{-1/2}\sum_{j,k=1}^{2}\{\partial_j((\epsilon_i h_i)^{-1/2}h_i^{jk}\partial_k)\} + 2i\partial_t \\ &\quad - 2g_2(t)\sum_{j,k=1}^{2} h_i^{jk}x^j x^k\Bigg]\Psi(\mathbf{x},t) = 0, \quad i = 1,2 , \end{aligned} \tag{1.3}$$

where the two equations differ in the definition of their metrics:

$$\begin{aligned} &(h_1^{jk}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (h_2^{jk}) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \\ &\epsilon_1 = +1, \quad \epsilon_2 = -1, \quad h_i = \det(h_i^{jk}) . \end{aligned} \tag{1.4}$$

The symbols used are

$$\partial_j = \frac{\partial}{\partial x^j}, \qquad \partial_t = \frac{\partial}{\partial t}, \qquad \mathbf{x} = (x^1, x^2) .$$

When $i = 1$, the equation $Q_{(1)}\Psi = 0$ will lead to the bound states for the class I Morse oscillator; when $i = 2$, the equation $Q_{(2)}\Psi = 0$ will lead to its unbound states.

In Sec. III we obtain the bound states of the time-dependent class I Morse oscillator. In Sec. IV we treat its unbound states. In each case we employ an algebraic approach based on procedures similar to those of Alhassid *et al.*[9] The extension of this method to the time-dependent regime is not automatic, however, since we are dealing with evolution equations. We develop a novel technique to connect the algebraic structure of the Morse potential with the evolution equation (1.1). This treatment leads naturally to R-separation of variables[18] in which time is an ignorable or cyclic variable. Finally, in Sec. V, we compare briefly the algebraic approach that we have taken to those employed by Gerry[24] and Dattoli *et al.*[25]

 0022-2488/88/061396-09$02.50

II. CALCULATION OF SYMMETRY OPERATORS

A first order symmetry operator L of a differential equation like Eq. (1.3) can be expressed as a combination of the first order partial derivatives of the independent variables,[18]

$$L = A(\mathbf{x},t)\partial_t + B^1(\mathbf{x},t)\partial_1 + B^2(\mathbf{x},t)\partial_2 + C(\mathbf{x},t)\,, \quad (2.1)$$

where each coefficient depends upon the independent variables. The operator L must obey the commutation relation,

$$[L,Q_{(i)}] = \lambda_{(i)}(\mathbf{x},t)Q_{(i)}\,, \quad i=1,2\,, \quad (2.2)$$

where $\lambda_{(i)}(\mathbf{x},t)$ is an arbitrary function of x^1, x^2, and t. The condition on L ensures that if $\Psi(\mathbf{x},t)$ is a solution of (1.3), then $L\Psi(\mathbf{x},t)$ is also a solution. The set of all such operators, L, forms a Lie algebra.[18] Indeed the set of operators (1.2) generate a multiplier representation of the corresponding Lie group. The set of symmetry operators for (1.3) will be calculated for the case where $i=1$ in Sec. II A, and for $i=2$ in Sec. II B. The symmetry algebra for $Q_{(i)}$ is given the symbol $\mathfrak{S}_i$. By examining the commutation relations between the elements of the algebra, the structure of the symmetry algebra $\mathfrak{S}_i$ is determined.

A. Symmetries of $Q_{(1)}$

When $i=1$, the differential equation (1.3) is the two-dimensional Schrödinger equation,

$$Q_{(1)}\Psi(\mathbf{x},t) = [\partial_{11}+\partial_{22}+2i\,\partial_t - 2V(\mathbf{x},t)]\Psi(\mathbf{x},t) = 0\,, \quad (2.3)$$

where, for the moment, we have allowed the potential to have a more general form than in (1.3). We look for symmetry operators of this equation which obey the commutation relation (2.2).

When Eqs. (2.1) and (2.3) are substituted into Eq. (2.2), and the coefficients of like partial derivatives are collected, the following set of coupled partial differential equations is obtained:

$$A_1 = A_2 = 0\,, \quad (2.4a)$$

$$A_{11} + A_{22} + 2iA_t = 2i\lambda_{(1)}\,, \quad (2.4b)$$

$$2B^1{}_1 = 2B^2{}_2 = \lambda_{(1)}\,, \quad (2.4c)$$

$$B^2{}_1 + B^1{}_2 = 0\,, \quad (2.4d)$$

$$B^1{}_{11} + B^1{}_{22} + 2iB^1{}_t + 2C_1 = 0\,, \quad (2.4e)$$

$$B^2{}_{11} + B^2{}_{22} + 2iB^2{}_t + 2C_2 = 0\,, \quad (2.4f)$$

$$C_{11} + C_{22} + 2iC_t + 2B^1V_1 + 2B^2V_2 + 2AV_t + 2\lambda_{(1)}V = 0\,, \quad (2.4g)$$

where $A_1 = (\partial A/\partial x^1)$, etc. Solving for the coefficients A, B^1, B^2, and C, we have

$$A(\mathbf{x},t) = A(t)\,, \quad (2.5a)$$

$$B^1(\mathbf{x},t) = \tfrac{1}{2}\dot{A}(t)x^1 + bx^2 + d^1(t)\,, \quad (2.5b)$$

$$B^2(\mathbf{x},t) = \tfrac{1}{2}\dot{A}(t)x^2 - bx^1 + d^2(t)\,, \quad (2.5c)$$

$$C(\mathbf{x},t) = -i\{\tfrac{1}{4}\ddot{A}(t)[(x^1)^2+(x^2)^2] + \dot{d}^1(t)x^1 + \dot{d}^2(t)x^2\} + f(t)\,. \quad (2.5d)$$

Differentiation with respect to time of a purely time-dependent term is denoted by a dot above the term symbol.

Substituting Eqs. (2.5) into Eq. (2.4g) gives an equation that restricts the form of the potential $V(\mathbf{x},t)$,

$$\dot{A}V + AV_t + \{\tfrac{1}{2}\dot{A}x^1 + bx^2 + d^1\}V_1 + \{\tfrac{1}{2}\dot{A}x^2 - bx^1 + d^2\}V_2 = (i/2)\dddot{A} - \tfrac{1}{4}\ddot{A}[(x^1)^2+(x^2)^2] - \ddot{d}^1x^1 - \ddot{d}^2x^2 - i\dot{f}\,. \quad (2.6)$$

The most general solution to Eq. (2.6) is a potential of the following form:

$$V(\mathbf{x},t) = \overline{V}(\mathbf{x},t) + g_2{}^1(t)(x^1)^2 + g_2{}^2(t)(x^2)^2 + g_1{}^1(t)x^1 + g_1{}^2(t)x^2 + g_0(t)\,, \quad (2.7)$$

where $\overline{V}(\mathbf{x},t)$ solves the homogeneous equation obtained from Eq. (2.6). The $g_2(t)$ terms represent an anisotropic harmonic oscillator, the $g_1(t)$ terms represent a time-dependent linear potential, and $g_0(t)$ is a purely time-dependent potential. For our purposes, we can set $\overline{V}(\mathbf{x},t)=0$, and take $g_1{}^1(t) = g_1{}^2(t) = g_0(t) = 0$. Thus

$$V(\mathbf{x},t) = g_2{}^1(t)(x^1)^2 + g_2{}^2(t)(x^2)^2\,, \quad (2.8a)$$

and we obtain Eq. (1.3) for $i=1$, where

$$Q_{(1)} = \partial_{11} + \partial_{22} + 2i\,\partial_t - 2\{g_2{}^1(t)(x^1)^2 + g_2{}^2(t)(x^2)^2\}\,. \quad (2.8b)$$

Substituting Eq. (2.8a) into Eq. (2.6), and collecting the coefficients of like powers of x^1 and x^2, we obtain

$$4A\dot{g}_2{}^j + 8\dot{A}g_2{}^j + \ddot{A} = 0\,, \quad j=1,2\,, \quad (2.9a)$$

$$\ddot{d}^j + 2d^jg_2{}^j = 0\,, \quad j=1,2\,, \quad (2.9b)$$

$$\dot{f} = \tfrac{1}{2}\ddot{A}\,. \quad (2.9c)$$

Equations (2.9a) imply that either $A=0$, or $g_2{}^2 = g_2{}^1$. Since the former condition eliminates the time dependence of L, it is of little interest here. Thus the harmonic oscillator is isotropic. Equation (2.9a) is a third order differential equation, so the general solution for A can be written as a linear combination of its three linearly independent solutions Φ_j,

$$A(t) = \sum_{j=1}^{3}\beta_j\Phi_j\,, \quad (2.10)$$

where the β's are arbitrary constants. For closure of the algebra $\mathfrak{S}_1$,

$$\psi_{jk} = [\Phi_j\dot{\Phi}_k - \dot{\Phi}_j\Phi_k] \quad (2.11)$$

must also be a solution to Eq. (2.9a).[22] Furthermore, as shown in Ref. 22, if χ_1 and χ_2 are linearly independent solutions to Eqs. (2.9b), then we can choose

$$\Phi_1 = (\chi_1)^2,\quad \Phi_2 = (\chi_2)^2,\quad \Phi_3 = 2\chi_1\chi_2\,, \quad (2.12)$$

where the value of the Wronskian of the solution to Eq. (2.9b) is

$$\mathscr{W}(\chi_1,\chi_2) = \chi_1\dot{\chi}_2 - \dot{\chi}_1\chi_2 = 1\,. \quad (2.13)$$

The solutions to Eqs. (2.9b) are then

$$d^j = \beta_4^j\chi_1 + \beta_5^j\chi_2\,, \quad j=1,2\,. \quad (2.14)$$

If the expressions for A, d^1, and d^2, Eqs. (2.10) and (2.14), are substituted into Eq. (2.9c), and the resulting equation for $\dot{f}(t)$ is integrated then

$$f = \frac{1}{2}\sum_{j=1}^{3}\beta_j\dot{\Phi}_j + i\beta_6\,. \quad (2.15)$$

Substituting the relevant expressions for A, d^1, d^2, and f into Eqs. (2.5) gives the final expressions for the coefficients of L. Here, $A(t)$ is given by Eq. (2.10) and

$$B^1(\mathbf{x},t) = \sum_{j=1}^{3} \beta_j \left\{ \frac{1}{2} \ddot{\Phi} x^1 \right\} + \beta_4{}^1 \chi_1 + \beta_5{}^1 \chi_2 + b x^2 , \tag{2.16}$$

$$B^2(\mathbf{x},t) = \sum_{j=1}^{3} \beta_j \left\{ \frac{1}{2} \ddot{\Phi} x^2 \right\} + \beta_4{}^2 \chi_1 + \beta_5{}^2 \chi_2 - b x^1 , \tag{2.17}$$

$$\begin{aligned} C(\mathbf{x},t) = -i \Bigg[& \sum_{j=1}^{3} \beta_j \left\{ \frac{1}{4} \dddot{\Phi}_j [(x^1)^2 + (x^2)^2] + \frac{i}{2} \ddot{\Phi}_j \right\} \\ & + \{ \beta_4{}^1 \dot{\chi}_1 + \beta_5{}^1 \dot{\chi}_2 \} x^1 \\ & + \{ \beta_4{}^2 \dot{\chi}_1 + \beta_5{}^2 \dot{\chi}_2 \} x^2 - \beta_6 \Bigg] . \end{aligned} \tag{2.18}$$

A basis for the symmetry algebra $\mathfrak{S}_1$ can be found by setting, in turn, each of the nine arbitrary constants equal to one and all others equal to zero. The resulting basis is

$$\begin{aligned} L_j = {} & \Phi_j \, \partial_t + \tfrac{1}{2} \dot{\Phi}_j \{ x^1 \, \partial_1 + x^2 \, \partial_2 \} \\ & - (i/4) \ddot{\Phi}_j [(x^1)^2 + (x^2)^2] + \tfrac{1}{2} \dot{\Phi}_j , \quad j = 1,2,3 , \end{aligned} \tag{2.19}$$

$$L_4 = x^2 \, \partial_1 - x^1 \, \partial_2 , \tag{2.20}$$

$$\begin{aligned} & L_5 = \chi_1 \, \partial_1 - i \dot{\chi}_1 x^1 , \quad L_6 = \chi_2 \, \partial_1 - i \dot{\chi}_2 x^1 , \\ & L_7 = \chi_1 \, \partial_2 - i \dot{\chi}_1 x^2 , \quad L_8 = \chi_2 \, \partial_2 - i \dot{\chi}_2 x^2 , \\ & L_9 = E = i . \end{aligned} \tag{2.21}$$

The algebra $\mathfrak{S}_1$ is nine dimensional.

To determine the structure of algebra, the commutation relations must be examined. The operators $\{L_1, L_2, L_3\}$ satisfy the commutation relations,

$$[L_1, L_2] = L_3 , \quad [L_1, L_3] = 2L_1 , \quad [L_2, L_3] = -2L_2 , \tag{2.22}$$

forming an $\mathbf{sl}(2,\mathbb{R})$ Lie algebra.[18,20] Note that for the case where $g_2 = 0$, these generators can be interpreted as a space-time dilation, a conformal symmetry, and time translation.[22]

The generator L_4 belongs to $\mathfrak{o}(2)$, representing symmetry with respect to rotation. Since

$$[L_j, L_4] = 0 , \quad j = 1,2,3, \tag{2.23}$$

the structure of the algebra includes the direct sum of $\mathbf{sl}(2,\mathbb{R})$ with $\mathfrak{o}(2)$.

It is convenient to define the generators,

$$\begin{aligned} \mathscr{B}_1{}^1 &= -(\chi_1 \, \partial_1 - i \dot{\chi}_1 x^1) = -L_5 , \\ \mathscr{B}_2{}^1 &= +(\chi_2 \, \partial_1 - i \dot{\chi}_2 x^1) = +L_6 , \\ \mathscr{B}_1{}^2 &= -(\chi_1 \, \partial_2 - i \dot{\chi}_1 x^2) = -L_7 , \\ \mathscr{B}_2{}^2 &= +(\chi_2 \, \partial_2 - i \dot{\chi}_2 x^2) = +L_8 . \end{aligned} \tag{2.24}$$

The new operators obey the commutation relations,

$$\begin{aligned} & [\mathscr{B}_1{}^1, \mathscr{B}_2{}^1] = E , \quad [\mathscr{B}_1{}^2, \mathscr{B}_2{}^2] = E , \\ & [\mathscr{B}_1{}^1, \mathscr{B}_1{}^2] = [\mathscr{B}_2{}^1, \mathscr{B}_2{}^2] = 0 . \end{aligned} \tag{2.25}$$

The generators $\{\mathscr{B}_1{}^1, \mathscr{B}_2{}^1, \mathscr{B}_1{}^2, \mathscr{B}_2{}^2, E\}$ with commutation relations (2.25), form a two-dimensional Heisenberg–Weyl algebra $\mathfrak{w}_2$. With respect to the operators $\{L_1, L_2, L_3, L_4\}$, the commutation relations are

$$\begin{aligned} & [L_1, \mathscr{B}_1{}^1] = 0 , \quad [L_2, \mathscr{B}_1{}^1] = +\mathscr{B}_2{}^1 , \quad [L_3, \mathscr{B}_1{}^1] = -\mathscr{B}_1{}^1 , \quad [L_4, \mathscr{B}_1{}^1] = +\mathscr{B}_1{}^2 , \\ & [L_1, \mathscr{B}_1{}^2] = 0 , \quad [L_2, \mathscr{B}_1{}^2] = +\mathscr{B}_2{}^2 , \quad [L_3, \mathscr{B}_1{}^2] = -\mathscr{B}_1{}^2 , \quad [L_4, \mathscr{B}_1{}^2] = -\mathscr{B}_1{}^1 , \\ & [L_1, \mathscr{B}_2{}^1] = -\mathscr{B}_1{}^1 , \quad [L_2, \mathscr{B}_2{}^1] = 0 , \quad [L_3, \mathscr{B}_2{}^1] = +\mathscr{B}_2{}^1 , \quad [L_4, \mathscr{B}_2{}^1] = +\mathscr{B}_2{}^2 , \\ & [L_1, \mathscr{B}_2{}^2] = -\mathscr{B}_1{}^2 , \quad [L_2, \mathscr{B}_2{}^2] = 0 , \quad [L_3, \mathscr{B}_2{}^2] = +\mathscr{B}_2{}^2 , \quad [L_4, \mathscr{B}_2{}^2] = -\mathscr{B}_2{}^1 . \end{aligned} \tag{2.26}$$

Since the commutation of any element in $\mathfrak{S}_1$ with any element in $\mathfrak{w}_2$ generates an element in $\mathfrak{w}_2$, the Heisenberg–Weyl algebra $\mathfrak{w}_2$ is an invariant subalgebra of $\mathfrak{S}_1$. Thus the structure of the algebra $\mathfrak{S}_1$ can be written as

$$\{\mathbf{sl}(2,\mathbb{R}) \oplus \mathfrak{o}(2)\} \square \mathfrak{w}_2 ,$$

where $\oplus$ denotes a direct sum, while $\square$ denotes a semidirect sum. $\mathfrak{S}_1$ is called the Schrödinger algebra.

Now we have a set of symmetries for the time-dependent Schrödinger equation for the two-dimensional isotropic harmonic oscillator. Note that each of the operators in Eqs. (2.19)–(2.21) is skew-adjoint. It is more advantageous to work with a subalgebra[23] $\mathbf{S}_1$ of $\mathfrak{S}_1$, namely $\{M_3, J_+{}^1, J_-{}^1, J_+{}^2, J_-{}^2, I\}$, where the elements of $\mathbf{S}_1$ are a complexified form of the symmetry operators, Eqs. (2.19)–(2.21):

$$\begin{aligned} M_3 &= i(L_1 + L_2) \\ &= i\{\varphi \, \partial_t + \tfrac{1}{2} \dot{\varphi} \{ x^1 \, \partial_1 + x^2 \, \partial_2 \} \\ &\quad - (i/4) \ddot{\varphi} [(x^1)^2 + (x^2)^2] + \tfrac{1}{2} \dot{\varphi} \} , \\ J_+{}^1 &= (2^{-1/2}) \{ \mathscr{B}_1{}^1 + i \mathscr{B}_2{}^1 \} = -\xi^* \, \partial_1 + i \dot{\xi}^* x^1 , \\ J_-{}^1 &= (2^{-1/2}) \{ -\mathscr{B}_1{}^1 + i \mathscr{B}_2{}^1 \} = \xi \, \partial_1 - i \dot{\xi} x^1 , \\ J_+{}^2 &= (2^{-1/2}) \{ \mathscr{B}_1{}^2 + i \mathscr{B}_2{}^2 \} = -\xi^* \, \partial_2 + i \dot{\xi}^* x^2 , \\ J_-{}^2 &= (2^{-1/2}) \{ -\mathscr{B}_1{}^2 + i \mathscr{B}_2{}^2 \} = \xi \, \partial_2 - i \dot{\xi} x^2 , \\ I &= -iE = 1 , \end{aligned} \tag{2.27}$$

where

$$\begin{aligned} & \xi(t) = (2^{-1/2})(\chi_1(t) + i \chi_2(t)) , \\ & \xi^*(t) = (2^{-1/2})(\chi_1(t) - i \chi_2(t)) , \\ & \varphi = 2 \xi \xi^* , \quad \mathscr{W}(\xi(t), \xi^*(t)) = -i . \end{aligned} \tag{2.28}$$

With this choice M_3 is now Hermitian, and each pair $J_\pm{}^k$, $k=1,2$, are adjoints.

The commutation relations are

$$[M_3,J_\pm{}^k]=\pm J_\pm{}^k,$$
$$[J_+{}^k,J_-{}^k]=-I,\quad k=1,2, \tag{2.29}$$
$$[J_+{}^1,J_+{}^2]=[J_+{}^1,J_-{}^2]=[J_+{}^2,J_-{}^1]=0.$$

Also, we shall need the Casimir operator $\mathscr{C}_{(1)}$, of the subalgebra $\mathbf{S}_1$. It is defined as

$$\mathscr{C}_{(1)}=J_+{}^1J_-{}^1+J_+{}^2J_-{}^2-M_3I=-\tfrac{1}{2}(\varphi Q_{(1)}+2). \tag{2.30}$$

The Casimir operator has the property that it commutes with each of the elements of the subalgebra $\mathbf{S}_1$.

B. Symmetries of $Q_{(2)}$

When $i=2$, the differential equation (1.3) is

$$Q_{(2)}\Psi(\mathbf{x},t)=[\partial_{11}-\partial_{22}+2i\partial_t - 2g_2(t)[(x^1)^2-(x^2)^2]]\Psi(\mathbf{x},t)=0, \tag{2.31}$$

where the spatial form of the potential has been specified. From Eq. (2.1), a symmetry operator of Eq. (2.31) has the form

$$\mathbb{L}=\mathbb{A}(\mathbf{x},t)\partial_t+\mathbb{B}^1(\mathbf{x},t)\partial_1+\mathbb{B}^2(\mathbf{x},t)\partial_2+\mathbb{C}(\mathbf{x},t), \tag{2.32}$$

and obeys the commutation relation

$$[\mathbb{L},Q_{(2)}]=\lambda_{(2)}(\mathbf{x},t)Q_{(2)}. \tag{2.33}$$

Block letters have been used here to distinguish the symmetries of $Q_{(2)}$ from those of $Q_{(1)}$. Substitution of Eqs. (2.31) and (2.32) into Eq. (2.33) generates the following set of coupled partial differential equations:

$$\mathbb{A}_1=\mathbb{A}_2=0, \tag{2.34a}$$
$$\mathbb{A}_{11}-\mathbb{A}_{22}+2i\mathbb{A}_t=2i\lambda_{(2)}, \tag{2.34b}$$
$$2\mathbb{B}_1{}^1=2\mathbb{B}^2{}_2=\lambda_{(2)}, \tag{2.34c}$$
$$\mathbb{B}^2{}_1-\mathbb{B}^1{}_2=0, \tag{2.34d}$$
$$\mathbb{B}^1{}_{11}-\mathbb{B}^1{}_{22}+2i\mathbb{B}^1{}_t+2\mathbb{C}_1=0, \tag{2.34e}$$
$$\mathbb{B}^2{}_{11}-\mathbb{B}^2{}_{22}+2i\mathbb{B}^2{}_t+2\mathbb{C}_2=0, \tag{2.34f}$$
$$\mathbb{C}_{11}-\mathbb{C}_{22}+2i\mathbb{C}_t + 4\mathbb{B}^1g_2x^1-4\mathbb{B}^2g_2x^2+2\mathbb{A}\dot{g}_2[(x^1)^2-(x^2)^2] = -2g_2[(x^1)^2-(x^2)^2]\lambda_{(2)}. \tag{2.34g}$$

Following a procedure similar to the one in Sec. II A, the solutions to these coupled equations are

$$\mathbb{A}(t)=\sum_{j=1}^{3}\beta_j\Phi_j,$$
$$\mathbb{B}^1(\mathbf{x},t)=\frac{1}{2}\sum_j\beta_j\dot{\Phi}_jx^1+bx^2+\beta_4{}^1\chi_1+\beta_5{}^1\chi_2,$$
$$\mathbb{B}^2(\mathbf{x},t)=\frac{1}{2}\sum_j\beta_j\dot{\Phi}_jx^2+bx^1+\beta_4{}^2\chi_1+\beta_5{}^2\chi_2, \tag{2.35}$$
$$\mathbb{C}(\mathbf{x},t)=-i\Big[\sum_j\beta_j\Big\{\frac{1}{4}\ddot{\Phi}_j[(x^1)^2-(x^2)^2]+\Big(\frac{i}{2}\Big)\dot{\Phi}_j\Big\} + \{\beta_4{}^1\dot{\chi}_1+\beta_5{}^1\dot{\chi}_2\}x^1 - \{\beta_4{}^2\dot{\chi}_1+\beta_5{}^2\dot{\chi}_2\}x^2-\beta_6\Big].$$

The Φ_j are solutions to a differential equation for $\mathbb{A}$ identical in form to Eq. (2.9a). Furthermore, χ_1 and χ_2 solve the same homogeneous equation (2.9b), with Wronskian (2.13) and we have, as in Eqs. (2.12),

$$\Phi_1=(\chi_1)^2,\quad \Phi_2=(\chi_2)^2,\quad \Phi_3=2\chi_1\chi_2. \tag{2.36}$$

A basis for $\mathfrak{S}_2$, the symmetry algebra of $Q_{(2)}$, is

$$\mathbb{L}_j=\Phi_j\partial_t+\tfrac{1}{2}\dot{\Phi}_j[x^1\partial_1+x^2\partial_2] - (i/4)\ddot{\Phi}_j[(x^1)^2-(x^2)]+\tfrac{1}{2}\dot{\Phi}_j,\quad j=1,2,3, \tag{2.37}$$

$$\mathbb{L}_4=x^2\partial_1+x^1\partial_2, \tag{2.38}$$

$$\mathbb{L}_5=\chi_1\partial_1-i\dot{\chi}_1x^1,\quad \mathbb{L}_6=\chi_2\partial_1-i\dot{\chi}_2x^1, \tag{2.39}$$
$$\mathbb{L}_7=\chi_1\partial_2+i\dot{\chi}_1x^2,\quad \mathbb{L}_8=\chi_2\partial_2+i\dot{\chi}_2x^2,\quad \mathbb{L}_9=E=i.$$

The operators $\{\mathbb{L}_1,\mathbb{L}_2,\mathbb{L}_3\}$ satisfy the commutation relations (2.22) and so form an $\mathbf{sl}(2,\mathbb{R})$ Lie algebra; $\{\mathbb{L}_4\}$ is a basis for $\mathfrak{o}(1,1)$ and $\{\mathbb{L}_5,\mathbb{L}_6,\mathbb{L}_7,\mathbb{L}_8,E\}$ is a basis for a two-dimensional Heisenberg–Weyl algebra, $\mathfrak{w}_2$. We have $\mathfrak{S}_2=\{\mathbf{sl}(2,\mathbb{R})\oplus\mathfrak{o}(1,1)\}\square\mathfrak{w}_2$, and as Lie algebras, $\mathfrak{S}_2$ is not isomorphic to $\mathfrak{S}_1$.

As above, we confine ourselves to the subalgebra $\mathbf{S}_2$ of a complexified symmetry algebra $\{\mathbb{M}_3,H_+{}^1,H_-{}^1,H_+{}^2,H_-{}^2,I\}$,

$$\mathbb{M}_3=i(\mathbb{L}_1+\mathbb{L}_2) = i\{\varphi\partial_t+\tfrac{1}{2}\dot{\varphi}[x^1\partial_1+x^2\partial_2] - (i/4)\ddot{\varphi}[(x^1)^2-(x^2)^2]+\tfrac{1}{2}\dot{\varphi}\},$$
$$H_+{}^1=(2^{-1/2})(-\mathbb{L}_5+i\mathbb{L}_6)=-\xi^*\partial_1+i\dot{\xi}^*x^1,$$
$$H_-{}^1=(2^{-1/2})(+\mathbb{L}_5+i\mathbb{L}_6)=\xi\partial_1-i\dot{\xi}x^1, \tag{2.40}$$
$$H_+{}^2=(2^{-1/2})(-\mathbb{L}_7+i\mathbb{L}_8)=-\xi^*\partial_2-i\dot{\xi}^*x^2,$$
$$H_-{}^2=(2^{-1/2})(+\mathbb{L}_7+i\mathbb{L}_8)=\xi\partial_2+i\dot{\xi}x^2,$$
$$I=-iE=1,$$

where $\xi(t)$, $\xi^*(t)$, and φ are defined in (2.28). The commutation relations are

$$[\mathbb{M}_3,H_\pm{}^k]=\pm H_\pm{}^k,\quad k=1,2,$$
$$[H_+{}^1,H_-{}^1]=-I,\quad [H_+{}^2,H_-{}^2]=+I, \tag{2.41}$$
$$[H_+{}^1,H_+{}^2]=[H_+{}^1,H_-{}^2]=[H_+{}^2,H_-{}^1]=0.$$

The operator $\mathbb{M}_3$ is self-adjoint and the $H_{\pm}{}^{k}$ are adjoints.

The Casimir operator of S_2 is

$$C_{(2)} = H_+{}^1 H_-{}^1 - H_+{}^2 H_-{}^2 - \mathbb{M}_3 I = -\tfrac{1}{2}(\varphi Q_{(2)} + 2) . \tag{2.42}$$

III. THE BOUND STATES

A procedure[3,9] for generating the one-dimensional Schrödinger equation with a time-independent Morse potential is to select a pair of creation, $a^\dagger$, $b^\dagger$, and annihilation a, b, operators that obey the commutation relations

$$\begin{aligned} &[a^\dagger,a] = -1, \quad [b^\dagger,b] = -1 , \\ &[a^\dagger,b] = [b^\dagger,a] = [b^\dagger,a^\dagger] = 0 . \end{aligned} \tag{3.1}$$

The operators are realized on a two-dimensional harmonic oscillator space,

$$\begin{aligned} &a = (2^{-1/2})(x^1 + \partial_1), \quad a^\dagger = (2^{-1/2})(x^1 - \partial_1) , \\ &b = (2^{-1/2})(x^2 + \partial_2), \quad b^\dagger = (2^{-1/2})(x^2 - \partial_2) , \end{aligned} \tag{3.2}$$

and then recast in such a way that linear combinations of their bilinear products generate the group SU(2). In their analysis, Alhassid *et al.*[9] look for simultaneous eigenfunctions of one of the generators of SU(2), and of a fourth operator, the number operator $\mathcal{N}$. In the realization of su(2) which they use, the Casimir operator $\mathcal{C} = \frac{1}{4}\mathcal{N}(\mathcal{N}+2)$, and so $\mathcal{N}$ commutes with each of the elements of su(2). However, note that $\mathcal{N}$ is not a member of su(2). A change of variables in the eigenvalue equation for $\mathcal{N}$ produces the one-dimensional time-independent Schrödinger equation for a Morse potential.

To derive the Schrödinger equation with a time-dependent Morse potential, we use an analogous procedure to that of the time-independent case. From the calculation of the symmetries of $Q_{(1)}$ in Sec. II A, we have the operators $\{J_+{}^1, J_+{}^2, J_-{}^1, J_-{}^2\}$, which will serve as the creation and annihilation operators $\{a^\dagger, b^\dagger, a, b\}$. From Eq. (2.29), the $J_{\pm}{}^{k}$ satisfy the same commutation relations as in Eq. (3.1). In fact, when $g_2 = \frac{1}{2}$, $\xi = (2^{-1/2})e^{it}$, $\xi^* = (2^{-1/2})e^{-it}$, and from Eq. (2.27),

$$\begin{aligned} J_+{}^1 &= (2^{-1/2})e^{-it}(x^1 - \partial_1) = e^{-it}a^\dagger , \\ J_-{}^1 &= (2^{-1/2})e^{it}(x^1 + \partial_1) = e^{it}a , \\ J_+{}^2 &= (2^{-1/2})e^{-it}(x^2 - \partial_2) = e^{-it}b^\dagger , \\ J_-{}^2 &= (2^{-1/2})e^{it}(x^2 + \partial_2) = e^{it}b . \end{aligned} \tag{3.3}$$

In the limit when $t \to 0$, the operators $J_{\pm}{}^{k}$ agree with definition (3.2).

According to Alhassid *et al.*,[9] the combinations of the generators $J_{\pm}{}^{k}$,

$$\begin{aligned} F_x &= \tfrac{1}{2}(J_+{}^1 J_-{}^2 + J_+{}^2 J_-{}^1) \\ &= \tfrac{1}{2}(-\varphi\partial_{12} + i\tfrac{1}{2}\dot{\varphi}(x^2\partial_1 + x^1\partial_2) + 2\dot{\xi}\dot{\xi}^* x^1 x^2) , \\ F_y &= -(i/2)(J_+{}^1 J_-{}^2 - J_+{}^2 J_-{}^1) \\ &= -(i/2)(x^1\partial_2 - x^2\partial_1) , \\ F_z &= \tfrac{1}{2}(J_+{}^1 J_-{}^1 - J_+{}^2 J_-{}^2) \\ &= -(\varphi/2)(\partial_{11} - \partial_{22}) + (i/2)\dot{\varphi}(x^1\partial_1 - x^2\partial_2) \\ &\quad + \dot{\xi}\dot{\xi}^*[(x^1)^2 - (x^2)^2] , \end{aligned} \tag{3.4}$$

lead a realization of the su(2) Lie algebra, with commutation relations

$$[F_x,F_y] = iF_z, \quad [F_y,F_z] = iF_x, \quad [F_z,F_x] = iF_y . \tag{3.5}$$

For the time-independent harmonic oscillator with $g_2 = \frac{1}{2}$, the function $\varphi = 2\xi\xi^* = 1$ and the time dependence of these three operators vanishes.

The number operator $\mathcal{N}$ commutes with each element in su(2), and may be expressed as

$$\begin{aligned} \mathcal{N} &= J_+{}^1 J_-{}^1 + J_+{}^2 J_-{}^2 \\ &= -(\varphi/2)(\partial_{11} + \partial_{22}) + 2i\xi^*\dot{\xi} \\ &\quad + (i/2)\dot{\varphi}(x^1\partial_1 + x^2\partial_2) \\ &\quad + \dot{\xi}\dot{\xi}^*[(x^1)^2 + (x^2)^2] . \end{aligned} \tag{3.6}$$

Using the definition of φ and the Wronskian, Eq. (2.28), the following substitutions can be made in Eq. (3.6):

$$2\xi^*\dot{\xi} = (\dot{\varphi}/2) + i , \quad \dot{\xi}\dot{\xi}^* = (\ddot{\varphi}/4) + g_2\varphi . \tag{3.7}$$

Thus

$$\begin{aligned} \mathcal{N} &= -(\varphi/2)(\partial_{11} + \partial_{22}) \\ &\quad + (i/2)\dot{\varphi} - 1 + (i/2)\dot{\varphi}(x^1\partial_1 + x^2\partial_2) \\ &\quad + \{(\ddot{\varphi}/4) + g_2\varphi\}[(x^1)^2 + (x^2)^2] . \end{aligned} \tag{3.8}$$

Following the procedure of Alhassid *et al.*[9] we construct simultaneous eigenvalue equations for the operators $\mathcal{N}$ and F_y. Let their eigenstates be denoted by $|n,m\rangle$, where n is an integer, and

$$\mathcal{N}|n,m\rangle = n|n,m\rangle , \tag{3.9a}$$

$$F_y|n,m\rangle = m|n,m\rangle, \quad -(n/2) \leqslant m \leqslant (n/2) . \tag{3.9b}$$

Transforming to polar coordinates,

$$x^1 = r\cos\theta, \quad x^2 = r\sin\theta, \quad 0 \leqslant r < \infty, \quad 0 \leqslant \theta < 2\pi , \tag{3.10}$$

the two operators of interest become

$$\begin{aligned} \mathcal{N} &= -(\varphi/2)\{(1/r)\partial_r(r\partial_r) + (1/r^2)\partial_{\theta\theta}\} \\ &\quad + (i/2)\dot{\varphi} - 1 + (i/2)\dot{\varphi}r\partial_r \\ &\quad + \{(\ddot{\varphi}/4) + g_2\varphi\}r^2 , \end{aligned} \tag{3.11a}$$

$$F_y = -(i/2)\partial_\theta . \tag{3.11b}$$

Next, make the substitution

$$r^2 = (n+1)\exp(-\rho) , \quad -\infty < \rho < \infty , \tag{3.12}$$

into the first eigenvalue equation (3.9a). Then we obtain

$$\begin{aligned} &[-(\varphi/2)\{4(n+1)^{-1}e^{\rho}\partial_{\rho\rho} + (n+1)^{-1}e^{\rho}\partial_{\theta\theta}\} \\ &\quad - i\dot{\varphi}\partial_\rho + (i/2)\dot{\varphi} - 1 \\ &\quad + (n+1)\{(\ddot{\varphi}/4) + g_2\varphi\}e^{-\rho}]\Psi_{nm}(\rho,\theta,t) \\ &\quad = n\Psi_{nm}(\rho,\theta,t) . \end{aligned} \tag{3.13}$$

Taking into account Eq. (3.11b), the second eigenvalue equation (3.9b) permits $\Psi_{nm}(\rho,\theta,t)$ to be written as

$$\Psi_{nm}(\rho,\theta,t) = \langle\rho,\theta,t|n,m\rangle = e^{2im\theta}R_n{}^m(\rho,t) . \tag{3.14}$$

Using this form of the wave function in Eq. (3.13), and multiplying through by $(2\varphi)^{-1}(n+1)e^{-\rho}$, we have

$$\left\{-\partial_{\rho\rho}-\frac{(n+1)}{2\varphi}\left[i\dot{\varphi}e^{-\rho}\partial_\rho+(n+1)\left(\frac{\ddot{\varphi}}{4}+g_2\varphi\right)e^{-2\rho}+\left(\frac{i}{2}\dot{\varphi}-n-1\right)e^{-\rho}\right]\right\}R_n{}^m(\rho,t)$$
$$=-m^2R_n{}^m(\rho,t),\quad -\left(\frac{n}{2}\right)\leqslant m\leqslant\left(\frac{n}{2}\right). \tag{3.15}$$

One can easily show that Eq. (3.15) reduces to the standard time-dependent equation by setting the coefficient $g_2=\frac{1}{2}$. In this case, Eq. (3.15) becomes

$$[-\partial_{\rho\rho}+(n+1)^2\tfrac{1}{4}\{e^{-2\rho}-2e^{-\rho}\}]R_n{}^m(\rho,t)$$
$$=-m^2R_n{}^m(\rho,t)\,, \tag{3.16}$$

which has the form of the Schrödinger equation with a Morse potential as given by Alhassid *et al.*[9]

We are now faced with the task of solving Eq. (3.15). In the time-independent case, the two eigenvalue equations (3.9) allow full separation of the wave function into ρ-dependent and θ-dependent functions. Equation (3.16) is solvable in terms of confluent hypergeometric functions. However, the time-dependent equation (3.15) is a complicated mixture of ρ- and t-dependent variables. Normally, one does not view an equation like Eq. (3.15) as an evolution equation. The evolution equation can be obtained by making the variable substitutions of Eqs. (3.10) and (3.12) into Eq. (2.8b); thus

$$[\partial_{\rho\rho}-m^2+(i/2)(n+1)e^{-\rho}\partial_t$$
$$-(g_2/2)(n+1)^2e^{-2\rho}]R_n{}^m=0\,. \tag{3.17}$$

We have called Eq. (3.17) the Schrödinger equation for a class I time-dependent Morse potential.

To make the connection between Eqs. (3.15) and (3.17) we must deviate from the procedure outlined by Alhassid *et al.*[9] We need to remove the explicit time dependence in Eq. (3.15). To do this, we find a similarity variable ζ, which depends on both ρ and t. This change of variables will permit both Eqs. (3.15) and (3.17) to be transformed into the same ordinary differential equation in ζ. We can find such a similarity variable by exploiting the relationship between $\mathcal{N}$ and $Q_{(1)}$. Recall from Eq. (2.30),

$$\mathcal{N}+M_3I=-\tfrac{1}{2}\varphi Q_{(1)}-1\,, \tag{3.18}$$

where we have used Eq. (3.6). Furthermore, we have

$$[\mathcal{N},\varphi Q_{(1)}]=0\,. \tag{3.19}$$

Since $\mathcal{N}$ and $Q_{(1)}$ commute, we can require that they have common eigenfunctions $R_n{}^m(\rho,t)$. Thus $R_n{}^m(\rho,t)$ lies on the solution space of $Q_{(1)}$:

$$Q_{(1)}R_n{}^m(\rho,t)=0\,. \tag{3.20}$$

From Eqs. (3.9a), (3.18), and (3.20) we can write

$$(\mathcal{N}+\tfrac{1}{2}\varphi Q_{(1)})R_n{}^m(\rho,t)=-(M_3+1)R_n{}^m(\rho,t)$$
$$=nR_n{}^m(\rho,t)\,. \tag{3.21}$$

We have replaced the second order differential operator $\mathcal{N}$ with a first order symmetry operator M_3 which contains the partial derivative with respect to time. Explicitly, Eq. (3.21) is

$$\{-i\dot{\varphi}\,\partial_\rho+(i/2)\dot{\varphi}-1$$
$$+\tfrac{1}{4}(n+1)\ddot{\varphi}e^{-\rho}+i\varphi\,\partial_t\}R_n{}^m(\rho,t)$$
$$=nR_n{}^m(\rho,t)\,. \tag{3.22}$$

Now, we have a first order partial differential equation that may be solved by the method of characteristics (see Ref. 23). The similarity variable is

$$\zeta=(1/\varphi)e^{-\rho}\,, \tag{3.23}$$

and the solution has the form,

$$R_n{}^m(\rho,t)=a_n(\zeta)e^{\Re}\,, \tag{3.24}$$

where a_n is a function of the similarity variable ζ. Here, $\Re$ is defined by

$$\Re=i\tfrac{1}{2}\ln\varphi+\tfrac{1}{4}(n+1)\dot{\varphi}\zeta-(n+1)\Upsilon\,, \tag{3.25}$$

where

$$\Upsilon=\int^t\frac{1}{\varphi}=\frac{i}{2}\ln\left(\frac{\xi^*}{\xi}\right). \tag{3.26}$$

Equation (3.21) implies that the time variable is an ignorable variable. Integration of Eq. (3.22) leads not only to the similarity variable ζ but also to the R-factor $\Re$ in Eqs. (3.24) and (3.25). Thus the differential equation (3.17) is R-separable.[18]

Expressing $\Re$ in terms of the similarity variable ζ, we have

$$\Re=-2(n+1)^{-1}\zeta^{-1}\{\zeta\,\partial_\zeta+\zeta^2\,\partial_{\zeta\zeta}-m^2\}+i\dot{\varphi}\zeta\,\partial_\zeta$$
$$+(n+1)\{(\ddot{\varphi}/4)+g_2\varphi\}\zeta\varphi+(i/2)\dot{\varphi}-1\,. \tag{3.27}$$

Substituting Eqs. (3.24) and (3.27) into the eigenvalue equation (3.9a), and multiplying from the left by $e^{-\Re}$ gives

$$e^{-\Re}\mathcal{N}e^{\Re}a_n(\zeta)$$
$$=[-2(n+1)^{-1}\zeta^{-1}\{\zeta^2\,\partial_{\zeta\zeta}+\zeta\,\partial_\zeta-m^2\}$$
$$+\tfrac{1}{8}(n+1)\zeta\{2\ddot{\varphi}\varphi+8g_2\varphi^2-\dot{\varphi}^2\}-1]$$
$$\times a_n(\zeta)=na_n(\zeta)\,. \tag{3.28}$$

Then by the definition of φ and the Wronskian, Eq. (2.28), the explicit time dependence vanishes since

$$2\ddot{\varphi}\varphi+8g_2\varphi^2-\dot{\varphi}^2=4\,. \tag{3.29}$$

Rearranging Eq. (3.28) and dividing through by $-2(n+1)^{-1}\zeta^{-1}$ leaves

$$[\zeta^2d_{\zeta\zeta}+\zeta d_\zeta-m^2-\tfrac{1}{4}(n+1)^2\{\zeta^2-2\zeta\}]a_n(\zeta)=0\,. \tag{3.30}$$

We have a second-order ordinary differential equation in one variable ζ. Equation (3.30) can also be obtained from the relationship

$$e^{-\Re}Q_{(1)}e^{\Re}a_n(\zeta)=0\,. \tag{3.31}$$

If we let

$$a_n(\eta)=\{\eta/(n+1)\}^m\exp\{-\tfrac{1}{2}\eta\}W(\eta),$$
$$\eta=(n+1)\zeta\,, \tag{3.32}$$

then the equation reduces to

$$\{\eta\,\partial_{\eta\eta}+(\alpha_1-\eta)\partial_\eta-\delta_1\}W(\eta)=0\,,$$
$$\alpha_1=2m+1,\quad \delta_1=m-(n/2)\,. \tag{3.33}$$

Equation (3.33) is Laplace's equation, and the solutions

$W(\eta)$ are confluent hypergeometric functions which depend implicitly on time through the variable $\eta=(n+1)\zeta=(n+1)e^{-\rho}/\varphi(t)$. Combining Eq. (3.32) and the solution to Eq. (3.33), we get

$$a_n(\zeta)=\zeta^m\exp\{-(n+1)(\zeta/2)\}\times W(m-(n/2),2m+1,(n+1)\zeta). \quad (3.34)$$

The wave function $R_n{}^m(\zeta,t)$ is then

$$\begin{aligned}R_n{}^m(\zeta,t)&=\Omega\exp i\{(i/2)\ln\varphi+\tfrac{1}{4}(n+1)\dot\varphi\zeta-(n+1)\Upsilon\}\\&\quad\times\zeta^m\exp\{-(n+1)(\zeta/2)\}\\&\quad\times W(m-(n/2),2m+1,(n+1)\zeta).\end{aligned} \quad (3.35)$$

The constant Ω can be determined by normalization of the wave function.

Observe that Eqs. (3.30) and (3.33) are identical in form to those obtained by Alhassid *et al.*[9] for the time-independent case. However, in contrast to these results, the wave function (3.35) contains not only the time-dependent factor $e^{i\Re}$, but the confluent hypergeometric functions W are implicitly time dependent as well. In the event that $g_2=\frac{1}{2}$, we recover the usual time-independent Schrödinger equation for the Morse potential and its solutions.

In the time-independent case, the depth of the well determines the number of energy levels, n. See Eq. (2.25) of Ref. 9. In the time-dependent case, the depth of the well is

$$D=-[(i/2)\varphi-n-1]^2/[8\varphi\{(\ddot\varphi/4)+g_2\varphi\}]. \quad (3.36)$$

Therefore, the depth of the well varies with time, as does the number of energy levels.

IV. THE UNBOUND STATES

To generate the Schrödinger equation for the unbound states of the time-independent Morse potential, Alhassid *et al.*[9] recast the bilinear products of the creation and annihilation operators, $\{a^\dagger,b^\dagger,a,b\}$ to form the generators of the su(1,1) Lie algebra. The unbound states form a representation space of su(1,1) algebra, viewed as the analytic continuation of su(2).

In the time-dependent case, we shall work with the elements $\{H_+{}^1,H_-{}^1,H_+{}^2,H_-{}^2\}$ from Eq. (2.40). If we set $g_2=\frac{1}{2}$, $\xi=(2^{-1/2})e^{it}$, and $\xi^*=(2^{-1/2})e^{-it}$, we can compare the operators $H_\pm{}^k$ to the time-independent creation and annihilation operators $\{a^\dagger,b^\dagger,a,b\}$ as follows:

$$\begin{aligned}H_+{}^1&=(2^{-1/2})e^{-it}(x^1-\partial_1)=e^{-it}a^\dagger,\\H_-{}^1&=(2^{-1/2})e^{it}(x^1+\partial_1)=e^{it}a,\\H_+{}^2&=-(2^{-1/2})e^{-it}(x^2-\partial_2)=-e^{-it}b,\\H_-{}^2&=-(2^{-1/2})e^{it}(x^2+\partial_2)=-e^{it}b^\dagger.\end{aligned} \quad (4.1)$$

Following the procedure of Ref. 9, we construct the products,

$$\begin{aligned}K_x&=\tfrac{1}{2}(H_+{}^1H_-{}^2+H_+{}^2H_-{}^1),\\K_y&=-(i/2)(H_+{}^1H_-{}^2-H_+{}^2H_-{}^1),\\K_z&=\tfrac{1}{2}(H_+{}^1H_-{}^1+H_+{}^2H_-{}^2).\end{aligned} \quad (4.2)$$

The operators $\{K_x,K_y,K_z\}$ with their commutation relations

$$[K_x,K_y]=-iK_z,\quad [K_z,K_x]=iK_y,\quad [K_z,K_y]=-iK_x \quad (4.3)$$

form the algebra su(1,1). In this realization, the Casimir operator $\mathscr{C}=\frac{1}{4}(\mathscr{D}+1)(\mathscr{D}-1)$, where $\mathscr{D}$ is called the difference operator and $\mathscr{D}$ commutes with each of the elements in su(1,1). Here, $\mathscr{D}$ is not a member of su(1,1) and has the form

$$\mathscr{D}=H_+{}^1H_-{}^1-H_+{}^2H_-{}^2+1. \quad (4.4)$$

Now we construct the simultaneous eigenvalue equations

$$\mathscr{D}|\mu,k\rangle=\mu|\mu,k\rangle, \quad (4.5a)$$

$$K_y|\mu,k\rangle=k|\mu,k\rangle. \quad (4.5b)$$

Since su(1,1) is noncompact, the spectra of the operators $\mathscr{D}$ and K_y are continuous. In Cartesian coordinates, the operators of interest are

$$\begin{aligned}\mathscr{D}&=-(\varphi/2)(\partial_{11}-\partial_{22})+(i/2)\dot\varphi\\&\quad+(i/2)\dot\varphi(x^1\partial_1+x^2\partial_2)\\&\quad+\{\tfrac{1}{4}\ddot\varphi+g_2\varphi\}[(x^1)^2-(x^2)^2],\\K_y&=(-i/2)(x^2\partial_1+x^1\partial_2).\end{aligned} \quad (4.6)$$

Instead of transforming to polar coordinates, we transform to hyperbolic coordinates

$$x^1=r\cosh\theta,\quad x^2=r\sinh\theta,\quad 0\leqslant r<\infty,\quad 0\leqslant\theta<\infty. \quad (4.7)$$

In these coordinates,

$$\begin{aligned}\mathscr{D}&=-(\varphi/2)(\partial_{rr}+(1/r)\partial_r-(1/r^2)\partial_{\theta\theta})\\&\quad+(i/2)\dot\varphi+(i/2)\dot\varphi r\partial_r+\{\tfrac{1}{4}\ddot\varphi+g_2\varphi\}r^2,\end{aligned} \quad (4.8a)$$

$$K_y=(-i/2)\partial_\theta. \quad (4.8b)$$

Substituting Eq. (4.8a) into the eigenvalue equation (4.5a), and making the transformation

$$r^2=\mu e^{-\rho},\quad -\infty<\rho<\infty, \quad (4.9)$$

we have

$$\begin{aligned}&[-(\varphi/2)\{(4/\mu)e^\rho\partial_{\rho\rho}+(1/\mu)e^\rho\partial_{\theta\theta}\}-i\dot\varphi\partial_\rho\\&\quad+(i/2)\dot\varphi-1+\mu\{\tfrac{1}{4}\ddot\varphi+g_2\varphi\}e^{-\rho}]\Psi_{\mu k}(\rho,\theta,t)\\&\quad=\mu\Psi_{\mu k}(\rho,\theta,t).\end{aligned} \quad (4.10)$$

From Eq. (4.8b), the second eigenvalue equation (4.5b) permits the wave function $\Psi_{\mu k}(\rho,\theta,t)$ to be written as

$$\Psi_{\mu k}(\rho,\theta,t)=\langle\rho,\theta,t|\mu,k\rangle=e^{2ik\theta}R_\mu{}^k(\rho,t). \quad (4.11)$$

Using this form of the wave function in Eq. (4.10) and rearranging, gives

$$\begin{aligned}&\left\{-\partial_{\rho\rho}-\frac{\mu}{2\varphi}\left[i\dot\varphi e^{-\rho}\partial_\rho+\mu\left(\frac{\ddot\varphi}{4}+g_2\varphi\right)e^{-2\rho}\right.\right.\\&\quad\left.\left.+\left(\frac{i}{2}\dot\varphi-\mu\right)e^{-\rho}\right]\right\}R_\mu{}^k(\rho,t)\\&\quad=+k^2R_\mu{}^k(\rho,t).\end{aligned} \quad (4.12)$$

If $g_2=\frac{1}{2}$, then $\varphi=1$, and Eq. (4.12) reduces to the time-

independent Schrödinger equation describing the unbound states of a Morse potential,

$$[-\partial_{\rho\rho}+(\mu^2/4)(e^{-2\rho}-2e^{-\rho})]R_\mu{}^k(\rho) = +k^2R_\mu{}^k(\rho)\,. \tag{4.13}$$

Analogous to the procedure of Sec. III, we can solve (4.12) by taking advantage of the relationship between the difference operator $\mathscr{D}$ and $Q_{(2)}$. From the Casimir operator of the subalgebra S_2, Eq. (2.42),

$$\{\mathscr{D}+(\varphi/2)Q_{(2)}\}R_\mu{}^k = M_3R_\mu{}^k\,. \tag{4.14}$$

In ρ and t variables, $Q_{(2)}$ is

$$Q_{(2)} = \partial_{\rho\rho}+k^2+(i/2)\mu e^{-\rho}\partial_t-(g_2/2)\mu^2e^{-2\rho}\,. \tag{4.15}$$

Furthermore, since

$$[\mathscr{D},\varphi Q_{(2)}] = 0\,, \tag{4.16}$$

we can require that $R_\mu{}^k(\rho,t)$ lies in the solution space of $Q_{(2)}$. Thus

$$Q_{(2)}R_\mu{}^k = 0\,, \tag{4.17}$$

and, from Eq. (4.14) and the eigenvalue equation (4.5a),

$$M_3R_\mu{}^k = \mu R_\mu{}^k\,. \tag{4.18}$$

Explicitly

$$[-i\dot\varphi\,\partial_\rho+i\varphi\,\partial_t+(i\dot\varphi/2)+\tfrac14\ddot\varphi\mu e^{-\rho}]R_\mu{}^k = \mu R_\mu{}^k\,. \tag{4.19}$$

Solving Eq. (4.19) by the method of characteristics,[23] we obtain the similarity variable

$$(1/\varphi)e^{-\rho} = \zeta\,, \tag{4.20}$$

and the $\mathscr{R}$-factor,

$$\mathscr{R}(\zeta,t) = (i/2)\ln\varphi+(\mu/4)\dot\varphi\zeta-\mu\Upsilon\,, \tag{4.21}$$

where Υ is defined in Eq. (3.26). Hence

$$R_\mu{}^k(\zeta,t) = e^{i\mathscr{R}}a_\mu(\zeta)\,, \tag{4.22}$$

where $a_\mu(\zeta)$ is a function of ζ. To determine the specific form of $a_\mu(\zeta)$ we perform the transformation

$$e^{-i\mathscr{R}}\mathscr{D}e^{i\mathscr{R}}a_\mu(\zeta) = [-2(\mu\zeta)^{-1}\{\zeta^2\partial_{\zeta\zeta}+\zeta\partial_\zeta+k^2\} + (\mu/2)\zeta]a_\mu(\zeta) = \mu a_\mu(\zeta)\,. \tag{4.23}$$

Rearranging Eq. (4.23), we have

$$[\zeta^2\partial_{\zeta\zeta}+\zeta\partial_\zeta+k^2-(\mu^2/4)\{\zeta^2-2\zeta\}]a_\mu(\zeta) = 0\,. \tag{4.24}$$

We have a second order differential equation in one variable ζ with solutions

$$a_\mu(\zeta) = \zeta^{\pm ik}\exp\{-(\eta/2)\}W_\pm(\eta)\,,\quad \eta=\mu\zeta\,, \tag{4.25}$$

where $W(\eta)$ solves

$$\{\eta\partial_{\eta\eta}+(\alpha_2-\eta)\partial_\eta-\delta_2\}W_\pm(\eta) = 0\,,\qquad \alpha_2 = \pm 2ik+1,\quad \delta_2 = \tfrac12\{\pm 2ik+1-\mu\}\,. \tag{4.26}$$

Thus

$$a_\mu(\zeta) = U\zeta^{ik}\exp\{-(\mu/2)\zeta\}W_+(\tfrac12\{2ik+1-\mu\},2ik+1,\mu\zeta) + V\zeta^{-ik}\exp\{-(\mu/2)\zeta\}W_-(\tfrac12\{-2ik+1-\mu\},-2ik+1,\mu\zeta)\,. \tag{4.27}$$

The ratio of constants V/U can be determined by imposing the boundary condition that $a_\mu(\zeta\to\infty)=0$, since as $\zeta\to\infty$, $\rho\to-\infty$, and the Morse potential goes to infinity. Thus

$$\frac{V}{U} = \frac{\Gamma(2ik+1)\Gamma(\tfrac12\{-2ik+1-\mu\})}{\Gamma(-2ik+1)\Gamma(\tfrac12\{2ik+1-\mu\})}\,. \tag{4.28}$$

The reflection coefficient is $(|V|^2/|U|^2)=1$.

As we found in Sec. III for Eq. (3.17) for the bound states, Eq. (4.15) for the unbound states is also R-separable. The time variable is ignorable because of Eq. (4.18).

V. DISCUSSION

It is of interest to compare the approach we have adopted to algebraic methods employed by others[24,25] in solving time-dependent problems.

The evolution of a quantum system[26] may be given by $|\psi(t)\rangle = U(t,t_0)|\psi(t_0)\rangle$, where $U(t,t_0)$ is the evolution operator and is usually expressed in terms of the system Hamiltonian. If the Hamiltonian can be written as a linear combination of SU(1,1) or SU(2) generators, in which the coefficients are time dependent, then the technique of Dattoli *et al.*[25] or the propagator approach of Gerry[24] are appropriate. In the method we have used, the Schrödinger equation for the Morse potential of both the bound states is related to the Casimir operators for su(2) and su(1,1), respectively, and is not realized as a linear combination of the group generators. The finite number of bound states for the Morse oscillator are described by the irreducible unitary representations of su(2). The spectrum of unbound states is described by the continuous series[10] $C_k{}^{1/2}$, one of the continuous irreducible unitary representations of su(1,1). The discrete series for su(1,1) in which the spectrum is bounded below, does not play a role in our approach. However, the discrete series is the irreducible representation exploited by Gerry[24] and Dattoli *et al.*[25] The common ground of the two approaches is an interesting area for possible future research.

ACKNOWLEDGMENT

We gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

[1] P. M. Morse, Phys. Rev. **34**, 57 (1929).
[2] A. O. Barut, A. Inomata, and R. Wilson, J. Math. Phys. **28**, 605 (1987).
[3] R. D. Levine, Chem. Phys. Lett. **95**, 87 (1983).
[4] M. Berrondo and A. Palma, J. Phys. A **13**, 773 (1980).
[5] R. D. Levine and C. E. Wulfman, Chem. Phys. Lett. **60**, 372 (1979).
[6] F. Iachello, Chem. Phys. Lett. **78**, 581 (1981).
[7] O. S. Van Roosmalen, A. E. L. Dieperink, and F. Iachello, Chem. Phys. Lett. **85**, 32 (1982).
[8] Y. Alhassid, Chem. Phys. Lett. **108**, 79 (1984).
[9] Y. Alhassid, F. Gursey, and F. Iachello, Ann. Phys. (NY) **148**, 346 (1983); Chem. Phys. Lett. **99**, 27 (1983); Phys. Rev. Lett. **50**, 873 (1983).
[10] Y. Alhassid, F. Gursey, and F. Iachello, Ann. Phys. (NY) **167**, 181 (1986); A. Frank and K. B. Wolf, J. Math. Phys. **26**, 973 (1985).
[11] Y. Alhassid and J. Wu, Chem. Phys. Lett. **109**, 81 (1984).

[12]S. Lee, Am. J. Phys. **53**, 753 (1985).
[13]R. Montemayor and L. Urrutia, Am. J. Phys. **51**, 641 (1983).
[14]G. L. Hofacker and R. D. Levine, Chem. Phys. Lett. **33**, 404 (1975).
[15]M. D. Pattengill and J. C. Polanyi, Chem. Phys. **3**, 1 (1974).
[16]P. J. Kuntz, M. Nemeth, J. C. Polanyi, S. D. Rosner, and C. E. Young, J. Chem. Phys. **44**, 1168 (1966).
[17]K. L. Sebastian, J. Chem. Phys. **76**, 1075 (1982).
[18]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison–Wesley, Reading, MA, 1977); E. G. Kalnins and W. Miller, Jr., J. Math. Phys. **28**, 1005 (1987).
[19]J. Cizek and J. Paldus, Int. J. Quantum Chem. **12**, 875 (1977).
[20]C. P. Boyer, Helv. Phys. Acta **47**, 589 (1974).
[21]H. R. Lewis, Jr., J. Math. Phys. **9**, 1976 (1968); Phys. Rev. Lett. **18**, 510, 636 (1967).
[22]D. R. Truax, J. Math. Phys. **22**, 1959 (1981).
[23]D. R. Truax, J. Math. Phys. **23**, 43 (1982).
[24]C. G. Gerry, Phys. Rev. A **31**, 2721 (1985); **33**, 2207 (1986).
[25]G. Dattoli, A. Torre, and R. Caloi, Phys. Rev. A **33**, 2789 (1986).
[26]K. Gottfried, *Quantum Mechanics I.* (Benjamin, Reading, MA, 1974).

Nonlinearly coupled oscillators in quantum mechanics: A normal form approach

Martin Kummer and Raghuramaiah Gompa
University of Toledo, Toledo, Ohio 43606

(Received 25 September 1987; accepted for publication 3 February 1988)

Quantum mechanical Hamiltonians of the form of n harmonic oscillators coupled via an interaction of the form ϵ times a polynomial in the position and momenta variables are studied in a rigorous Hilbert space setting. In particular, normal form theory is used to define the mth approximation to the associated Schrödinger initial value problem and it is shown that it deviates in norm from the exact solution by a term of order $\epsilon^{m+1}|t|$ (t = time) provided only that the initial vector is confined to an appropriate dense subspace. The main concentration is on the case in which there exist no resonances of order $\leqslant m$ between the frequencies of the n oscillators, but the case of two oscillators in 1:1 resonance is also taken up.

I. INTRODUCTION

In recent years normal form theory has become a very versatile tool in the study of the flow that a classical Hamiltonian induces in those regions of phase space in which it is predominantly deterministic (e.g., periodic or quasiperiodic) rather than "chaotic." Pioneered by Birkhoff[1] and Siegel (e.g., see Ref. 2) normal form theory has been the subject or tool of investigation in so many works in recent years that it is impossible to quote all of them. Besides mentioning some we feel are representative[3–6] we refer to the recent works by van der Meer[7] and Sanders and Verhulst[8] and the bibliographies contained therein.

It seems that the study of the corresponding quantum mechanical theory, although known on a purely formal level [see Ref. 10(a)], has not been given the attention it deserves in the mathematical literature. The present paper attempts to fill this gap. Our work was stimulated by the pioneering work of Lemlih and Ellison[9] who, by presenting a rigorous analysis of the time averaging method in a quantum mechanical model problem, set the stage for an elevation of quantum mechanical normal form theory from a purely formal theory[10] to a rigorous mathematical theory (see, however, Ref. 11).

Rather than developing the theory in the framework of an abstract Hilbert space formalism, as has been done by one of the present authors,[12] in the present paper we confine our discussion to the class of quantum mechanical Hamiltonians that describe n harmonic oscillators with polynomial interaction. Working within the framework of the Heisenberg algebra $\mathscr{H}_n$ of creation and annihilation operators a_k, c_k ($k = 1,2,...,n$) a Hamiltonian of our class has the following explicit form:

$$H = H^0 + \epsilon V, \tag{1.1}$$

where

$$H^0 = \sum_{k=1}^{n} \omega_k \left(c_k a_k + \frac{1}{2}\right) \tag{1.2}$$

and V is a polynomial in the a_k's and c_k's.

In Sec. II we bring the interaction into a "standard" form, which is characterized by the property that each c_k is paired off with the corresponding a_k to yield the number operator $N_k := c_k a_k$ whenever this is possible and the operators that stay "single" in the process are moved to the left.

In Sec. III we discuss normal form theory, first abstractly and then in the context of our class of Hamiltonians, under the assumption that the perturbing interaction potential V is already in standard form. As in the corresponding classical theory (see, e.g., Refs. 6 and 13) we find that the theory of graded Lie algebras provides an adequate setting for our discussion.

Using the coefficients in a formal ϵ expansion of the normal form as well as those of the normalizing transformation we define the mth-order approximation $\phi_m(t)$ to the exact solution $\psi(t)$ of the Schrödinger initial value problem

$$i\dot{\psi} = (H_0 + \epsilon V)\psi, \quad \psi(0) = \psi_0, \tag{1.3}$$

and in Sec. IV we formulate conditions under which an estimate of the form

$$\|\phi_m(t) - \psi(t)\| = O(\epsilon^{m+1}|t|) \tag{1.4}$$

holds for ϵ varying in some small interval $[0,\epsilon_0)$.

It turns out that an appropriate framework for the rigorous study of our family of Hamiltonians is provided by a certain Hilbert space of analytic functions described by Bargmann[14–16] (see, also, Ref. 17). The simplest case for which this approach allows us to obtain estimates of the form (1.4) is the case in which the frequencies ω_k ($k = 1,2,...,n$) of the harmonic oscillators are subjected to some nonresonance condition. Of course, as in the time averaging technique, studied by means of a model problem described by Lemlih and Ellison,[9] the validity of an estimate of type (1.4) requires that the initial vector ψ_0 be confined to a suitable dense linear submanifold of our Hilbert space. This submanifold will be referred to as "initial domain."

Quite different from the nonresonant case are the low-order resonances, in particular the resonance in which the frequencies of the n harmonic oscillators all agree. In the case of $n = 2$ this is called the 1:1 resonance and the classical counterpart has been studied by one of the present authors in Ref. 5 (see, also, Ref. 18). We briefly sketch this case at the end of Sec. IV by exhibiting a suitable initial domain. The

 0022-2488/88/061405-12$02.50

one remarkable feature of this case, however, is the fact that the Hilbert space under the time evolution operator of its normal form breaks up into invariant subspaces exactly the same way as under the natural representation of the group SU(2) (see Ref. 16), namely, into a direct sum of $(2j+1)$-dimensional invariant subspaces ($j = 0, \frac{1}{2}, 1, \ldots$).

Finally, in Sec. V we touch upon an alternative method of approximation also based on the normal form technique: The difference consists in the form in which we write the normalizing transformation. Whereas in the first normal form method we write the normalizing transformation as an infinite product of exponentials, in the alternative method we write it directly as a formally unitary transformation. There is a trade-off between these two methods. Whereas in the alternative method the explicit expressions for the approximants are much simpler than in the first normal form method, the initial vector ψ_0 in the alternative method has to be restricted more severely.

The reader might question how our rigorous approach to time-dependent quantum mechanical perturbation theory fits in with those approaches existing in the literature, notably with the extensive work of Kato[19]: It appears to us that Chap. IX, Sec. 6 (p. 506) of the second edition of his work bears some relationship to our work. However, since Kato's[19] asymptotic formulas are based on straightforward perturbation theory rather than on normal form theory, they are already different in the first approximation. It would be an interesting task to make a detailed comparison between Kato's[19] and our approaches. However, such a comparison would lead us in a different direction and could be the subject of a separate study.

II. PREPARATION OF THE HAMILTONIAN

Our goal is to find approximate solutions of the Schrödinger equation associated with the quantum mechanical Hamiltonian

$$H = \sum_{k=1}^{n} \frac{p_k^2}{2M_k} + \frac{1}{2} \sum_{k=1}^{n} M_k \omega_k^2 q_k^2 + \epsilon P(q,p), \tag{2.1}$$

which describes n nonlinearly coupled oscillators. Here $P(q,p)$ is a polynomial in the position and momentum variables (q_k, p_k) $(k = 1,2,\ldots,n)$ and the coupling parameter ϵ is sufficiently small so that we can view the interaction term as a perturbation of the Hamiltonian H^0 which results from (2.1) by setting $\epsilon = 0$ and which describes n uncoupled harmonic oscillators with frequencies $\omega_1, \omega_2, \ldots, \omega_n$.

The position and momentum variables $(q_k)_{k=1}^n$, $(p_k)_{k=1}^n$ satisfy the Heisenberg commutation relations. This means that they all commute, with the exception of q_k, p_k $(k = 1,2,\ldots,n)$ which are subjected to the commutation relations

$$[p_k, q_k] = -i. \tag{2.2}$$

The well-known transformation of variables

$$q_k = (2M_k\omega_k)^{-1/2}(a_k + c_k),$$
$$p_k = (M_k\omega_k/2)^{1/2}\,(1/i)(a_k - c_k) \tag{2.3}$$

transforms our Hamiltonian into the form

$$H = H^0 + \epsilon V(a,c), \quad H^0 = \sum_{k=1}^{n} \omega_k \left(N_k + \frac{1}{2}\right), \tag{2.4}$$

where $N_k := c_k a_k$ $(k = 1,2,\ldots,n)$ and V is a polynomial in the noncommutative variables c_k, a_k. The latter are commonly known as creation and annihilation operators and the particular combination $N_k = c_k a_k$ is known as kth number operator. As a consequence of the Heisenberg commutation relations the a_k's commute among themselves, and so do the c_k's, whereas the commutation relations

$$[a_k, c_j] = \delta_{kj} \quad (k,j = 1,2,\ldots,n) \tag{2.5}$$

hold.

In the following the free algebra over $\mathbb{C}$, generated by the $2n+1$ elements $a_1, a_2, \ldots, a_n$, $c_1, c_2, \ldots, c_n$, 1 modulo the commutation relations (2.5) will be denoted by $\mathscr{H}_n$ and referred to as Heisenberg algebra (for a quantum mechanical system of n degrees of freedom). Clearly, the complex conjugation extends to the involutive antiautomorphism denoted by the dagger which interchanges a_k and c_k, i.e.,

$$a_k^\dagger = c_k, \quad c_k^\dagger = a_k \quad (k = 1,2,\ldots,n). \tag{2.6}$$

We find it convenient to introduce the notation

$$b_k^l = \begin{cases} c_k^l, & \text{for } l > 0, \\ 1, & \text{for } l = 0, \\ a_k^{-l}, & \text{for } l < 0. \end{cases} \tag{2.7}$$

Here on the rhs c_k^l is the lth power of the variable c_k, whereas a_k^{-l} is the $|l|$th power of the variable a_k. For $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{Z}^n$, we set

$$b^\alpha = b_1^{\alpha_1} \cdot b_2^{\alpha_2} \cdot \cdots \cdot b_n^{\alpha_n}. \tag{2.8}$$

Note that since in b^α for $k = 1,2,\ldots,n$ either a_k or c_k, but not both, occur, there are no relations between the b^α's so that the Heisenberg algebra contains the free right module over the ring $\mathbb{C}[N]$ of polynomials in the number operators $N = (N_1, N_2, \ldots, N_n)$, generated by the b^α's, i.e., elements of the form

$$V = \sum_{\alpha \in \Delta} b^\alpha P_\alpha(N), \quad P_\alpha \in \mathbb{C}[N], \tag{2.9}$$

where Δ is a finite subset of $\mathbb{Z}^n$. In fact, it is not difficult to see that every element of the Heisenberg algebra $\mathscr{H}_n$ has a unique representation (2.9). In order to see this we simply use the commutation relations to pair off the operators a_k, c_k in any product of annihilation and creation operators and move all those operators that remain "single" in the process to the left. The fact that there are no relations among the b^α's implies that the representation (2.9) is unique. We now list some formulas that help to bring a given algebra element into the form (2.9), which we shall call "standard." For this purpose we first confine ourselves to the case $n = 1$, taking up the case of general n later.

By induction we easily prove the following pairing off formula:

$$a^m c^m = Z_m(N) \tag{2.10}$$

$(m \geqslant 0)$, where $Z_m(N) \in \mathbb{C}[N]$ is the polynomial

$$Z_m(N) := (N+1)_m = \Gamma(N+1+m)\Gamma(N+1)^{-1}. \tag{2.11}$$

Here $(N+1)_m := (N+1)(N+2)\cdots(N+m)$ is the Pochhammer symbol. Using our definition (2.7) we can write (2.10) in the form

$$b^{-m}b^m = Z_m(N) \quad (m \geqslant 0). \tag{2.12}$$

Formula (2.11) remains valid for $m \leqslant 0$ if we extend the definition of $Z_m(N)$ as follows:

$$Z_m(N) = [\Gamma(N+1+m)\Gamma(N+1)^{-1}]^{\operatorname{sgn} m}, \tag{2.13}$$

where

$$\operatorname{sgn} m = \begin{cases} 1, & m > 0, \\ -1, & m \leqslant 0. \end{cases}$$

We note that $Z_{-m}(N)$ for $m > 0$ is also a polynomial, namely,

$$Z_{-m}(N) = N(N-1)\cdots(N-m+1) := m!\binom{N}{m} \tag{2.14}$$

in obvious notation. Also, for all $m \in \mathbb{Z}$ we have

$$Z_{-m}(N+m) = Z_m(N). \tag{2.15}$$

In bringing a given element of the algebra into standard form the following formula, whose verification we leave to the reader, turns out to be useful:

$$b^k b^l = b^{k+l} b_{k,l}(N). \tag{2.16}$$

Here the polynomial $b_{k,l} \in \mathbb{C}[N]$ is given by the following formula:

$$b_{k,l}(N) = \begin{cases} 1, & \text{for } kl \geqslant 0, \\ Z_k(N+l), & \text{for } kl < 0 \text{ and } |l| \geqslant |k|, \\ Z_l(N), & \text{for } kl < 0 \text{ and } |k| \geqslant |l|. \end{cases} \tag{2.17}$$

If the nonlinear part of the potential $V(a,c)$ happens to be a Hermite polynomial its standard form is especially simple. This follows from the Weyl identity of *formal* power series, valid on the Heisenberg algebra $\mathscr{H}_1$:

$$\exp[-\epsilon^2/2]\exp[\epsilon(a+c)] = \exp(\epsilon c)\cdot\exp(\epsilon a), \tag{2.18}$$

which can be written in the form

$$\sum_{m=0}^{\infty} \frac{1}{m!} h_m\left(\frac{a+c}{\sqrt{2}}\right)\left(\frac{\epsilon}{\sqrt{2}}\right)^m = \sum_{r,s=0}^{\infty} \frac{c^r a^s}{r!s!}\epsilon^{r+s} = \sum_{r,s=0}^{\infty} \frac{b^{r-s} b_{r,-s}(N)}{r!s!}\epsilon^{r+s}.$$

Here $b_{r,-s}$ was defined in (2.17) and h_m is the Hermite polynomial of order m. Comparing terms of order m, we have

$$(m!)^{-1} h_m((M\omega)^{1/2}q)2^{-m/2} = {\sum_{k=-m}^{+m}}' b^k \frac{b_{(m+k)/2,(k-m)/2}(N)}{((m+k)/2)!((m-k)/2)!}. \tag{2.19}$$

On the rhs of (2.19) the prime following the summation symbol indicates that the sum has to be extended only over all those values of the running index which have the same parity as the upper limit. This convention will be in force throughout this paper.

We now turn to the standard form of a simple power of q, or what is the same up to a numerical factor of a power of $(a+c)$. For this purpose we abbreviate expression (2.19) by C_m and then verify that by comparing equal powers of ϵ in the relation that results from multiplying (2.18) by $\exp(\epsilon^2/2)$ we obtain

$$(a+c)^m = m! {\sum_{r=0}^{m}}' C_r \left[\left(\frac{m-r}{2}\right)!\right]^{-1} \left(\frac{1}{2}\right)^{(m-r)/2}. \tag{2.20}$$

Inserting expression (2.20) into (2.19) and changing the order of the two sums we obtain

$$(a+c)^m = {\sum_{k=-m}^{+m}}' b^k P_k^m(N), \tag{2.21}$$

where the polynomial $P_k^m(N)$ is given by the expression

$$P_k^m(N) = m! {\sum_{r \geqslant |k|}^{m}}' \left(\frac{1}{2}\right)^{(m-r)/2} \times \left[\left(\frac{m-r}{2}\right)!\right]^{-1} \frac{b_{(r+k)/2,(k-r)/2}(N)}{((r+k)/2)!((r-k)/2)!}.$$

For $k \geqslant 0$ we set $\frac{1}{2}(r-k) = p$, $\frac{1}{2}(r+k) = p+k \geqslant p$ and we realize that, in view of (2.14) and (2.17), $b_{p+k,-p}(N) = Z_{-p}(N) = p!\binom{N}{p}$. Hence, for $k \geqslant 0$,

$$P_k^m(N) = \sum_{p=0}^{(m-k)/2} \left(\frac{1}{2}\right)^{(m-k)/2-p} \times \frac{m!}{[(m-k)/2-p]!(p+k)!}\binom{N}{p}. \tag{2.22}$$

For $k < 0$ a similar argument shows that

$$P_k^m(N) = P_{-k}^m(N+k). \tag{2.23}$$

Example:

$$(a+c)^4 = b^{-4}P_4^4(N-4) + b^{-2}P_2^2(N-2) + P_0^4(N) + b^2 P_2^4(N) + b^4 P_4^4(N).$$

Here

$$P_4^4(N) = 1,$$

$$P_2^4(N) = \sum_{p=0}^{1} \left(\frac{1}{2}\right)^{1-p} \frac{4!}{(1-p)!(p+2)!}\binom{N}{p} = \frac{1}{2}\frac{4!}{2!} + \frac{4!}{3!}N = 6 + 4N,$$

$$P_0^4(N) = \sum_{p=0}^{2} \left(\frac{1}{2}\right)^{2-p} \frac{4!}{(2-p)!p!}\binom{N}{p} = \frac{1}{4}\cdot\frac{4!}{2!} + \frac{1}{2}\cdot 4!N + \frac{4!}{2!}\cdot\frac{N(N-1)}{2} = 3 + 12N + 6N^2 - 6N = 3 + 6N + 6N^2.$$

Hence

$$(a+c)^4 = a^4 + a^2(4N-2) + (3 + 6N + 6N^2) + c^2(6+4N) + c^4.$$

Another useful formula for bringing an expression into standard form is that for any $m \in \mathbb{Z}$ and any polynomial $P \in \mathbb{C}[N]$ the identity

$$b^m P(N+m) = P(N)b^m \tag{2.24}$$

holds. In the proof of formula (2.24) one has to treat the cases $m > 0$ (i.e., $b = c$) and $m < 0$ (i.e., $b = a$) separately.

In both cases the proof proceeds by induction on m, as well as on the degree of the polynomial $P(N)$.

Thus far we have worked within $\mathscr{H}_1$. Generalizing to the case of $\mathscr{H}_n$ we introduce in $\mathbb{Z}^n$ a partial order by means of the definition

$$\alpha \leqslant \beta \text{ iff } \alpha_i \leqslant \beta_i \quad (i = 1,2,...,n).$$

Accordingly, $\Omega^n := \{\alpha \in \mathbb{Z}^n: \alpha \geqslant 0\}$ is the set of n-tuples of non-negative integers.

For $\alpha,\beta \in \mathbb{Z}^n$ we define

$$Z_\alpha(N) = \prod_{l=1}^{n} Z_{\alpha_l}(N_l), \tag{2.25}$$

$$b_{\alpha,\beta}(N) = \prod_{l=1}^{n} b_{\alpha_l,\beta_l}(N_l). \tag{2.26}$$

Formulas (2.15), (2.16), and (2.24) generalize to

$$Z_{-\alpha}(N+\alpha) = Z_\alpha(N), \tag{2.27}$$

$$b^\alpha b^\beta = b^{\alpha+\beta} b_{\alpha,\beta}(N), \tag{2.28}$$

$$b^\alpha P(N+\alpha) = P(N) b^\alpha. \tag{2.29}$$

Also, formula (2.21) has an obvious generalization to $\mathscr{H}_n$, namely,

$$(a+c)^\alpha = \sum_{-\alpha<\gamma<\alpha}{}' b^\gamma P_\gamma^\alpha(N). \tag{2.30}$$

Here $P_\gamma^\alpha(N)$ is just the product of the polynomials $P_{\gamma_k}^{\alpha_k}(N_k)$ $(k = 1,2,...,n)$.

An element of the algebra is called formally self-adjoint if it is invariant under the involutive antiautomorphism †. Applying this automorphism to the element (2.9) in standard form yields

$$V^\dagger = \sum_{\alpha\in\Delta} \overline{P_\alpha(N)}\, b^{-\alpha} = \sum_{\alpha\in\Delta} b^{-\alpha}\, \overline{P_\alpha(N-\alpha)}$$

$$= \sum_{\alpha\in-\Delta} b^\alpha\, \overline{P_{-\alpha}(N+\alpha)}. \tag{2.31}$$

Hence V is formally self-adjoint iff $\Delta = -\Delta$ and $\overline{P_{-\alpha}(N+\alpha)} = P_\alpha(N)$. Clearly, q^m is formally self-adjoint and it is this property which underlies the symmetry relation (2.23) of the associated polynomials of its standard representation.

From now on we shall assume that our Hamiltonian (1.1) has the form $H^0 + \epsilon V$, where

$$H^0 = \sum_{k=1}^{n} \omega_k \left(N_k + \frac{1}{2}\right) = \langle \omega, N \rangle + \xi. \tag{2.32}$$

Here $\langle\ ,\ \rangle$ is the usual inner product of $\mathbb{R}^n$ and $\xi := \frac{1}{2}\sum_{k=1}^n \omega_k$. Moreover, we shall assume that the potential V is in standard form and is formally self-adjoint. Explicitly, this means that there exists a finite reflection invariant subset Δ of $\mathbb{Z}^n$ so that V has the form $\sum_{\alpha\in\Delta} V_\alpha$, where

$$V_\alpha = b^\alpha P_\alpha(N) = b^\alpha\, \overline{P_{-\alpha}(N+\alpha)}. \tag{2.33}$$

We also find it convenient to introduce the subalgebra $\mathscr{A}(V)$ of the Heisenberg algebra $\mathscr{H}_n$ generated by the V_α's $(\alpha\in\Delta)$, as well as the Lie subalgebra $\mathscr{L}$ generated by the V_α's and H^0. These definitions are meaningful since $\mathscr{H}_n$ can be viewed as an associative algebra as well as a Lie algebra, with the Lie bracket the commutator. In order to indicate that $\mathscr{H}_n$ or an (associative) subalgebra $\mathscr{A}$ is viewed in this fashion the letter L is attached to the symbol as a superscript. Notice that $\mathscr{L} = \mathscr{L}_0 \oplus \mathscr{L}(V)$, where $\mathscr{L}_0 = \text{span } H^0$ and $\mathscr{L}(V) \subset \mathscr{A}(V)^L$ is an ideal in $\mathscr{L}$. Both $\mathscr{L}(V)$ and $\mathscr{A}(V)$ are graded by the natural numbers, with $\mathscr{L}_p(V)$ and $\mathscr{A}_p(V)$ being the linear hulls of all monomials of order p in the V_α's $(\alpha\in\Delta)$. In formulas,

$$\mathscr{A}_p(V) := \text{span}(V_{\alpha_1\alpha_2\cdots\alpha_p}),$$

$$\mathscr{L}_p(V) := \text{span}\{(\text{ad } V)_{\alpha_1\alpha_2\cdots\alpha_{p-1}} V_{\alpha_p}\}, \tag{2.34}$$

where $\alpha_1,\alpha_2,...,\alpha_p$ range over Δ and where we have set

$$V_{\alpha_1\alpha_2\cdots\alpha_p} = V_{\alpha_1} V_{\alpha_2} \cdots V_{\alpha_p}, \tag{2.35a}$$

$$(\text{ad } V)_{\alpha_1\alpha_2\cdots\alpha_p} = \text{ad } V_{\alpha_1} \cdot \text{ad } V_{\alpha_2} \cdots \cdot \text{ad } V_{\alpha_p}. \tag{2.35b}$$

(We adhere to the common usage of the symbol "ad," according to which it associates to each element V of the Lie algebra the linear map defined by the brackets; in formulas, ad $VA = [V,A]$.)

We also present the formula that expresses the Lie monomials in terms of the ordinary ones, written in the form

$$(\text{ad } V)_{\alpha_1\alpha_2\cdots\alpha_{p-1}} V_{\alpha_p}$$

$$= \sum_{\substack{\sigma\cup\tau=(\alpha_1,\alpha_2,...,\alpha_{p-1}) \\ \sigma\cap\tau=\phi}} (-1)^{|\tau|} V_{\vec{\sigma}\alpha_p\overleftarrow{\tau}}. \tag{2.36}$$

Here the summation is over all subsets $\vec{\sigma}$ of $\alpha_1,\alpha_2,...,\alpha_{p-1}$ arranged in the natural order of the corresponding subscripts, whereas $\overleftarrow{\tau}$ denotes the complementary set, arranged in the reverse order. Finally, $|\tau|$ denotes the order of the subset τ.

Since the operation ad is a homomorphism of a Lie algebra into its Lie algebra of linear endomorphisms its application to formula (2.36) yields

$$\text{ad}[(\text{ad } V)_{\alpha_1\alpha_2\cdots\alpha_{p-1}} V_{\alpha_p}]$$

$$= \sum_{\substack{\sigma\cup\tau=(\alpha_1,\alpha_2,...,\alpha_{p-1}) \\ \sigma\cap\tau=\phi}} (-1)^{|\tau|} (\text{ad } V)_{\vec{\sigma}\alpha_p\overleftarrow{\tau}}. \tag{2.37}$$

The claim made above, namely, that $\mathscr{L}$ is a graded Lie algebra, is immediate from formula (2.37).

Note that the vectors $V_{\alpha_1\alpha_2\cdots\alpha_p}$, as well as the vectors (2.35), are eigenvectors to the eigenvalue

$$\langle \omega, \alpha_1 + \alpha_2 + \cdots + \alpha_p \rangle$$

of the operator ad H^0, which thus is exhibited as an operator of degree 0. Moreover, if we introduce the submodule Y_0 (over $\mathbb{Z}$) of $\mathbb{Z}^n$ by means of the definition

$$Y_0 = \{\alpha\in\mathbb{Z}^n: \langle \alpha,\omega \rangle = 0\} \tag{2.38}$$

and let Y_1 be its complement in $\mathbb{Z}^n$ then we see that each homogeneous subspace of $\mathscr{L}(V)$ [or $\mathscr{A}(V)$] splits under ad H_0 into kernel and image, with the former spanned by all vectors (2.35) for which the sum $\alpha_1 + \alpha_2 + \cdots + \alpha_p$ belongs to Y_0 and the latter spanned by those vectors for which the same sum belongs to Y_1. Summarizing, we have

$$\mathscr{L}(V) = \mathscr{N} \oplus \mathscr{M}, \tag{2.39}$$

where $\mathscr{N} = \operatorname{Ker} \operatorname{ad} H^0$, $\mathscr{M} = \operatorname{Im} \operatorname{ad} H^0$, ad H^0 being restricted to $\mathscr{L}(V)$. Notice, also, that on account of the Jacobi identity, $\mathscr{N}$ is a Lie subalgebra of $\mathscr{L}$ and $\mathscr{M}$ is invariant under $\mathscr{L}$.

Before we close this section we give the standard form of a general element (2.35a), which is

$$V_{\alpha_1\alpha_2\cdots\alpha_p} = b^{\alpha_1+\alpha_2+\cdots+\alpha_p} P_{\alpha_1\alpha_2\cdots\alpha_p}(N), \tag{2.40}$$

where

$$P_{\alpha_1\alpha_2\cdots\alpha_p}(N) = \prod_{j=1}^{p} b_{\gamma_p\alpha_j}(N+\beta_{j,p}) P_{\alpha_j}(N+\beta_{j,p}) \tag{2.41}_p$$

and we have used the abbreviations

$$\gamma_j = \alpha_1 + \alpha_2 \cdots + \alpha_{j-1}, \quad \beta_{j,p} = \alpha_{j+1} + \cdots + \alpha_p,$$

which possess the special values $\gamma_1 = \beta_{p,p} = 0$. The proof of (2.40) is based on the formula

$$b^{\alpha} P(N) b^{\beta} Q(N) = b^{\alpha+\beta} b_{\alpha,\beta}(N) P(N+\beta) Q(N), \tag{2.42}$$

which expresses the product of two given elements in standard form again in standard form (here $P,Q \in \mathbb{C}[N]$; $\alpha,\beta \in \mathbb{Z}^n$). Clearly, formula (2.40) is valid in the case $p = 1$. Assuming its validity for a given p we invoke (2.42) to show that

$$P_{\alpha_1\alpha_2\cdots\alpha_p\alpha_{p+1}}(N) = b_{\gamma_{p+1},\alpha_{p+1}}(N) P_{\alpha_{p+1}}(N) P_{\alpha_1\alpha_2\cdots\alpha_p}(N+\alpha_{p+1}). \tag{2.43}$$

However, the rhs of Eq. (2.43) agrees precisely with what we obtain if we split off in $(2.41)_{p+1}$ the last term and express the remainder in terms of the lhs of $(2.41)_p$.

III. PERTURBATION THEORY

In the following we develop a perturbation theory for the Schrödinger equation

$$i\dot{\psi} = (H_0 + \epsilon V)\psi \tag{3.1}$$

based on normal form theory. Here we may think of H^0 as being $\langle \omega, N \rangle + \xi$ and of V being an element of $\mathscr{H}_n$ in standard form.

However, we find it convenient to formulate our approach to peturbation theory within the framework of an algebra $\mathscr{C}$ which is either associative or Lie depending on the context. We also introduce the algebra $\mathscr{C}^\infty$ of a formal power series in the parameter ϵ with coefficients in $\mathscr{C}$. An element $C \in \mathscr{C}^\infty$ will be written in the form

$$C = \sum_{j=0}^{\infty} C^{(j)} \epsilon^j, \quad C^{(j)} \in \mathscr{C}. \tag{3.2}$$

Moreover, we will use the notation

$$C^{[m]} = \sum_{j=0}^{m} C^{(j)} \epsilon^j, \quad C^{\{m\}} = \sum_{j=m+1}^{\infty} C^{(j)} \epsilon^j \tag{3.3}$$

so that

$$C = C^{[m]} + C^{\{m\}}. \tag{3.4}$$

Correspondingly, $\mathscr{C}^{[m]}$ will denote the subspace of polynomials in ϵ of degree $\leqslant m$ with coefficients in $\mathscr{C}$, whereas $\mathscr{C}^{\{m\}}$ will denote the ideal in $\mathscr{C}^\infty$ of a formal power series starting with a term on the order of at least $m+1$. We also note the formula

$$A^{[m]} \cdot B^{[m]} - (A \cdot B)^{[m]} = \epsilon^{m+1} \rho_m(A,B,\epsilon), \tag{3.5}$$

where

$$\rho_m(A,B,\epsilon) = \sum_{r=0}^{m-1} \epsilon^r \sum_{l=r+1}^{m} A^{(l)} B^{(m+1+r-l)}, \tag{3.6}$$

which will play a crucial role in the following discussion.

Note that for $A,S \in \mathscr{C}^\infty$ the expression $C = \exp(\epsilon S) A$ is well defined in terms of the usual power series as an element of $\mathscr{C}^\infty$ since each coefficient $C^{(j)}$ involves a finite sum only. In case $\mathscr{C} := \mathscr{L}$ is a Lie algebra, $\exp(\epsilon S)$ is written as $\exp(\epsilon \operatorname{ad} S)$ and is easily seen to be an automorphism of $\mathscr{C}^\infty$. If $\mathscr{C} := \mathscr{A}$ is an associative algebra and $\mathscr{A}^L$ is the corresponding Lie algebra (with the commutator being the Lie bracket) then

$$\exp(\epsilon \operatorname{ad} S) A = \exp(\epsilon S) A \exp(-\epsilon S) \tag{3.7}$$

holds for $A \in \mathscr{A}^\infty$. In particular, $\exp(\epsilon \operatorname{ad} S)$ is also an automorphism of the algebra $\mathscr{A}^\infty$.

Like normal form theory of classical mechanics (see Refs. 10 and 13) our approach to quantum mechanical perturbation theory is best formulated within the framework of an abstract Lie algebra $\mathscr{L}$ that splits under ad H^0:

$$\mathscr{L} = \mathscr{M} \oplus \mathscr{N}, \tag{3.8}$$

with $\mathscr{M} = \operatorname{Im} \operatorname{ad} H^0$ and $\mathscr{N} = \operatorname{Ker} \operatorname{ad} H^0$. If $A \in \mathscr{L}$, $\tilde{A}$ and $\hat{A}$ will denote the components of A along $\mathscr{M}$ and $\mathscr{N}$, respectively. The splitting (3.8) obviously carries over to the Lie algebra $\mathscr{L}^\infty$ of the formal power series in ϵ over $\mathscr{L}$; the same holds for the associated projections, which we will continue to denote by tilde and caret.

Our perturbation theory is based on the idea of bringing the perturbation ϵV into normal form, i.e., into a form in which it commutes with H^0 by means of a suitable transformation. For our purpose we find it best to write the normalizing transformation as a product of exponentials, each of which transforms the next higher order term into normal form.

If $S_1 \in \mathscr{L}$ is any solution of the commutator equation

$$[H^0, S_1] = \tilde{V},$$

then by comparing terms of order ϵ on both sides of

$$\exp(\epsilon \operatorname{ad} S_1)(H^0 + \epsilon V) = H^0 + \epsilon H^{(1)} + \epsilon^2 V_2, \quad V_2 \in \mathscr{L}^\infty, \tag{3.9}$$

we obtain $H^{(1)} = \hat{V}$. Assuming inductively that $S_1, S_2, \ldots, S_{m-1}$ and $H^{(1)}, H^{(2)}, \ldots, H^{(m-1)}$ belonging to $\mathscr{L}$ and $V_m \in \mathscr{L}^\infty$ have been constructed, we let S_m be a solution of

$$[H^0, S_m] = \tilde{V}_m^{(0)} \tag{3.10}$$

and we verify, by comparing terms of order ϵ^m on both sides of

$$\exp(\epsilon^m \operatorname{ad} S_m)(H^{[m-1]} + \epsilon^m V_m) = H^{[m]} + \epsilon^{m+1} V_{m+1} \tag{3.11}$$

with suitable $V_{m+1} \in \mathscr{L}^\infty$, that $H^{(m)} = \hat{V}_m^{(0)}$ holds.

Combining these transformations into a single one we may write

$$H^0 + \epsilon V = \mathscr{R}_m(S)(H^{[m]} + \epsilon^{m+1} V_{m+1}), \tag{3.12}$$

where

$$\mathscr{R}_m(S) := \exp(-\epsilon \operatorname{ad} S_1)\exp(-\epsilon^2 \operatorname{ad} S_2) \times \cdots \exp(-\epsilon^m \operatorname{ad} S_m). \tag{3.13}$$

Setting for $\eta = (\eta_1, \eta_2, \ldots, \eta_m) \in \Omega^m$ ($= m$-tuples of non-negative integers),

$$|\eta| = \sum_{r=1}^{n} |\eta_r|, \quad \eta^* = \sum_{r=1}^{n} r\eta_r, \quad n! = \eta_1! \cdot \eta_2! \cdots \eta_m!$$

and

$$(\operatorname{ad} S)^\eta = (\operatorname{ad} S_1)^{\eta_1} \cdot (\operatorname{ad} S_2)^{\eta_2} \cdot \cdots \cdot (\operatorname{ad} S_m)^{\eta_m},$$

we can write the lth coefficient of $\mathscr{R}_m(S)$ in the form

$$\mathscr{R}_m^{(l)}(S) = \sum_{\substack{\eta \in \Omega^m \\ \eta^* = l}} \frac{(-1)^{|\eta|}}{\eta!} (\operatorname{ad} S)^\eta. \tag{3.14}$$

Comparing terms of order ϵ^{m+1} $(m \geqslant 1)$ on both sides of (3.12) we obtain the formula

$$V_{m+1}^{(0)} = -\sum_{l=1}^{m+1} \mathscr{R}_m^{(l)}(S) H^{(m+1-l)}, \tag{3.15}$$

which expresses the coefficient of ϵ^0 in V_{m+1} in terms of H^0, $H^{(1)}, \ldots, H^{(m)}$, and $S_1, S_2, \ldots, S_m$. A similar formula could be derived for the higher order coefficients of V_{m+1}, but since they are not needed we refrain from presenting them here.[12] For $m = 1$ we obtain

$$V_2^{(0)} = -\mathscr{R}_1^{(1)}(S) H^{(1)} - \mathscr{R}_1^{(2)}(S) H^0,$$

where $\mathscr{R}_1^{(1)}(S) = -\operatorname{ad} S_1$, $\mathscr{R}_1^{(2)}(S) = \frac{1}{2}(\operatorname{ad} S_1)^2$, and therefore

$$V_2^{(0)} = \operatorname{ad} S_1 \hat{V} + \tfrac{1}{2} \operatorname{ad} S_1 (V - \hat{V}) = \tfrac{1}{2} \operatorname{ad} S_1 (V + \hat{V})$$

and $H^{(2)} = \hat{V}_2^{(0)}$. Since on account of the Jacobi identity $\mathscr{M}$ is stable under $\mathscr{N}$ we may write

$$H^{(2)} = \tfrac{1}{2}(\operatorname{ad} S_1 \tilde{V})\hat{}, \tag{3.16}$$

provided only that we normalize S_1 to lie in $\mathscr{M}$. (Compare with Ref. 10.)

The foregoing development is all valid in the abstract framework of a Lie algebra $\mathscr{L}$ and its power series extension $\mathscr{L}^\infty$. We now return to the situation of our quantum mechanical problem in which $\mathscr{L}$ is realized as a Lie subalgebra of $\mathscr{H}_n^L$, generated by H^0 and $(V_\alpha)_{\alpha \in \Delta}$. Since $V \in \mathscr{L}_1(V)$ it follows from an easy induction argument based on formula (3.15) that in fact S_p and $H^{(p)}$ are in $\mathscr{L}_p(V) \subset \mathscr{A}_p(V)$, i.e., they lie in the linear hull of vectors $V_{\alpha_1 \cdots \alpha_p}$ $(\alpha_1, \alpha_2, \ldots, \alpha_p \in \Delta)$. Moreover, if we normalize S by the requirement $\hat{S} = 0$, an induction argument based on formula (3.15) shows that if V is formally self-adjoint so is iS_p and $H^{(p)}$.

We now proceed to show how the transformation of $H_0 + \epsilon V$ into normal form can be used to define approximate solutions to the Schrödinger initial value problem

$$i\dot{\psi} = (H_0 + \epsilon V)\psi, \quad \psi(0) = \psi_0. \tag{3.17}$$

Moreover, we shall describe a procedure that allows us to estimate the associated error. Finally, we shall take up the task of making these constructions rigorous. For the present we assume that our Heisenberg algebra is realized as the subalgebra of an operator algebra over some function space on which also the operator

$$U_m(t) = \exp(-iH^{[m]}t) \tag{3.18}$$

makes sense. The mth approximant to the initial value problem is defined by means of the formula

$$\phi_m(t) = [\mathscr{R}_m^{[m]}(S) U_m(t)]\psi_0. \tag{3.19}$$

Explicitly, for $m = 0,1,2,3$, the approximations are

$$\begin{aligned}
\phi_0(t) &= U_0(t)\psi_0, \\
\phi_1(t) &= (1 - \epsilon T_1) U_1(t), \\
\phi_2(t) &= \{1 - \epsilon T_1 + \epsilon^2[\tfrac{1}{2}T_1^2 - T_2]\} U_2(t)\psi_0, \\
\phi_3(t) &= \{1 - \epsilon T_1 + \epsilon^2[\tfrac{1}{2}T_1^2 - T_2] \\
&\quad + \epsilon^3[-\tfrac{1}{6}T_1^3 + T_1 T_2 - T_3]\} U_3(t)\psi_0.
\end{aligned} \tag{3.20}$$

Here we have abbreviated $T_k = \operatorname{ad} S_k$ $(k = 1,2,3)$.

Proceeding informally at first, we assume that we can differentiate $\phi_m(t)$ and that the derivative is the expected one. Introducing the quantity

$$\mu_m(t) := \sum_{r=0}^{m-1} \epsilon^r \sum_{l=r+1}^{m} \mathscr{R}_m^{(l)}(S)(H^{(m+1+r-l)} U_m(t))\psi_0, \tag{3.21}$$

we find, in view of (3.5),

$$\begin{aligned}
i\dot{\phi}_m &= \mathscr{R}_m^{[m]}(S)(H^{[m]} U_m(t))\psi_0 \\
&= [\mathscr{R}_m(S)(H^{[m]} U_m(t))]^{[m]}\psi_0 + \epsilon^{m+1}\mu_m(t).
\end{aligned} \tag{3.22}$$

Here we view $U_m(t)$ as being of order ϵ^0. Since $\exp(\epsilon^j \operatorname{ad} S)$, for $S \in \mathscr{H}_n$, $j \geqslant 1$, is an automorphism in the algebra of a formal power series with coefficients in the given operator algebra, we may write

$$\begin{aligned}
i\dot{\phi}_m &= [\mathscr{R}_m(S) H^{[m]} \mathscr{R}_m(S) U_m(t)]^{[m]}\psi_0 + \epsilon^{m+1}\mu_m(t) \\
&= [(H^0 + \epsilon V)\mathscr{R}_m(S) U_m(t)]^{[m]}\psi_0 + \epsilon^{m+1}\mu_m(t) \\
&= (H^0 + \epsilon V)\phi_m(t) + \epsilon^{m+1}(\mu_m - r_m),
\end{aligned}$$

where $r_m(t) = V[\mathscr{R}_m^{(m)}(S) U_m(t)]\psi_0$. Summarizing, we see that ϕ_m satisfies the initial value problem

$$i\dot{\phi}_m = (H^0 + \epsilon V)\phi_m + \epsilon^{m+1} g_m(t), \quad \phi_m(0) = \psi_0, \tag{3.23}$$

which differs from the Schrödinger initial value problem by the presence of the term $g_m(t)$. Explicitly,

$$\begin{aligned}
g_m(t) = &\sum_{r=0}^{m-1} \epsilon^r \sum_{l=r+1}^{m} \mathscr{R}_m^{(l)}(S)(H^{(m+1+r-l)} U_m(t))\psi_0 \\
&- V[\mathscr{R}_m^{(m)}(S) U_m(t)]\psi_0.
\end{aligned} \tag{3.24}$$

In Sec. IV it will become clear that these derivations can be justified rigorously in a Hilbert space setting. It then follows from an argument similar to the one of Lemlih and Ellison[9] that $\phi_m(t)$ deviates in norm from the exact solution $\psi(t)$ of the Schrödinger initial value problem by a term of order $\epsilon^{m+1}|t|$ for $\epsilon \downarrow 0$, i.e.,

$$\|\phi_m(t) - \psi(t)\| = \mathscr{O}(\epsilon^{m+1}|t|), \tag{3.25}$$

provided only that the initial vector ψ_0 is confined to a suitable dense subspace of the Hilbert space.

IV. RIGOROUS DETERMINATION OF THE INITIAL DOMAIN

In order to make the foregoing development rigorous we introduce the space $\mathscr{E}_n$ of entire functions of n complex variables $z := (z_1, z_2, \ldots, z_n)$ and realize the algebra elements a_k and c_k as operators over $\mathscr{E}_n$ of partial differentiation with respect to z_k and multiplication by z_k, respectively, i.e.,

$$(a_k f)(z) = \frac{\partial f}{\partial z_k}(z), \quad (c_k f)(z) = z_k f(z). \tag{4.1}$$

On the space $\mathscr{E}_n$ we introduce a pseudonorm by means of the definition

$$\|f\|^2 := \int |f(z)|^2 \, d\mu_n(z), \tag{4.2}$$

where the measure $d\mu_n(z)$ on $\mathbb{C}^n$ is given by the expression

$$d\mu_n(z) = \pi^{-n} e^{-\bar{z}\cdot z} \, d^n x \, d^n y.$$

Here

$$\bar{z}\cdot z = \sum_{k=1}^{n} |z_k|^2, \quad z_k = x_k + i y_k,$$

$$d^n x = dx_1 \, dx_2 \cdots dx_n,$$

etc. Obviously, $\mathscr{E}_n$ contains elements of infinite norm. However, it is well known that the linear subspace

$$\mathscr{F}_n = \{ f \in \mathscr{E}_n : \|f\| < \infty \} \tag{4.3}$$

of $\mathscr{E}_n$ is a Hilbert space. In fact, Bargmann[14–16] has amply demonstrated that the Hilbert space $\mathscr{F}_n$ is better suited for many applications in physics than the more conventional Hilbert space $L^2(\mathbb{R}^n)$ (see, also, Ref. 17).

Moreover, Bargmann[14–16] has demonstrated that there exists a unitary integral transformation from $L^2(\mathbb{R}^n)$ to $\mathscr{F}_n$ whose kernel essentially agrees with the generating function of the Hermite functions. Explicitly, the transformation is given by the formula

$$f(z) = \int A_n(z,q) \psi(q) d^n q, \tag{4.4}$$

where

$$A_n(z,q) := \pi^{-n/4} \exp\left[-\tfrac{1}{2}(z^2 + q^2) + \sqrt{2}\, z\cdot q \right]$$

$$= \sum_{\eta > 0} u_\eta(z) \Phi_\eta(q). \tag{4.5}$$

The symbols in expression (4.5) have the following meaning:

$$u_\eta(z) := (\eta!)^{-1/2} z^\eta, \quad \eta \in \Omega^n, \tag{4.6}$$

is a complete orthonormal system of functions in $\mathscr{F}_n$ and

$$\Phi_\eta(q) = (2^{|\eta|}\eta!)^{-1/2} \pi^{-n/4} \exp\left(-\frac{1}{2} q^2 \right) \prod_{l=1}^{n} h_{\eta_l}(q_l) \tag{4.7}$$

[$h_m(q)$ = Hermite polynomial of order m] is such a system in $L^2(\mathbb{R}^n)$.

The Heisenberg algebra $\mathscr{H}_n$, as well as the subalgebras $\mathscr{A}(V)$, $\mathscr{L}$, now appear in the disguise of algebras of differential operators over $\mathscr{E}_n$. By restriction these operators become unbounded operators on the Hilbert space $\mathscr{F}_n$. In fact, any differential operator A over $\mathscr{E}_n$ restricts to an operator with natural maximal domain

$$\mathscr{D}(A) := \{ f \in \mathscr{F}_n : Af \in \mathscr{F}_n \}. \tag{4.8}$$

Since the complete orthonormal system $u_\eta(z)$, $\eta \in \Omega^n$ [see (4.6)], is contained in $\mathscr{D}(A)$ the linear hull of these vectors constitutes a dense invariant domain for all our operators. This implies in particular that $\mathscr{D}(A)$, as defined in (4.8), is dense in $\mathscr{F}_n$.

Notice that if $A, B \in \mathscr{H}_n$ then

$$\mathscr{D}(A) \cap \mathscr{D}(B) \subset \mathscr{D}(A + B).$$

Hence, if for θ: $\Omega^n \to [0, \infty)$ we associate the domain

$$\mathscr{B}(\theta) := \left\{ f \in \mathscr{F}_n : \sum_{\eta \in \Omega^n} |\langle f, u_\eta \rangle|^2 \theta(\eta) < \infty \right\}, \tag{4.9}$$

then the domain

$$\bigcap_{\alpha \in \Delta} \mathscr{B}(\theta_\alpha) \quad \text{with} \quad \theta_\alpha(\eta) = Z_\alpha(\eta) |P_\alpha(\eta)|^2 \tag{4.10}$$

is contained in the maximal domain $\mathscr{D}(V)$ of the operator $V = \Sigma_{\alpha \in \Delta} b^\alpha P_\alpha(N)$.

We have seen in Sec. II that the operators

$$V = \sum_{\alpha \in \Delta} b^\alpha P_\alpha(N), \quad W = \sum_{\alpha \in -\Delta} b^\alpha \, \overline{P_{-\alpha}(N + \alpha)} \tag{4.11}$$

are formally adjoints of each other. In fact, it turns out that they are adjoint of each other as operators on the Hilbert space $\mathscr{F}_n$, i.e., we shall prove the following theorem.

Theorem: If V and W are as in (4.11) then $V^* = W$ as operators over $\mathscr{F}_n$.

Remark: Remember $\mathscr{D}(V^*) = \{ f \in \mathscr{F}_n : g \to \langle f, Vg \rangle$ is a continuous linear functional on $\mathscr{D}(V) \}$ and $V^* f$ is by definition the unique vector in $\mathscr{F}_n$, so that

$$\langle f, Vg \rangle = \langle V^* f, g \rangle, \tag{4.12}$$

for all $g \in \mathscr{D}(V)$.

The proof of this theorem is based on the following lemma.

Lemma: For any $f \in \mathscr{E}_n$ we have

$$Wf = \sum_{\eta \in \Omega^n} \langle V u_\eta, f \rangle u_\eta. \tag{4.13}$$

We first deduce the theorem from the lemma.

Assume $f \in \mathscr{D}(V^*)$. Then from (4.12) for all $\eta \in \Omega$,

$$\langle V u_\eta, f \rangle = \langle u_\eta, V^* f \rangle.$$

Hence, from (4.13), $Wf = V^* f \in \mathscr{F}_n$. This proves $V^* \subset W$. Now, assume $f \in \mathscr{D}(W)$. Then, by definition of $\mathscr{D}(W)$, $Wf \in \mathscr{F}_n$ and from (4.13) we deduce

$$\langle V u_\eta, f \rangle = \langle u_\eta, Wf \rangle$$

for all $\eta \in \Omega^n$. Now, let $g \in \mathscr{D}(V)$. Since g is an entire function we have

$$Vg = \sum_{\eta \in \Omega^n} \langle g, u_\eta \rangle V u_\eta \in \mathscr{F}_n.$$

By continuity of the inner product we find

$$\langle Vg,f\rangle = \sum_{\eta\in\Omega^n} \langle g,u_\eta\rangle\langle Vu_\eta,f\rangle$$

$$= \sum_{\eta\in\Omega^n} \langle g,u_\eta\rangle\langle u_\eta,Wf\rangle = \langle g,Wf\rangle.$$

It follows that $f\in\mathscr{D}(V^*)$ and $V^*f = Wf$, i.e., $W\subset V^*$. By combining this result with the converse inclusion the deduction of the theorem from the lemma is complete. Finally, the statement of the lemma follows from the following computation:

$$Wf = \sum_{\alpha\in\Delta}\sum_{\eta\in\Omega^n} \langle u_\eta,f\rangle [Z_{-\alpha}(\eta)]^{1/2}\overline{P_\alpha(\eta-\alpha)}\, u_{\eta-\alpha}$$

$$= \sum_{\alpha\in\Delta}\sum_{\eta\in\Omega^n} \langle u_{\eta+\alpha},f\rangle [Z_\alpha(\eta)]^{1/2}\overline{P_\alpha(\eta)}\, u_\eta$$

$$= \sum_{\eta\in\Omega^n} \langle Vu_\eta,f\rangle u_\eta.$$

Corollary 1: All elements of $\mathscr{H}_n$ define closed operators over $\mathscr{F}_n$.

Corollary 2: An element of $\mathscr{H}_n$ defines a self-adjoint operator over $\mathscr{F}_n$ iff it is formally self-adjoint.

Notice that by our definition of the domain of an operator $A\in\mathscr{H}_n$ neither $\mathscr{D}(BA)\subset\mathscr{D}(A)$ nor $A\mathscr{D}(BA)\subset\mathscr{D}(B)$ need to be true. In the following we shall denote the maximal domain of $V_{\alpha_1\alpha_2\cdots\alpha_p}$ simply by $\mathscr{D}(\alpha_1\alpha_2\cdots\alpha_p)$. According to what we have just stated it need not be true that $\mathscr{D}(\alpha_k\alpha_{k+1}\cdots\alpha_p)$ $(k=1,2,\ldots,p)$ is a chain of domains, each contained in the next one, nor need it be true that V_{α_p} maps $\mathscr{D}(\alpha_1\alpha_2\cdots\alpha_p)$ into $\mathscr{D}(\alpha_1\alpha_2\cdots\alpha_{p-1})$. However, if we define

$$\delta(\alpha_1\alpha_2\cdots\alpha_p) = \bigcap_{k=1}^{p}\mathscr{D}(\alpha_k\alpha_{k+1}\cdots\alpha_p) \tag{4.14}$$

then these statements are true with $\mathscr{D}$ replaced by δ. The first statement is a consequence of our definition and the truth of the second is seen as follows. Assume $f\in V_{\alpha_p}\delta(\alpha_1\alpha_2\cdots\alpha_p)$. Then there exists

$$g\in\delta(\alpha_1\alpha_2\cdots\alpha_p)\subset\delta(\alpha_k\alpha_{k+1}\cdots\alpha_p)\subset\mathscr{D}(\alpha_p)$$

so that $V_{\alpha_p}g = f$. Hence $f\in\mathscr{F}_n$, and

$$V_{\alpha_k\alpha_{k+1}\cdots\alpha_{p-1}}f = V_{\alpha_k\alpha_{k+1}\cdots\alpha_{p-1}\alpha_p}g\in\mathscr{F}_n,$$

and therefore $f\in\mathscr{D}(\alpha_k\alpha_{k+1}\cdots\alpha_{p-1})$, for $k=1,2,\ldots,p-1$. This is exactly what we wanted to show.

Notice that, according to (2.40) and (4.9),

$$\mathscr{D}(\alpha_1\alpha_2\cdots\alpha_p) = \mathscr{B}(\theta_{\alpha_1\alpha_2\cdots\alpha_p}), \tag{4.15}$$

where

$$\theta_{\alpha_1\alpha_2\cdots\alpha_p}(\eta) := Z_{\alpha_1+\alpha_2+\cdots+\alpha_p}(\eta)|P_{\alpha_1\alpha_2\cdots\alpha_p}(\eta)|^2. \tag{4.16}$$

Moreover, we introduce the domain

$$\mathscr{D}^0(\alpha_1\alpha_2\cdots\alpha_p) := \mathscr{D}(V_{\alpha_1\alpha_2\cdots\alpha_p}H^{00}) = \mathscr{B}(\theta^0_{\alpha_1\alpha_2\cdots\alpha_p}), \tag{4.17}$$

where $H^{00} = \langle\omega,N\rangle$ and therefore

$$\theta^0_{\alpha_1\alpha_2\cdots\alpha_p}(\eta) = |\langle\omega,\eta\rangle|^2\theta_{\alpha_1\cdots\alpha_p}(\eta). \tag{4.18}$$

Finally, we set

$$\delta^0(\alpha_1\alpha_2\cdots\alpha_p) = \mathscr{D}^0(\alpha_1\alpha_2\cdots\alpha_p)\cap\delta(\alpha_1\alpha_2\cdots\alpha_p). \tag{4.19}$$

For any $r=0,1,2,\ldots,p$, we have

$$V_{\alpha_1\cdots\alpha_r}H^0V_{\alpha_{r+1}\cdots\alpha_p} = V_{\alpha_1\alpha_2\cdots\alpha_p}(H^{00}+\xi^p_r),$$

where $\xi^p_r = \xi + \Sigma_{i=r+1}^p\langle\alpha_i,\omega\rangle$. Since, for any $A,B\in\mathscr{H}_n$,

$$\mathscr{D}(A+B)\cap\mathscr{D}(A) = \mathscr{D}(B)\cap\mathscr{D}(A),$$

it follows that

$$\mathscr{D}(V_{\alpha_1\cdots\alpha_r}H^0V_{\alpha_{r+1}\cdots\alpha_p})\cap\mathscr{D}(\alpha_1\cdots\alpha_p) = \mathscr{D}^0(\alpha_1\cdots\alpha_p)\cap\mathscr{D}(\alpha_1\cdots\alpha_p)$$

and therefore

$$\mathscr{D}(V_{\alpha_1\cdots\alpha_r}H^0V_{\alpha_{r+1}\cdots\alpha_p})\cap\delta(\alpha_1\alpha_2\cdots\alpha_p) = \delta^0(\alpha_1\alpha_2\cdots\alpha_p).$$

Setting $r=0$ in the last relation and reasoning as above in the case of $\delta(\alpha_1\alpha_2\cdots\alpha_p)$ we also have

$$V_{\alpha_p}\delta^0(\alpha_1\alpha_2\cdots\alpha_p)\subset\delta^0(\alpha_1\alpha_2\cdots\alpha_{p-1}).$$

Finally, we introduce the domains

$$\delta_p(\Delta) = \bigcap_{\alpha_1,\ldots,\alpha_p\in\Delta}\delta(\alpha_1\alpha_2\cdots\alpha_p) \tag{4.20}$$

and similarily $\delta^0(\Delta)$. Notice that

$$V_{\alpha_p}\delta_p(\Delta)\subset V_{\alpha_p}\bigcap_{\alpha_1,\ldots,\alpha_{p-1}\in\Delta}\delta(\alpha_1\alpha_2\cdots\alpha_{p-1}\alpha_p)$$

$$\subset\delta_{p-1}(\Delta)$$

and similarily for δ^0_p. More generally, if p, q are two natural numbers such that $q<p$ and if $A\in\mathscr{A}_q(V)$ then

$$A\delta_p(\Delta)\subset\delta_{p-q}(\Delta),\quad A\delta^0_p(\Delta)\subset\delta^0_{p-q}(\Delta). \tag{4.21}$$

Our fundamental estimate is based on the following Gronwall-like lemma (compare with Ref. 9).

Main Lemma: Let $\mathscr{K}$ be a Hilbert space. Let I be an open interval containing 0. Assume ϕ, ψ are two continuously differentiable functions defined on I with values in the domain $\mathscr{D}(A)$ of a symmetric operator A over $\mathscr{K}$. Assume further that ϕ, ψ satisfy the differential equations

$$i\dot\phi = A\phi + g,\quad i\dot\psi = A\psi, \tag{4.22}$$

where $g: I\to\mathscr{K}$ is a continuous function. Then the following estimate holds:

$$\|\phi(t)-\psi(t)\|\leqslant\|\phi(0)-\psi(0)\| + \left|\int_0^t\|g(s)\|ds\right|. \tag{4.23}$$

In order not to interrupt our main argument we relegate the proof of this lemma to the Appendix. Of course, for the application we have in mind the symmetric operator is $A = H_0 + \epsilon V$, $\phi = \phi_m$ [defined as the mth approximation in (3.19)] and the error term g_m is defined in (3.24). In order to state the main theorem of this paper we need the concept of nonresonance of order m.

Definition: The potential V is called nonresonant of order m with respect to H^0 if

$$\underbrace{\Delta+\Delta+\cdots+\Delta}_{k}\cap Y_0 = \{0\}\quad (1\leqslant k\leqslant m). \tag{4.24}$$

Using this concept our main theorem can now be stated as follows.

Main Theorem: Assume V is nonresonant of order m with respect to H^0. Then the approximate solution ϕ_m [defined in (3.19)] satisfies an estimate of type (3.25) provided only ψ_0 is restricted to the domain

$$\delta_{2m}(\Delta)\cap\delta_m^0(\Delta).$$

The proof of the main theorem is broken up into a series of lemmas.

Lemma 1: If V is nonresonant of order p with respect to H^0 then $\delta_p(\Delta)$ is invariant under $U_p(t,\epsilon) := \exp(-iH^{[p]}(\epsilon)t)$.

Proof: Since by assumption V is (formally) self-adjoint an easy induction shows that the same is true for each element $H^{(p)}$, $p\geqslant 1$, hence also for $H^{[p]}$ and ϵ real. However, we know that formal self-adjointness implies genuine self-adjointness on the maximal domain. By Stone's theorem $\exp(-iH^{[p]}t)$ is unitary. In case V is nonresonant of order p $H^{[p]}$ is seen easily to be a function $C(N,\epsilon)$ of the number operators N_k $(k=1,2,\ldots,n)$ and ϵ only. Since $\delta_p(\Delta)$ is the intersection of sets of type $\mathscr{B}(\theta)$ [see (4.9)] it is sufficient to prove that such a set is invariant under $\exp(-iC(N,\epsilon)t)$. However, in view of the identity

$$\langle u_\eta,\exp(-iC(N,\epsilon)t)f\rangle = \exp(-iC(\eta,\epsilon)t)\langle u_\eta f\rangle$$

this is self evident.

Lemma 2:

$$\phi_m(t) := \sum_{l=0}^{m}\epsilon^l\sum_{\alpha^*=l}(-1)^{|\alpha|}\frac{(\mathrm{ad}\,S)^\alpha}{\alpha!}U_m(t,\epsilon)\psi_0$$

is well defined for $\psi_0\in\delta_m(\Delta)$.

Proof: Evidently, it is sufficient to prove that $(\mathrm{ad}\,S)^\alpha U_m(t,\epsilon)\psi_0$ with $\alpha^*=l$ is well defined for $\psi_0\in\delta_l(\Delta)$. Remembering that

$$(\mathrm{ad}\,S)^\alpha = (\mathrm{ad}\,S_1)^{\alpha_1}(\mathrm{ad}\,S_2)^{\alpha_2}\cdots(\mathrm{ad}\,S_l)^{\alpha_l}$$

it is clear that

$$(\mathrm{ad}\,S)^\alpha U_m(t,\epsilon) = \sum_{0\leqslant\beta\leqslant\alpha}\binom{\alpha}{\beta}S^{\alpha-\beta}U_m(t,\epsilon)S^{\overleftarrow{\beta}}(-1)^\beta,$$

where $\overleftarrow{\beta}=(\beta_l,\beta_{l-1},\ldots,\beta_1)$. Since $S^{\overleftarrow{\beta}}$ maps $\delta_l(\Delta)$ into $\delta_{l-\beta^*}(\Delta)$, which is left invariant by $U_m(t,\epsilon)$ and mapped into $\delta_{l-\alpha^*}(\Delta)=\delta_0(\Delta)=\mathscr{F}_n$ by $S^{\alpha-\beta}$, Lemma 2 is proved.

Lemma 3: Let $(\chi_\eta)_{\eta\in\Omega^n}$ be a set of continuous complex valued functions defined on an open interval J about 0, each of which is bounded by a positive constant M_η. If $\Sigma_{\eta\in\Omega^n}M_\eta<\infty$ then $\Sigma_{\eta\in\Omega^n}\chi_\eta(h)$ is well defined and continuous on J. In particular,

$$\lim_{h\to 0}\sum_{\eta\in\Omega^n}\chi_\eta(h)=\sum_{\eta\in\Omega^n}\chi_\eta(0).$$

Proof: By adding

$$\sum_{\substack{\eta\in\Omega^n\\|\eta|=k}}\chi_\eta=\chi_k,\qquad \sum_{\substack{\eta\in\Omega^n\\|\eta|=k}}\mu_\eta=M_k\quad\left(|\eta|:=\sum_{i=1}^n\eta_i\right),$$

the general case is reduced to the case $n=1$, in which it is an immediate consequence of the Cauchy criterion.

Lemma 4: The formula $v(t)=A\exp(-iH^{00}t)\psi_0$ for $A\in\mathscr{A}_p(V)$ and $\psi_0\in\delta_p^0(\Delta)$ is differentiable, with the expected derivative $v_1(t):=-iAH^{00}\exp(-\mathrm{iH^{00}t})\psi_0$.

Proof: By linearity it is obviously sufficient to prove Lemma 4 for the case of $A=V_{\alpha_1\alpha_2\cdots\alpha_p}$. Since

$$[v(t+h)-v(t)]/h-v_1(t) = V_{\alpha_1\cdots\alpha_p}\exp(-iH^{00}t)[(e^{-iH^{00}h}-1)/h+iH^{00}]\psi_0,$$

we find

$$\left|\left|\frac{v(t+h)-v(t)}{h}-v_1(t)\right|\right|^2 = \sum_{\eta\in\Omega^n}|\langle u_\eta,\psi_0\rangle|^2\theta_{\alpha_1\cdots\alpha_p}(\eta)\times\left|\frac{e^{-i\langle\eta,\omega\rangle h}-1}{h}+i\langle\eta,\omega\rangle\right|^2. \tag{4.25}$$

From $e^{-iah}-1=-ia\int_0^h e^{-iat}dt$ we conclude that $|e^{-iah}-1|\leqslant|h||a|$. Hence the η term in the sum (4.25) is bounded by the constant

$$M_\eta=4|\langle u_\eta,\psi_0\rangle|^2\theta_{\alpha_1\cdots\alpha_p}(\eta)|\langle\eta,\omega\rangle|^2.$$

Since by definition of $\delta_p^0(\Delta)$: $\Sigma_{\eta\in\Omega^n}M_\eta$ is convergent the statement of Lemma 4 follows from Lemma 3.

Lemma 4(a): The expression $A\exp(-iH^{[m]}t)B\psi_0$ with $A\in\mathscr{A}_p(V)$, $B\in\mathscr{A}_q(V)$, $\psi_0\in\delta_m^0(\Delta)\cap\delta_{2m}(\Delta)$ $(p+q=m)$ is differentiable with respect to t and the derivative is the expected one.

Proof: We have $\psi_1=B\psi_0\in\delta_p(\Delta)\cap\delta_{m+p}(\Delta)$. Now proceed as in Lemma 4 (with ψ_0 replaced by ψ_1).

Lemma 5: The function $g_m(t)$, defined in (3.24), is continuous for $\psi_0\in\delta_{2m}(\Delta)$.

Proof: Apparently it is sufficient to prove that $A\exp(-iH^{[m]}t)B\psi_0$ is continuous for $A\in\mathscr{A}_p(V)$, $B\in\mathscr{B}_q(V)$ with $p+q=2m$. Since under our assumption $\psi_1=B\psi_0$ belongs to $\delta_p(\Delta)$ it is sufficient to prove that $v(t):=V_{\alpha_1\alpha_2\cdots\alpha_p}e^{-iH^{[m]}t}\psi_1$ is continuous for $\psi_1\in\delta_p(\Delta)$. However, this follows from Lemma 3 in view of the formula

$$\|v(t+h)-v(t)\|^2=\sum_{\eta\in\Omega^n}|\langle u_\eta,\psi_1\rangle|^2\theta_{\alpha_1\alpha_2\cdots\alpha_p}(\eta)\times|\exp[-iH^{[m]}(\eta,\epsilon)h]-1|^2.$$

Here we have used the fact that $H^{[m]}$ is a function of N, ϵ only.

Proof of the Main Theorem from the lemmas: Each term of $\phi_m(t)$ is of the type described in Lemma 4(a) with $p+q\leqslant m$. Hence the argument sketched at the end of Sec. III is valid, i.e., ϕ_m satisfies the Schrödinger equation modulo an error term $g_m(t)$, which according to Lemma 5 is continuous. Since the operator $H^0+\epsilon V$ is self-adjoint on its maximal domain it is symmetric on $\delta_m^0(\Delta)\cap\delta_{2m}(\Delta)$ so that the hypotheses of our Main Lemma are satisfied. Its conclusion implies the result of our Main Theorem.

Corollary: The domain $\delta_m^0(\Delta)\cap\delta_{2m}(\Delta)$ for which the conclusion of our Main Theorem is valid can be simplified to read as

$$\mathscr{D}^0_m(\Delta)\cap\delta_{2m}(\Delta), \tag{4.26}$$

where

$$\mathscr{D}^0_m(\Delta)=\bigcap_{\alpha_1\alpha_2,\ldots,\alpha_m\in\Delta}\mathscr{D}^0(\alpha_1\alpha_2\cdot\cdot\cdot\alpha_m).$$

Proof: Clearly, the domain (4.26) contains the domain $\delta^0_m(\Delta)\cap\delta_{2m}(\Delta)$; we only have to show that (4.26) is also contained in it. However, obviously

$$\delta_{2m}(\Delta)\subset\delta(\alpha_1\alpha_2\cdots\alpha_m),$$

for all $\alpha_1,\alpha_2,\ldots,\alpha_m\in\Delta$. Hence

$$\delta_{2m}(\Delta)\cap\mathscr{D}^0(\alpha_1\alpha_2\cdots\alpha_m)\subset\delta^0(\alpha_1\alpha_2\cdots\alpha_m).$$

Forming the intersection over all $\alpha_1,\alpha_2,\ldots,\alpha_m\in\Delta$ yields

$$\delta_{2m}(\Delta)\cap\mathscr{D}^0_m(\Delta)\subset\delta^0_m(\Delta).$$

The conclusion follows. This simplification of the initial domain is achieved without sacrificing precision. Still, the description of the initial domain given thus far may appear too unwieldly to be of practical use. For this reason we now proceed to construct a subdomain of the initial domain (4.26) which bears a simple relationship to the potential $V=\Sigma_{\alpha\in\Delta}V_\alpha$. For this purpose we note that there exists a positive constant $C(\Delta)$ and a positive integer $r(\Delta)$ so that, for all $\alpha\in\Delta$,

$$0\leqslant\theta_\alpha(\eta)=Z_\alpha(\eta)|P_\alpha(\eta)|^2\leqslant C(\Delta)(|\eta|+1)^{r(\Delta)}, \tag{4.27}$$

where $|\eta|=\Sigma_{i=1}\eta_i$. Here the integer

$$r(\Delta):=\max_{\alpha\in\Delta}[2\deg P_\alpha(\eta)+|\alpha|]$$

is nothing but the degree of V viewed as a polynomial in the creation and annihilation operators. Going through the construction of the initial domain (4.26) with the estimate (4.27) in mind we see that under the assumption $mr(\Delta)\geqslant 2$ our initial domain contains the simple subdomain $\mathscr{B}_{2m}(\Delta)$, where, for integral p,

$$\mathscr{B}_p(\Delta):=\left\{f\in\mathscr{F}_n:\ \sum_{\eta\in\Omega^n}|\langle f,u_\eta\rangle|^2(|\eta|+1)^{pr(\Delta)}<\infty\right\}. \tag{4.28}$$

Remark: In the case $n=1$, $|\eta|+1$ may be replaced by η.

It turns out that the domain (4.28) is sufficiently small to be a candidate for the initial domain even in case of some resonances. In order to illustrate this point and also in order to discuss a case which in a certain sense is quite opposite to the nonresonant case, we take up the case of the 1:1 resonance, which one of the present authors has studied previously within the context of classical mechanics.[5] In quantum mechanics this system is described by a Hamiltonian

$$H^0=N_1+N_2+1$$

and a potential interaction in standard form

$$V=\sum_{(k,l)\in\Delta}b_1^k b_2^l P_{k,l}(N_1,N_2), \tag{4.29}$$

where Δ is a finite reflection invariant subset of $\mathbb{Z}^2$ and

$$P_{k,l}(N_1,N_2)=\overline{P_{-k,-l}(N_1+k,N_2+l)}$$

are polynomials in the number operators $N_1=c_1a_1$, $N_2=c_2a_2$. Notice that since Y_0 is the submodule (over $\mathbb{Z}$) of $\mathbb{Z}^2$ generated by $(1,-1)$ any potential (4.29) with $\Delta\neq\{0\}$ is resonant with respect to H^0. As in the classical context the normal form is a polynomial in the Hopf "variables" which, however, in the quantum mechanical context are bilinear forms in the creation and annihilation operators, namely (see Refs. 5, 6, and 16),

$$\mathbf{M}=(\tfrac{1}{2}(c_1a_2+c_2a_1),(1/2i)(c_1a_2-c_2a_1),\tfrac{1}{2}(N_1-N_2)),$$
$$J=\tfrac{1}{2}(N_1+N_2)\quad[\mathbf{M}^2=J(J+1)],$$

which generate the U(2) action over $\mathscr{F}_2$. Accordingly, the Hilbert space $\mathscr{F}_2$, under the action of the evolution operator, associated with the normal form, breaks up the same way as under the action of SU(2), namely, as follows:

$$\mathscr{F}_2=\sum_{j\geq 0}\oplus\mathscr{P}_{2j}.$$

Here $\mathscr{P}_{2j}$ is the space of complex homogeneous polynomials of degree $2j$ in two complex variables z_1, z_2. Notice that the domain (4.28) with $p=2m$ in the present case becomes

$$\mathscr{B}_{2m}(\Delta)=\left\{f\in\mathscr{F}_n:\ \sum_{j=0}^{\infty}\sum_{\mu=-j}^{+j}{}'|\langle f,e^j_\mu\rangle|^2(2j+1)^{2mr(\Delta)}<\infty\right\}, \tag{4.30}$$

where $e^j_\mu=\mu_{j+\mu}j_{-\mu}$. Notice that this domain, and, more generally, each domain of this type with $2m$ replaced by any positive integer p, is invariant under the evolution operator of the normal form. For this reason our arguments carry over from the nonresonant case to the 1:1 resonance and show that (4.30) is a valid initial domain for the mth approximant in the 1:1 resonance.

V. A SECOND METHOD OF APPROXIMATION

Finally, we will describe an alternative method of approximation to the solution of the Schrödinger initial value problem

$$i\dot{\phi}=(H_0+\epsilon V)\phi,\quad\phi(0)=\psi_0. \tag{5.1}$$

This method is also based on the normal form technique. However, instead of writing the normalizing transformation in the form of a product of exponentials, as has been done previously, we simply write it in the form of a power series $S(\epsilon)$ (compare with Ref. 11). The method is best described in the setting of an abstract associative algebra $\mathscr{A}$ an its formal power series extension $\mathscr{A}^\infty$. Assuming that $H_0,V\in\mathscr{A}$ we look for a transformation $S\in\mathscr{A}^\infty$ so that

$$(H^0+\epsilon V)S=SH,\quad S^{(0)}=\mathbb{I}, \tag{5.2}$$

and $H=H^0+H^{(1)}$ commutes with H^0. A sufficient condition to make this construction possible is precisely the splitting assumption made in Sec. II, except that $\mathscr{A}^L$ takes over the role of $\mathscr{L}$. Again this splitting assumption extends to the algebra of a formal power series and the components of an element in the kernel $\mathscr{N}$ and image $\mathscr{M}$ of ad H_0 will again be denoted by a caret and a tilde, respectively. Comparing

terms of order ϵ^m on both sides of Eq. (5.2) yields

$$-\operatorname{ad} H^0 S^{(m)} + H^{(m)} = V_m, \tag{5.3}$$

where

$$V_m = -\sum_{l=1}^{m-1} S^{(l)} H^{(m-l)} + V S^{(m-1)}. \tag{5.4}$$

In particular, $V_1 = V$, $S^{(1)}$ is a solution of

$$\operatorname{ad} H^{(0)} S^{(1)} = -\tilde{V}, \tag{5.5}$$

and $H^{(1)} = \hat{V}$. Also,

$$V_2 = -S^{(1)} H^{(1)} + V S^{(1)} = [\hat{V}, S^{(1)}] + \tilde{V} S^{(1)} \tag{5.6}$$

and $S^{(2)}$ is a solution of

$$\operatorname{ad} H^{(0)} S^{(2)} = -\tilde{V}_2$$

and $H^{(2)} = \hat{V}_2$. If we normalize $S^{(1)}$ by the condition $\hat{S}^{(1)} = 0$ then $[\hat{V}, S^{(1)}]\hat{} = 0$ and the expression for $H^{(2)}$ simplifies to

$$H^{(2)} = (\tilde{V} S^{(1)})\hat{}.$$

Although $H^{(2)}$ is still formally self-adjoint if V and $H^{(0)}$ have this property the normalization $\hat{S} = 0$ in general is not the correct one to guarantee the formal self-adjointness of H for real ϵ. This will be achieved as follows.

We normalize S in such a manner that $\hat{S}$ is formally self-adjoint and for $m \geqslant 1$ satisfies

$$\hat{S}^{(m)} = -\frac{1}{2} \sum_{l=1}^{m-1} [S^{(l)} (S^{(m-l)})^\dagger]\hat{}.$$

In order to see that this normalization achieves the desired result we first recognize that (5.2) implies the relation

$$\operatorname{ad} H^0 S S^\dagger = S(H - H^\dagger) S^\dagger + \epsilon [S S^\dagger, V].$$

Clearly, $(S S^\dagger)^{(0)} = \mathbb{I}$ and $H^0 = (H^\dagger)^{(0)}$. Now assume $(S S^\dagger)^{[m-1]} = \mathbb{I}$ and $H^{[m-1]} = (H^\dagger)^{[m-1]}$. Then it follows that

$$\operatorname{ad} H^0 (S S^\dagger)^{(m)} = H^{(m)} - (H^{(m)})^\dagger.$$

Since the rhs belongs to $\mathcal{N}$, whereas the lhs belongs to $\mathcal{M}$ we see that $(S S^\dagger)^{(m)} \in \mathcal{N}$ and $H^{[m]} = (H^\dagger)^{[m]}$. Hence

$$\begin{aligned}(S S^\dagger)^{(m)} &= [(S S^\dagger)^{(m)}]\hat{} \\ &= 2\hat{S}^{(m)} + \sum_{l=1}^{m-1} [S^{(l)} (S^{(m-l)})^\dagger]\hat{} = 0\end{aligned}$$

and therefore $(S S^\dagger)^{[m]} = \mathbb{I}$. In particular, $\hat{S}^{(1)} = 0$, $\hat{S}^{(2)} = -\frac{1}{2}[S^{(1)} S^{(1)\dagger}]\hat{}$, etc.

We define the mth approximant ϕ_m by means of the formula

$$\phi_m(t) = S^{[m]} \exp(-i H^{[m]} t) T^{[m]} \psi_0, \quad T = S^\dagger. \tag{5.7}$$

On account of (3.5) we have

$$\phi_m(0) = \psi_0 + \epsilon^{m+1} \rho_m(S, T, \epsilon) \psi_0,$$

i.e., ϕ_m satisfies the desired initial condition only up to a factor of order ϵ^{m+1}. Again invoking (3.5) we find

$$\begin{aligned}&S^{[m]} H^{[m]} - H^0 S^{[m]} - \epsilon V S^{[m]} \\ &\quad = \epsilon^{m+1} [\rho_m(S, H, \epsilon) - V S^{(m)}].\end{aligned}$$

The terms in brackets on the rhs will be abbreviated by $g_m(t, \epsilon)$. We use this relation in order to show that $\phi_m(t)$ again satisfies the Schrödinger equation modulo a term of order ϵ^{m+1}. Indeed, formal differentiation yields

$$\begin{aligned}i\dot{\phi}_m &= S^{[m]} H^{[m]} \exp(-i H^{[m]} t) T^{[m]} \psi_0 \\ &= (H^0 + \epsilon V) \phi_m + \epsilon^{m+1} g_m(t, \epsilon).\end{aligned}$$

It can be shown[12] by a similar analysis, as presented in Sec. IV, that in a rigorous Hilbert space setting the appropriate initial domains for our alternative method take the form

$$\delta_{3m}(\Delta) \cap \mathscr{D}^0_{2m}(\Delta),$$

with the simple subdomain $\mathscr{B}_{3m}(\Delta)$ [see (4.28)]. Moreover, for ψ_0 restricted to this domain, we obtain an estimate of the form

$$\|\phi_m(t) - \psi(t)\| \leqslant \epsilon^{m+1} M(1 + |t|).$$

Notice that this method, although formally simpler than the method described in Sec. III, leads to smaller initial domains and an error term that does not vanish for $t = 0$.

Of course, the cause for this behavior of the alternative method is the presence of the operator $T^{[m]}$ in the definition of the mth approximant (5.7). Its function is to guarantee that ϕ_m does not only satisfy the Schrödinger equation, but also the initial condition modulo a term of order ϵ^{m+1}.

ACKNOWLEDGMENTS

The authors gratefully acknowledge discussions on the algebraic aspects of the normal form with Alberto Baider; no doubt, they have strongly influenced the writing of Sec. III. Thanks also go to Sheila A-Lee for her superb typing of the manuscript.

APPENDIX: PROOF OF THE MAIN LEMMA

Set $u = \phi - \psi$. Then we compute

$$\frac{d}{dt} \langle u, u \rangle = -2 \operatorname{Im} \langle Au + g, u \rangle = -2 \operatorname{Im} \langle g, u \rangle.$$

Here we used the symmetry of the operator A.

Setting $v_\delta(t) = [\langle u, u \rangle + \delta]^{1/2}$ for $\delta > 0$ and using the Schwarz inequality, we find

$$-\|g(t)\| \leqslant \dot{v}_\delta(t) \leqslant \|g(t)\|.$$

Integrating, we obtain

$$v_\delta(t) \leqslant v_\delta(0) + \left| \int_0^t \|g(s)\| ds \right|.$$

The conclusion of the lemma follows in the limit $\delta \downarrow 0$.

[1] G. D. Birkhoff, *Dynamical Systems* (American Mathematical Society, New York, 1927), Chap. 3.
[2] C. L. Siegel and J. K. Moser, *Lectures on Celestial Mechanics* (Springer, New York, 1971), pp. 209–220.
[3] J. K. Moser, "Lectures on Hamiltonian mechanics," Mem. Am. Math. Soc. **81**, 1 (1968).
[4] R. Cushman, R. Deprit, and R. Mosak, "Normal form and representation theory," J. Math. Phys. **24**, 2103 (1983).
[5] M. Kummer, "On resonant non-linearly coupled oscillators with two equal frequencies," Commun. Math. Phys. **48**, 53 (1976).
[6] R. Churchill, M. Kummer, and D. Rod, "On averaging, reduction and symmetry in Hamiltonian systems," J. Differ. Eqs. **49**, 359 (1983).
[7] J.-C. van der Meer, "The Hamilton Hopf bifurcation," in *Lecture Notes in Mathematics*, Vol. 1160 (Springer, Berlin, 1985).
[8] J. A. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical*

Systems, Applied Mathematical Sciences, Vol. 59 (Springer, New York, 1985), pp. 162–179.
[9]A. B. Lemlih and J. A. Ellison, "The method of averaging and the quantum anharmonic oscillator," Phys. Rev. Lett. **55**, 1950 (1985).
[10](a) M. Kummer, "How to avoid 'secular' terms in classical and quantum mechanics," Nuovo Cimento B **1**, 123 (1971); (b) M. K. Ali, "The quantum normal form and its equivalents," J. Math. Phys. **26**, 2565 (1985).
[11]N. D. Antoniou, "On the convergence of perturbation theories that avoid secular terms in quantum mechanics," Ph.D. thesis, University of Toledo, 1972.
[12]R. Gompa, "Approximations to the quantum mechanical time evolution," Ph.D. thesis, University of Toledo, 1987.
[13]A. Baider and R. Churchill, "Uniqueness and non-uniqueness of normal forms for vector fields," preprint, Dept. of Math. Sciences, Hunter College, New York, 1987.
[14]V. Bargmann, "On a Hilbert space of analytic functions and an associated integral transform," Part I, Comm. Pure Appl. Math. **14**, 187 (1961); Part II, Comm. Pure Appl. Math. **20**, 1 (1967).
[15]V. Bargmann, "Remarks on a Hilbert space of analytic functions," Proc. Natl. Acad. Sci. USA **48**, 199 (1962).
[16]V. Bargmann, "On the representation of the rotation group," Rev. Mod. Phys. **34**, 829 (1962).
[17]D. Babbit, "Certain Hilbert spaces of analytic functions associated with the Heisenberg group," *Studies in Mathematical Physics, Princeton Series in Physics*, edited by E. H. Lieb, B. Simon, and A. S. Wightman (Princeton U.P., Princeton, NJ, 1976).
[18]R. Cushman and D. Rod, "Reduction of the semisimple 1:1-resonance," Physica D **6**, 105 (1982).
[19]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976), 2nd ed.

Time-dependent rotated Hartree: Formal development

H.-D. Meyer, J. Kučar,[a] and L. S. Cederbaum
Theoretische Chemie, Physikalisch-Chemisches Institut, Im Neuenheimer Feld 253, D-6900 Heidelberg, West Germany

(Received 25 September 1987; accepted for publication 17 February 1988)

A new approximation for solving the time-dependent Schrödinger equation is proposed. It improves the time-dependent Hartree approximation by including time-dependent unitary operators acting on the Hartree product. This allows for the approximate description of the correlation between the various degrees of freedom. The mathematical structure of the new approach is analyzed and an alternative Lie algebraic derivation is presented. By adopting two different time-dependent variational principles two different sets of equations are obtained. Differences between the two resulting methods are discussed.

I. INTRODUCTION

Time-dependent methods in solving quantum dynamics have become increasingly popular during the last decade. Many severe problems of the time-independent formulation do not exist in a time-dependent picture, e.g., the treatment of continua and of the rearrangement and breakup processes in reactive scattering. Averaged quantities may be computed directly since one does not necessarily project the wave function onto final states. Hence time-dependent methods are particularly well suited for describing nonfully resolved experiments. Besides of the technical advantages, the time-dependent formulation often leads to a better understanding of the physical mechanism under discussion.

The applications we have in mind are (a) collisional vibrational excitation of a polyatomic molecule, e.g., $HCN(v) + He \rightarrow HCN(v') + He$, where v and v' denote the quantum numbers of vibration prior and post collision, respectively; (b) photodissociation, e.g., $H_2O(v) + h\nu \rightarrow H + OH(v')$; and (c) reactive scattering, e.g., $HCl + D \rightarrow HD + Cl$. The nuclear dynamics performs in all these cases on a single Born–Oppenheimer surface, i.e., the electronic motion decouples from the heavy particle motion (to a very good approximation). The former motion is assumed to be solved by quantum chemistry and the latter one is what we want to investigate.

Adopting a time-dependent formulation one has to choose among different numerical methods for solving the time-dependent Schrödinger equation. The best methods are, of course, the numerically exact ones.[1–9] However, the use of a numerically exact propagation scheme is limited to problems with a very small number of degrees of freedom (two or three). A considerable reduction of the computational effort is gained by resorting to approximate methods, e.g., to the mean field approaches. These are in particular the time-dependent Hartree (TDH) and time-dependent Hartree–Fock (TDHF) methods for treating distinguishable particles and fermions, respectively. In nuclear physics TDHF is one of the most popular time-dependent methods.[10] In the field of molecular physics TDH has been used less. However, semiclassical approximations to TDH have recently been investigated and applied.[11–13]

Another approach to solve the time-dependent Schrödinger equation approximately is the parametrized wave function method. In this method one adopts an ansatz for the wave function that contains a set of time-dependent parameters. The equations of motion for these parameters are determined by a time-dependent variational principle (VP). A thorough discussion of the mathematical structure of the parametrized wave function method can be found in Kramer and Saraceno.[14] A well-known example out of this category is the Gaussian wave packet propagation of Heller.[15–17]

Somewhat similar in spirit to the parametrized wave function method is the exponential operator approach. One writes the propagated wave function $\psi(t)$ as

$$\psi(t) = U(t)\psi(0). \tag{1.1}$$

The time-evolution operator $U(t)$ is written as an exponential or as a product of exponential operators. The exponential form guarantees unitarity although the propagation may be approximative. The first use of exponential time-evolution operators goes back to Magnus.[18] Using the Magnus formula one writes

$$U(t) = \exp\left(i \sum_{n=0}^{\infty} M_n\right), \tag{1.2}$$

where the Hermitian operators M_n are given as time integrals over i^n times the n-fold commutator of $H(t_1),\ldots,H(t_n)$. Depending on the problem there may be only a finite number of nonvanishing operators M_n. Otherwise the series must be truncated. However, if the Hamiltonian $H(t)$ belongs to a finite-dimensional Lie algebra then all M_n also belong to this Lie algebra and the infinite sum may analytically be reduced to a finite one. In this case the time-evolution operator becomes

$$U(t) = \exp\left(i \sum_{n=1}^{m} \alpha_n A_n\right), \tag{1.3}$$

where the operators $\{A_n\}$ form a basis of the m-dimensional Lie algebra. The parameters α_n may be determined by solving a set of differential equations. In a nice paper Pechukas and Light[19] have demonstrated the usefulness of the Magnus approach for solving certain model problems.

The Magnus formulation has two serious drawbacks: (i) the matrix elements of the time-evolution operator are often very hard to compute; and (ii) it can only be shown

[a] Permanent address: R. Boskovic Institute, Zagreb, Yugoslavia.

 0022-2488/88/061417-14$02.50

that Eqs. (1.1), (1.2), or (1.3) establish a local solution to the time-dependent Schrödinger equation (i.e., up to some finite t_0). A global solution may not exist! These two difficulties are overcome in the formulation of Wei and Norman.[20,21] They write the time-evolution operator as a product of exponentials

$$U(t) = \exp(i\alpha_m A_m)\cdots\exp(i\alpha_2 A_2)\exp(i\alpha_1 A_1), \quad (1.4)$$

where $\alpha_1,...,\alpha_m$ are real time-dependent parameters and where the Hermitian operators $A_1,...,A_m$ are assumed to be a basis of some Lie algebra. (Actually Wei and Norman did not require $A_1,...,A_m$ to be Hermitian and $\alpha_1,...,\alpha_m$ to be real and they considered a larger class of problems than just the time-dependent Schrödinger equation.) If the Hamiltonian $H(t)$ is in the Lie algebra then one can find parameters $\alpha_1,...,\alpha_m$ yielding an *exact* solution. Wei and Norman derived differential equations for the parameters and showed under which condition the solution is *global*. These conditions, in particular, concern the ordering of the product (1.4).

The Wei and Norman approach is not of too much practical use in the general case. This is because one generally cannot find a Lie algebra which on one hand contains the Hamiltonian and which on the other hand has a small sized basis of simple operators such that $\exp(i\alpha_j A_j)$ can be evaluated. If one approximates the Hamiltonian by a member of the Lie algebra, then this approximation is often very poor.

In this paper we shall combine the TDH and the Wei and Norman approaches yielding the time-dependent rotated Hartree (TDRH) formalism. By "rotated" we here denote a generalized rotation, i.e., a unitary transformation of the wave function; we do not restrict ourselves to physical rotations of the coordinates. In TDRH the unitary operators act on the Hartree product rather than on $\psi(t=0)$. Hence the unitary transformations have only to account for the correlations between the different degrees of freedom and not for the full motion of the wave packet. Since TDRH is derived from a VP it is necessarily an improvement over TDH.

The TDRH was recently applied successfully to a model problem.[22] Here we devote ourselves to the formal derivation of the new method. In Sec. II we shall derive the equations of motion from time-dependent VP's. In Sec. III we discuss the TDRH approach in terms of the Lie algebra formalism and introduce the new concept of a Δ algebra. The algebraic approach leads to the same equations of motion as derived in Sec. II. The Lie algebraic formulation, however, is of great importance because it illuminates the mathematical structure of the method. It shows a way how to determine the optimal ordering of the exponential operators and allows us to name the conditions under which a global solution exists. In Sec. IV we shall discuss the two-dimensional harmonic algebra as an example and in Sec. V we finally summarize our results.

II. THE GENERALIZED HARTREE METHOD

A. The ansatz

The goal of the present work is to solve the time-dependent Schrödinger equation

$$i\dot{\psi}_{ex} = H\psi_{ex} \quad (2.1)$$

in an approximate but systematic fashion. Here H denotes the Hamiltonian that may or may not depend on time and ψ_{ex} is the exact wave function depending on n coordinates. We shall assume $\hbar = 1$ throughout. In the TDH formalism the wave function is approximated by a separable product

$$\psi_{ex}(x_1,...,x_n,t) \approx \Phi(x_1,x_2...,x_n,t) = \prod_{j=1}^{n} \Phi_j(x_j,t). \quad (2.2)$$

The single-particle functions Φ_j are assumed to be normalized,

$$\langle \Phi_j | \Phi_j \rangle = 1, \quad (2.3)$$

and their time evolution is given by the mean field equations[23]

$$i\dot{\Phi}_j = (H^{(j)} - [(n-1)/n]\langle H \rangle)\Phi_j, \quad (2.4)$$

where the total energy reads

$$\langle H \rangle = \langle \Phi | H | \Phi \rangle \quad (2.5)$$

and where

$$H^{(j)} = \langle \Phi_1,...,\Phi_{j-1},\Phi_{j+1},...,\Phi_n | H | \Phi_1,...,\Phi_{j-1},\Phi_{j+1},...,\Phi_n \rangle \quad (2.6)$$

is the mean field Hamiltonian describing the motion of the particle j in the mean field spanned by the other particles. The phase factor $[(n-1)/n]\langle H \rangle$ is often removed from the mean field equation (2.4) and the sum of all phase factors is treated separately.[23]

The Hartree ansatz can be improved by enlarging the space of the trial functions. Our ansatz for the trial wave function ψ reads[22]

$$\psi(t) = U(t)\Phi(t), \quad (2.7)$$

where $U(t)$ is a unitary operator. This operator is written as a product

$$U = \prod_{k=1}^{m} U_k \equiv U_m \cdots U_2 U_1 \quad (2.8)$$

with

$$U_k(t) = \exp(i\alpha_k(t)A_k). \quad (2.9)$$

(Note that we order products from right to left!) The α_k denote time-dependent real parameters and the A_k are Hermitian operators.

The unitary transformations performed by the exponential operators can be considered as generalized rotations. Hence we shall refer to the method which follows from the ansatz (2.7)–(2.9) as the time-dependent Hartree method with generalized rotations or briefly time-dependent rotated Hartree (TDRH).

Some remarks on TDHF and its generalization to TDRHF are in order. In TDHF the trial function Φ is not of product form but is a Slater determinant. When generalizing TDHF to TDRHF one has to be careful not to destroy the antisymmetry. The natural way to achieve this is to restrict the generators of the rotations A_k to be second quantized operators, i.e., Hermitian combinations of fermion annihilation and creation operators.

B. Variational principles

The equations of motion for the parameters α_k as well as for the single-particle functions Φ_j are obtained by applying a time-dependent VP to the ansatz (2.7)–(2.9). In the following we shall discuss two different VP's. The first VP we discuss is due to McLachlan.[23] It states

$$\delta\|i\dot{\psi} - H\psi\|^2 = 0, \tag{2.10}$$

where $\|\cdot\|$ denotes the Hilbert space norm. Only the time derivative $\dot{\psi}$ is varied. This demand follows from the physical picture that, at each point of time, ψ is given as the initial value and the optimal approximation to $\dot{\psi}$ is desired. The McLachlan VP defines "optimal" as the smallest error in the norm. If the space of allowed variations of $\dot{\psi}$ is equal to the space of allowed variations of ψ, i.e., if

$$\{\delta\dot{\psi}\} = \{\delta\psi\}, \tag{2.11}$$

then one finds that Eq. (2.10) is equivalent to

$$\mathrm{Im}\left\langle \delta\psi \middle| i\frac{\partial}{\partial t} - H \middle| \psi \right\rangle = 0. \tag{2.12}$$

Our ansatz (2.7)–(2.9) implies that (2.11) and thus (2.12) are valid.

The other VP of interest is thoroughly investigated by Kramer and Saraceno.[14] It starts by introducing the Lagrangian

$$L = \left\langle \psi \middle| i\frac{\partial}{\partial t} - H \middle| \psi \right\rangle, \tag{2.13}$$

where ψ is assumed to be normalized. As usual one requires

$$\delta \int_{t_1}^{t_2} L\, dt = 0 \tag{2.14}$$

subject to the boundary conditions $\delta L(t_1) = \delta L(t_2) = 0$. Performing the partial integration leads to

$$\mathrm{Re}\left\langle \delta\psi \middle| i\frac{\partial}{\partial t} - H \middle| \psi \right\rangle = 0. \tag{2.15}$$

The two VP's lead to surprisingly similar equations.

If for each allowed variation $\delta\psi$ the variation $i\delta\psi$ is also allowed then both VP's reduce to

$$\left\langle \delta\psi \middle| i\frac{\partial}{\partial t} - H \middle| \psi \right\rangle = 0, \tag{2.16}$$

which is known as the Dirac–Frenkel VP.[24,25] The variational parameters α_k, however, are restricted to be real and hence the two VP's are inequivalent and produce different equations of motion. The VP due to McLachlan will be denoted as NVP (norm VP) and the second VP by LVP (Lagrangian VP). The NVP and LVP fulfill the imaginary and real parts of the Dirac–Frenkel VP (2.16), respectively. The Dirac–Frenkel VP itself is not applicable since it requires complex parameters α_k. The parameters α_k have been chosen to be real because otherwise the operators U_k are no longer unitary. The use of nonunitary operators leads to untraceable difficulties. Returning to the usual Hartree approach (2.4) we remark that in this case all three VP's (NVP, LVP, and Dirac–Frenkel) are equivalent and hence yield the same mean field equations. This is because the variational quantities Φ_j are complex.

C. Equations of motion

As for the TDH approach both VP's yield in our TDRH the same mean field equations of motion for the single-particle functions Φ_j. They read[22]

$$i\dot{\Phi}_j = \left\{ \tilde{H}^{(j)} - \frac{n-1}{n}\langle \tilde{H} \rangle + \sum_{k=1}^{m} \dot{\alpha}_k \left(\hat{A}_k^{(j)} - \frac{n-1}{n}\langle \hat{A}_k \rangle \right) \right\} \Phi_j, \tag{2.17}$$

where $\tilde{H}$ denotes the rotated Hamiltonian

$$\tilde{H} = U^\dagger H U, \tag{2.18}$$

and where $\hat{A}_k$ denotes the partially rotated generator

$$\hat{A}_k = \left[\prod_{l=1}^{k-1} \exp(i\alpha_l A_l) \right]^\dagger A_k \left[\prod_{l=1}^{k-1} \exp(i\alpha_l A_l) \right]. \tag{2.19}$$

The expectation value $\langle X \rangle$ and the mean field $X^{(j)}$ of the operator X are defined analogously to Eqs. (2.5) and (2.6), respectively. One recognizes that the first part of Eq. (2.17) is just the TDH formula except for the use of the rotated Hamiltonian. The second part arises because the mean field equation must account for the change of the wave function introduced by the change of the parameters α_k. This part is proportional to the velocities $\dot{\alpha}_k$.

Before we discuss the equations of motion for the parameters α_k we find it convenient to introduce the super operator Δ which is defined by

$$\Delta B = B + (n-1)\langle B \rangle - \sum_{j=1}^{n} B^{(j)}. \tag{2.20}$$

The super operator is idempotent, i.e.,

$$\Delta(\Delta B) = \Delta B, \tag{2.21}$$

and it produces operators with vanishing mean fields and vanishing expectation values

$$(\Delta B)^{(j)} = 0, \quad \langle \Delta B \rangle = 0. \tag{2.22}$$

The meaning of ΔB is simple. It describes that part of B which is nonseparable. The equation $\Delta B = 0$ holds if and only if B is a sum of single-particle operators.

The super operator Δ enables us to separate the set of operators into two classes. We shall call an operator B a *mixing* or a *separable* operator according to whether $\Delta B \neq 0$ or $\Delta B = 0$, respectively. Although the value of ΔB obviously depends on the choice of the Hartree product with respect to which the mean fields are defined, the distinction mixing/separable is independent of this choice! Finally we note two very useful rules, namely

$$\langle (\Delta A) B \rangle = \langle A(\Delta B) \rangle = \langle \Delta A \Delta B \rangle \tag{2.23}$$

and

$$\Delta(B_j B_l) = (B_j - \langle B_j \rangle)(B_l - \langle B_l \rangle), \tag{2.24}$$

where B_j and B_l denote single-particle operators operating on the jth and lth coordinate, respectively.

We now vary—according to the rules of NVP and LVP—the parameters α_k to obtain their equations of motion. This leads to[22]

$$\langle \Phi | [F, \Delta\hat{A}_k]_\pm | \Phi \rangle = 0, \tag{2.25}$$

where the auxiliary operator F is defined as

$$F = \Delta\tilde{H} + \sum_{k=1}^{m} \dot{\alpha}_k \Delta\hat{A}_k. \tag{2.26}$$

In Eq. (2.25) one has to use the anticommutator $[A,B]_+ = AB + BA$ if one adopts NVP and the commutator $[A,B]_- = AB - BA$ if LVP is used. Equation (2.25) can easily be rewritten by assuming a matrix notation

$$\mathsf{B}\dot{\boldsymbol{\alpha}} = \mathbf{b}, \tag{2.27a}$$

$$\mathsf{C}\dot{\boldsymbol{\alpha}} = \mathbf{c}, \tag{2.27b}$$

where $\dot{\boldsymbol{\alpha}} = (\dot{\alpha}_1, \dot{\alpha}_2, \ldots, \dot{\alpha}_m)^T$. The matrices B and C possess the elements

$$B_{kk'} = \langle [\Delta\hat{A}_k, \Delta\hat{A}_{k'}]_+ \rangle , \tag{2.28a}$$

$$C_{kk'} = i\langle [\Delta\hat{A}_k, \Delta\hat{A}_{k'}]_- \rangle , \tag{2.28b}$$

and the elements of the vectors $\mathbf{b}$ and $\mathbf{c}$ take the following appearance:

$$b_k = -\langle [\Delta\hat{A}_k, \Delta\tilde{H}]_+ \rangle , \tag{2.29a}$$

$$c_k = -i\langle [\Delta\hat{A}_k, \Delta\tilde{H}]_- \rangle . \tag{2.29b}$$

Equations (2.27a), (2.28a), (2.29a) refer to NVP and Eqs. (2.27b), (2.28b), (2.29b) to LVP. All quantities appearing in these equations are real. The matrices B and C are symmetric and antisymmetric, respectively; i.e.,

$$\mathsf{B}^T = \mathsf{B}, \tag{2.30a}$$

$$\mathsf{C}^T = -\mathsf{C}. \tag{2.30b}$$

D. Comparison of the variational principles

The most striking difference between the two VP's lies in their treatment of constants of motion. In particular, the LVP conserves the total energy of time-independent systems whereas NVP does not. Let I denote some time-independent constant of motion, i.e., $\partial I/\partial t = 0$ and $[I,H]_- = 0$. The time derivative of its expectation value is

$$\frac{d}{dt}\langle\psi|I|\psi\rangle = i\langle\Phi|[\Delta\tilde{I},F]_-|\Phi\rangle. \tag{2.31}$$

This result is independent of the VP used. It also holds for TDH for which F and $\Delta\tilde{I}$ simplify to ΔH and ΔI, respectively. Setting $I = H$ and using Eqs. (2.26) and (2.29) one finds

$$\dot{E} \equiv \frac{d}{dt}\langle\psi|H|\psi\rangle = \dot{\boldsymbol{\alpha}}^T\mathbf{c}. \tag{2.32}$$

This expression is valid independently of how $\dot{\boldsymbol{\alpha}}$ is determined. Using the LVP one finds that $\dot{E} = \dot{\boldsymbol{\alpha}}^T\mathsf{C}\dot{\boldsymbol{\alpha}} = 0$ vanishes because of the antisymmetry of the matrix C [cf. (2.27b), (2.30b)]. We now let I become again an arbitrary constant of motion. Using the LVP one then finds that the time derivative of $\langle\psi|I|\psi\rangle$ vanishes if $\Delta\tilde{I}$ is a linear combination of the $\Delta\hat{A}_k$ [cf. (2.25) and (2.31)]. This requirement is fulfilled if I is in the maximal embedding Lie algebra. (The embedding algebra will be introduced in the next section.) The NVP, on the other hand, does neither conserve the total energy nor the expectation value of a constant of motion (except for trivial ones with $\Delta\tilde{I} \equiv 0$; $I = 1$ is usually the only constant of motion with this property).

The constraint of conservation of total energy can easily be added to the NVP. This leads to the equations of motion

$$\dot{\boldsymbol{\alpha}} = (1 - \mathbf{d}\mathbf{c}^T)\mathsf{B}^{-1}\mathbf{b}, \tag{2.33}$$

where

$$\mathbf{d} = \mathsf{B}^{-1}\mathbf{c}/(\mathbf{c}^T\mathsf{B}^{-1}\mathbf{c}). \tag{2.34}$$

We have derived these expressions to show that LVP is *not* just NVP plus the constraint of conservation of energy.

From the above considerations one may get the impression that the LVP is superior over the NVP. However, if one investigates the error introduced into the wave function then the NVP seems to be the superior method. To discuss the error we introduce the effective Hamiltonian H_{eff} which propagates the approximate solution ψ. We define H_{eff} by

$$i\dot{\psi} = H_{\text{eff}}\psi. \tag{2.35}$$

One can easily show that

$$H_{\text{eff}} = H - UFU^{\dagger} \tag{2.36}$$

and hence

$$\|i\dot{\psi} - H\psi\| = \|F\Phi\| \tag{2.37}$$

holds. This shows that $\|F\Phi\|$ is a measure of the error. Equations (2.35)–(2.37) are valid for both VP's as well as for TDH for which F simplifies to ΔH.

The NVP is equivalent to requiring

$$\|F\Phi\|^2 = \text{Min}, \tag{2.38}$$

where the $\dot{\alpha}_k$ are to be varied. Hence the NVP tries to make $\|F\Phi\|$ as small as possible. This is very reasonable because—as seen above—$\|F\Phi\|$ is a measure of the error introduced in the wave function. We note that both VP's yield the same (exact) result, if the $\dot{\alpha}_k$ can be arranged such that $F\Phi = 0$. In general, however, the results obtained by employing the NVP or the LVP will be different. It is not clear to us in which sense the LVP provides an optimal result if an exact solution is not accessible.

We have shown under which conditions the new methods yield an exact solution (i.e., if $\|F\Phi\| = 0$). We have, however, not yet proved that a global solution exists. A global TDRH solution requires a global solution of the differential equation (2.27). Such a solution may not exist, e.g., if the matrices B or C become singular in the course of the integration. In the next section we shall investigate the origin of possible singularities of B or C and discuss how to avoid them.

III. USING LIE ALGEBRAS

A. General remarks

In Sec. III B we shall somewhat colloquially describe the Lie algebraic approach to TDRH. The algebraic approach offers an independent route to the equations of motion of TDRH. This way is more difficult to follow but it provides a considerably deeper understanding of the method. We start the discussion by assuming that the Hamiltonian belongs to a so-called embedding Lie algebra, i.e., by assuming that an exact solution is accessible. In Sec. III C we discuss how to proceed if the Hamiltonian is not a member of the embedding algebra. We shall find that the equations of motion, as derived in Sec. II, are recovered by the algebraic approach for both of the VP's. In Sec. III D we essentially repeat and extend the second part by using more mathematical rigor.

When discussing the algebraic approach we shall concentrate on the equations of motion for the parameters α. With respect to the mean field equations (2.17) we only remark that their solution exists and is unique, provided the velocities $\dot{\alpha}_k$ remain bounded. One then finds that the $\|\dot{\Phi}_j\|$ are bounded which guarantees the existence of the solution. The mean field operators appearing in (2.17) can be differentiated with respect to the single-particle functions Φ_j. Hence the Lipschitz condition can be satisfied yielding the uniqueness of the solution.

B. Outline of results

The basis of a Lie algebra has often been used[14,18,20,21] to serve as generators of the generalized rotations describing the time-evolution operator. Such a choice of generators allows for a rigorous and transparent theory. In the following we briefly discuss the properties of Lie algebra. For our purpose it is sufficient to define a Lie algebra as a complex linear space of linear operators which is closed under commutations, i.e., the commutator of two operators out of the algebra belongs to the algebra. From this property it follows[18] that the Lie algebra is invariant under rotations, i.e., if A and B belong to the algebra then A rotated by B, $\exp(-B)\,A\exp(B)$, also belongs to the Lie algebra. Hence if the Hamiltonian is in the algebra, then the rotated Hamiltonian $\tilde{H}$ is still in the algebra. All operators of the Lie algebra can be expanded in the finite basis and the exact quantal solution of, e.g., the Wei and Norman formulation becomes equivalent to solving a finite set of first-order differential equations.[14,18,20,21]

In the TDRH approach the situation is more complicated. The time evolution operator acts on the Hartree product which itself is a fairly good approximation to the exact wave function. Hence to yield an exact solution (we suppose the Hamiltonian being such that an exact solution is accessible) the unitary operator has only to account for the nonseparable parts of the wave function. Therefore only the mixing parts, $\Delta\tilde{H}$ and $\Delta\hat{A}_k$, of the rotated Hamiltonian $\tilde{H}$ and the partially rotated generators $\hat{A}_k$ appear in the equations of motion for the parameters α_k [cf. (2.27)–(2.29)]. This new situation leads to the introduction of the new concept of a Δ algebra. The Δ algebra $\mathscr{D}$ is defined as a linear space of mixing operators, i.e., $\mathscr{D}$ contains no separable operator except for the zero operator. In order to be a Δ algebra, $\mathscr{D}$ must be invariant under rotations (so called Δ invariance, see Sec. III C). For each Δ algebra there exist embedding Lie algebras. A Lie algebra $\mathscr{L}$ is called an embedding algebra of the Δ algebra $\mathscr{D}$ if $\mathscr{D}\subset\mathscr{L}$ and $\Delta\mathscr{L}=\Delta\mathscr{D}$; i.e., the super operator Δ projects any operator of $\mathscr{L}$ onto $\Delta\mathscr{D}$. ($\mathscr{D}$ and $\Delta\mathscr{D}$ are isomorphic. One merely distinguishes between the two sets because $\Delta\mathscr{D}$ is defined only with respect to the Hartree product.)

Let the generators of TDRH, A_k, be a basis of some m-dimensional Δ algebra $\mathscr{D}$. Because of the Δ invariance one may expand the partially rotated operators $\Delta\hat{A}_k$ by the unrotated ones, i.e.,

$$\Delta\hat{A}_k=\sum_{k'}D_{kk'}\Delta A_{k'}. \tag{3.1}$$

The matrix D is called "structure matrix." It depends on α as well as on the *ordering* of the generators. Now suppose that $\Delta\tilde{H}$ is in the Δ algebra, i.e.,

$$\Delta\tilde{H}=\sum_{k=1}^{m}\tilde{h}_k\Delta A_k. \tag{3.2}$$

By satisfying the linear equation

$$\tilde{\mathbf{h}}=-\mathsf{D}^T\dot{\boldsymbol{\alpha}} \tag{3.3}$$

the velocities $\dot{\alpha}_k$ are arranged such that the operator F [cf. (2.26)] vanishes indentically. Hence TDRH yields an *exact* solution if $\Delta\tilde{H}\in\Delta\mathscr{D}$. As known from the theory of differential equations,[26] Eq. (3.3) has a *global* solution if

$$\begin{aligned}&\text{(i)}\ \ \mathsf{D}\text{ is nonsingular for all }\boldsymbol{\alpha},\\ &\text{(ii)}\ \ \|\mathsf{D}^{-1T}\tilde{\mathbf{h}}\|\leqslant c_1+c_2\|\boldsymbol{\alpha}\|,\end{aligned} \tag{3.4}$$

where c_1 and c_2 are some constants and where $\|\cdot\|$ denotes the Euclidean norm of the $\mathbb{R}^m$. It is one of the major results of the algebraic approach to TDRH to emphasize that one has to avoid a singular structure matrix by choosing an optimal ordering of the generators A_k (see Sec. IV).

The requirement that $\Delta\tilde{H}$ lies in the Δ algebra is usually not easy to check. One has to consider the Δ projection of the *rotated* Hamiltonian $\tilde{H}$ rather than of the Hamiltonian itself (the Δ invariance does not help because $\tilde{H}\notin\mathscr{D}$ although $\Delta\tilde{H}\in\Delta\mathscr{D}$). However, if H belongs to an embedding Lie algebra $\mathscr{L}$ of the Δ algebra $\mathscr{D}$ then we can conclude $H\in\mathscr{L}\rightarrow\tilde{H}\in\mathscr{L}\rightarrow\Delta\tilde{H}\in\Delta\mathscr{D}$. Hence we arrive at the easy to check condition that TDRH yields an *exact* solution if H belongs to an embedding Lie algebra of $\mathscr{D}$.

C. Hamiltonians not belonging to the Lie algebra

In most of the problems which one wants to solve, the Hamiltonian H does not belong to an embedding Lie algebra or, more precisely, $\Delta\tilde{H}\notin\Delta\mathscr{D}$. In this case one constructs an approximation $\Delta\tilde{H}_{\text{app}}$ which belongs to the Δ algebra. The strategy now is to seek an exact solution to the approximate Hamiltonian which in turn establishes an approximate solution to the exact Hamiltonian. [By the way, the effective Hamiltonian for which we obtain the time evolved wave function is given by $H_{\text{eff}}=H-U(\Delta\tilde{H}-\Delta\tilde{H}_{\text{app}})U^{\dagger}$, see Eq. (2.36).] The most obvious construction of $\Delta\tilde{H}_{\text{app}}$ is making the difference between $\Delta\tilde{H}$ and $\Delta\tilde{H}_{\text{app}}$ as small as possible, where the measure is the norm with respect to the Hartree product, i.e.,

$$\|(\Delta\tilde{H}-\Delta\tilde{H}_{\text{app}})\Phi\|^2=\text{Min}. \tag{3.5}$$

We write

$$\Delta\tilde{H}_{\text{app}}=\sum_{k=1}^{m}\tilde{h}_k\Delta A_k, \tag{3.6}$$

where now $\tilde{h}_k$ are to be determined and are not given *a priori* as in Eq. (3.2). The solution to the problem (3.5) yields

$$\mathsf{B}_0\tilde{\mathbf{h}}=-\mathbf{b}_0, \tag{3.7}$$

where

$$b_{0k}=-\langle[\Delta A_k,\Delta\tilde{H}]_+\rangle, \tag{3.8}$$

$$B_{0kk'}=\langle[\Delta A_k,\Delta A_{k'}]_+\rangle. \tag{3.9}$$

The differential equation for the parameters (3.3) now reads

$$\dot{\alpha} = -\mathsf{D}^{-1T}\tilde{\mathbf{h}} = \mathsf{D}^{-1T}\mathsf{B}_0^{-1}\mathbf{b}_0$$
$$= (\mathsf{D}\mathsf{B}_0\mathsf{D}^T)^{-1}\mathsf{D}\mathbf{b}_0 = \mathsf{B}^{-1}\mathbf{b} \qquad (3.10)$$

with obvious definitions for B and **b**. It is easy to see that the quantities B and **b** of Eq. (3.10) are exactly the ones defined by Eqs. (2.28a) and (2.29a). Hence the prescription (3.5) to determine the Δ-algebra approximation to $\Delta\tilde{H}$ is strictly equivalent to the NVP.

Another way of finding a Δ-algebra approximation to $\Delta\tilde{H}$ is by requiring that $\Delta\tilde{H}$ and $\Delta\tilde{H}_{\rm app}$ have the same expectation values of the commutators with the Δ-algebra elements; i.e.,

$$\langle [\Delta\tilde{H},\Delta A_k]_-\rangle = \langle [\Delta\tilde{H}_{\rm app},\Delta A_k]_-\rangle . \qquad (3.11)$$

This condition is motivated by the fact that the structure of a Lie algebra is determined by its commutators. The condition (3.11) has the solution.

$$\mathsf{C}_0\tilde{\mathbf{h}} = -\mathbf{c}_0 \qquad (3.12)$$

with

$$C_{0kk'} = i\langle [\Delta A_k,\Delta A_{k'}]_-\rangle \qquad (3.13)$$

and

$$c_{0k} = -i\langle [\Delta A_k,\Delta\tilde{H}]_-\rangle . \qquad (3.14)$$

Similar to Eq. (3.10) one finds

$$\mathsf{C} = \mathsf{D}\mathsf{C}_0\mathsf{D}^T \qquad (3.15a)$$

and

$$\mathbf{c} = \mathsf{D}\mathbf{c}_0, \qquad (3.15b)$$

where C and **c** are given by Eqs. (2.28b) and (2.29b). This proves that the construction (3.11) is equivalent to the LVP.

We will now discuss the origin of the possible singularities of the matrices B and C. Equations (3.10) and (3.15) show that these matrices are singular if D or if B_0 (C_0) are singular. The matrix D is entirely defined by the structure of the Δ algebra and by the ordering of the generators. It is independent of the Hartree product Φ. The matrix elements of D can be derived explicitly (see Sec. IV). By a proper choice of the Δ algebra and of the ordering of the generators one can ensure that D is nonsingular once and for all.

The singular points of B_0 and C_0, on the other hand, do not depend on the ordering. They are even independent on the choice of the basis of $\mathscr{D}$. They do, however, depend on the Hartree product Φ. A singularity of B_0 indicates that Eq. (3.5) has no *unique* solution. (The existence of a solution is clear.) In fact, it is easy to show that B_0 is singular if and only if there exists a Hermitian operator $A\in\mathscr{D}$, $A\neq 0$, with

$$\Delta A\Phi = 0. \qquad (3.16)$$

A Hartree product Φ satisfying Eq. (3.16) is called a singular point of B_0. At a singular point one may add $\lambda\Delta A$ to $\Delta\tilde{H}_{\rm app}$ with an arbitrary λ and still satisfy Eq. (3.5). From Eq. (3.16), on the other hand, it follows that the time evolution of the wave function ψ is not changed by the above mentioned change of $\Delta\tilde{H}_{\rm app}$. Hence the singularities of B_0 are irrelevant physically. The (spurious) difficulty raised by the singularity of B_0 can be overcome by replacing B_0^{-1} by the pseudoinverse B_0^I.

Before we continue we shall briefly discuss the definition and the properties of a pseudoinverse.[27] Let M be a real matrix which does not need to be square and let P denote the projector on those eigenvectors of $\mathsf{M}^T\mathsf{M}$ which have vanishing eigenvalues. The pseudoinverse M^I is defined as

$$\mathsf{M}^I = (\mathsf{M}^T\mathsf{M} + \lambda\mathsf{P})^{-1}\mathsf{M}^T, \qquad (3.17)$$

where λ denotes some positive constant. (M^I does not depend on λ.) The pseudoinverse has the following properties[27]:

$$\mathsf{M}\mathsf{M}^I\mathsf{M} = \mathsf{M}, \qquad (3.18a)$$
$$\mathsf{M}^I\mathsf{M}\mathsf{M}^I = \mathsf{M}^I, \qquad (3.18b)$$
$$(\mathsf{M}^I)^I = \mathsf{M}, \qquad (3.18c)$$
$$(\mathsf{M}^T)^I = (\mathsf{M}^I)^T, \qquad (3.18d)$$
$$\mathsf{M}^I\mathsf{M} = 1 - \mathsf{P}, \qquad (3.18e)$$
$$\mathsf{M}\mathsf{M}^I = 1 - \mathsf{P}', \qquad (3.18f)$$

where P′ denotes the projector on the null space of $\mathsf{M}\mathsf{M}^T$. (To put it differently: 1 − P and 1 − P′ denote projectors on the row and column spaces of M, respectively.) If M^{-1} exists then $\mathsf{M}^I = \mathsf{M}^{-1}$. If the equation

$$\mathsf{M}\mathbf{x} = \mathbf{m} \qquad (3.19)$$

has a solution (i.e., if $\mathsf{P}'\mathbf{m} = \mathbf{0}$) then

$$\mathbf{x}_0 = \mathsf{M}^I\mathbf{m} \qquad (3.20)$$

is that solution of all possible solutions of Eq. (3.19) which has the smallest Euclidian norm. If, on the other hand, Eq. (3.19) has no solution (i.e., if $\mathsf{P}'\mathbf{m}\neq\mathbf{0}$) then $\mathbf{x}_0$ denotes the solution of minimal norm to the equation

$$\|\mathsf{M}\mathbf{x} - \mathbf{m}\|^2 = \text{Min}. \qquad (3.21)$$

Before we continue, we remark that we shall make further use of the pseudoinverse when we encounter the singular matrices D and C_0 of Eqs. (4.10) and (4.19) (see Sec. IV).

We now return to the equations of motion for the parameters α as determined by NVP. The equation

$$\tilde{\mathbf{h}} = -\mathsf{B}_0^I\mathbf{b}_0 \qquad (3.22)$$

provides a well defined exact solution to Eq. (3.7) [because $\mathsf{P}'\mathbf{b}_0 = \mathbf{0}$, see Eqs. (3.8), (3.9), and (3.16)]. In the present situation, however, it is not convenient to employ the pseudoinverse directly. If the matrix B_0 is regular but close to singular then $\tilde{\mathbf{h}}$ and hence $\dot{\alpha}$ may become very large which causes numerical difficulties as well as obscures the existence of a global solution. The pseudoinverse B_0^I—albeit well defined— cannot be uniformly bounded for all times. Guided by Eq. (3.17) we therefore preferred to replace B_0^{-1} by

$$\mathsf{B}_0^\epsilon = (\mathsf{B}_0^T\mathsf{B}_0 + \epsilon 1)^{-1}\mathsf{B}_0^T, \qquad (3.23)$$

which is a uniformly bounded approximation to the pseudoinverse. Here ϵ denotes a small positive number. The computed wave function will depend on ϵ. The convergence of $\psi(\epsilon,t)$ as $\epsilon\to 0$ is discussed at the end of this section. (Note that $\mathsf{B}_0^\epsilon\to\mathsf{B}^I$ as $\epsilon\to 0$.)

Returning to the LVP we find that the situation is similar but not quite as pleasant as above. The matrix C_0 is singular if and only if there exists a Hermitian operator $A\in\mathscr{D}$, $A\neq 0$ such that

$$\langle\Phi|[\Delta A,\Delta A_k]_-|\Phi\rangle = 0 \qquad (3.24)$$

holds for $k = 1,...,m$. This is a much weaker condition than (3.16) and hence singular points of C_0 are more likely than singular points of B_0. (A singular point of B_0 is necessarily a singular point of C_0.) Moreover, Eq. (3.11) may not have a solution at all (i.e., $P'\mathbf{c}_0 \neq \mathbf{0}$) if C_0 is singular and $H \notin \mathcal{L}_{\max}$. Hence the computed wave function $\psi(t)$ may depend on how one determines $\dot{\alpha}$ near the singularities of C_0. In practice one removes the singularities of C_0 by a procedure analogous to that of Eqs. (3.17) and (3.23). The above considerations again show that the NVP seems to yield better results than the LVP, at least if one considers the error in the wave function.

D. Mathematical details

In this section we shall give a more extensive discussion on Lie and Δ algebras as well as a more rigorous proof on the existence of a global exact solution. The motivation for all the definitions and lemmas to follow was provided above. The proofs of the lemmas are mostly not given here; they are either straightforward or can be found in the literature.[18,20,21,28–31] For sake of simplicity and brevity we shall not adopt the most general definition of a Lie algebra but consider only Lie algebras formed by linear operators. For sake of brevity we also shall drop the minus sign from the commutators in the following.

Definition 3.1: A set of linear operators $\mathcal{L}$ is called *closed* if for each $A,B \in \mathcal{L}$ follows $[A,B] \in \mathcal{L}$.

Definition 3.2: A set of linear operators $\mathcal{L}$ is called *invariant* if for each $A,B \in \mathcal{L}$ follows $e^{-B}Ae^{B} \in \mathcal{L}$.

Definition 3.3: A set of linear operators is called a *Lie algebra* if it is a closed linear space.

Lemma 3.1: A Lie algebra is invariant. An invariant linear space is a Lie algebra.

Definition 3.4: Let Δ be a super operator and $\mathcal{L}$ a set of operators. The sets $\Delta\mathcal{L}$, $[\mathcal{L}_1, \mathcal{L}_2]$, and $\mathrm{Comp}(\mathcal{L})$ (called the completion of $\mathcal{L}$) are defined as

$$\Delta\mathcal{L} = \{\Delta B \,|\, B \in \mathcal{L}\},$$

$$[\mathcal{L}_1,\mathcal{L}_2] = \{[A,B] \,|\, A \in \mathcal{L}_1,\ B \in \mathcal{L}_2\},$$

$$\mathrm{Comp}(\mathcal{L}) = \mathrm{span}(\mathcal{L} \cup [\mathcal{L},\mathcal{L}] \cup [\mathcal{L},[\mathcal{L},\mathcal{L}]] \cup [\mathcal{L},[\mathcal{L},[\mathcal{L},\mathcal{L}]]] \cup [[\mathcal{L},\mathcal{L}],[\mathcal{L},\mathcal{L}]] \cup \cdots),$$

where span $(\mathcal{L})$ denotes the complex linear space spanned by the set $\mathcal{L}$.

Lemma 3.2: Let $\mathcal{L}$ be a set of operators. Then $\mathrm{Comp}(\mathcal{L})$ is a Lie algebra. Moreover, $\mathcal{L}$ is a Lie algebra if and only if $\mathcal{L} = \mathrm{Comp}(\mathcal{L})$.

We now introduce the new concept of a Δ algebra. The super operator Δ which projects onto the mixing part of an operator is defined in the previous section [Eq. (2.20)].

Definition 3.5: A nonzero operator A is called mixing (separable) if $\Delta A \neq 0$ ($\Delta A = 0$). The zero operator is by definition both mixing and separable.

Lemma 3.3: The property mixing (separable) is independent of the Hartree product Φ used to construct the super operator Δ.

Lemma 3.4: Let $\mathcal{D}$ be a linear space of mixing operators and let $\{A_k\}$, $k = 1,...,m$, be a basis of $\mathcal{D}$. The mapping $\Delta: \mathcal{D} \to \Delta\mathcal{D}$ is isomorphic and $\{\Delta A_k\}$, $k = 1,...,m$, is a basis of $\Delta\mathcal{D}$.

Definition 3.6: A set of mixing operators $\mathcal{D}$ is called Δ *closed* if for any integer n and for $B_1,...,B_n \in \mathcal{D}$ follows $\Delta[B_1,[B_2,...,[B_{n-1},B_n]\cdots]] \in \Delta\mathcal{D}$.

Definition 3.7: A set of mixing operators $\mathcal{D}$ is called Δ *invariant* if for any integer n and for all $A, B_1,...,B_n \in \mathcal{D}$ follows $\Delta(e^{-B_1}e^{-B_2}\cdots e^{-B_n}Ae^{B_n}\cdots e^{B_2}e^{B_1}) \in \Delta\mathcal{D}$.

Definition 3.8: A set of mixing operators is called a Δ *algebra* if it is a Δ-closed linear space.

Lemma 3.5: A Δ algebra is Δ invariant. A Δ-invariant linear space of mixing operators is a Δ algebra.

Lemma 3.6: Let $\mathcal{L}$ be a Lie algebra. Then $\Delta\mathcal{L}$ is a Δ algebra.

Definition 3.9: Let $\mathcal{L}$ be a Lie algebra and let $\mathcal{D}$ be a Δ algebra. Here $\mathcal{L}$ is called an *embedding algebra* of $\mathcal{D}$ if $\mathcal{D} \subset \mathcal{L}$ and $\Delta\mathcal{L} = \Delta\mathcal{D}$.

Definition 3.10: The *minimal embedding algebra* $\mathcal{L}_{\min}$ of the Δ algebra $\mathcal{D}$ is defined as $\mathcal{L}_{\min} = \mathrm{Comp}(\mathcal{D})$.

Definition 3.11: Let $\mathcal{L}_{\max}$ be an embedding algebra of $\mathcal{D}$. Here $\mathcal{L}_{\max}$ is called a *maximal embedding algebra* if for any operator $A \notin \mathcal{L}_{\max}$ follows $\Delta\,\mathrm{Comp}(\{A\} \cup \mathcal{L}_{\max}) \neq \Delta\mathcal{D}$.

Lemma 3.7: The maximal embedding algebra is uniquely defined.

Lemma 3.8: For any embedding Lie algebra $\mathcal{L}$ there holds $\mathcal{L}_{\min} \subset \mathcal{L} \subset \mathcal{L}_{\max}$ and $\dim(\mathcal{L}_{\min}) \leqslant \dim(\mathcal{L}) \leqslant \dim(\mathcal{L}_{\max})$.

Having studied the properties of Lie and Δ algebras as well as their relations we come back to investigate the TDRH method. We shall assume that the generators of the rotations $\{A_k\}$, $k = 1,...,m$, are a basis of a Δ algebra $\mathcal{D}$. Since $\mathcal{D}$ is Δ invariant we can expand the mixing parts of the partially rotated generators $\Delta\hat{A}_k$ in the basis; i.e.,

$$\Delta\hat{A}_k = \sum_{k'=1}^{m} D_{kk'}\Delta A_{k'}, \tag{3.25}$$

where $\hat{A}_k$ and $D_{kk'}$ are functions of the parameter α. It is important to note that the structure matrix D is independent of the Hartree product Φ used to construct the super operator Δ. The structure matrix is real. We now extend the set of Hermitian operators $\{A_k\}$ such that $\{A_k\}$, $k = 1,...,M$, becomes a basis of the maximal embedding algebra $\mathcal{L}_{\max}$. The rotated basis $\tilde{A}_k = U^{\dagger}A_kU$ can be related to the original one via the matrix R, i.e.,

$$\tilde{A}_k = \sum_{k'=1}^{M} R_{kk'}A_{k'} \tag{3.26}$$

for all k between 1 and M. Let the Hamiltonian H be a member of the maximal embedding algebra $\mathcal{L}_{\max}$, i.e.,

$$H = \sum_{k=1}^{M} h_kA_k. \tag{3.27}$$

Since $\mathcal{L}_{\max}$ is invariant we may also expand the rotated Hamiltonian in the basis

$$\tilde{H} = \sum_{k=1}^{M} \tilde{h}_kA_k. \tag{3.28}$$

By comparison one finds

$$\tilde{h}_k = \sum_{k'=1}^{M} h_{k'} R_{k'k}. \tag{3.29}$$

Using the above equations we can now express F [cf. (2.26)] as

$$F = \sum_{k'=1}^{m} \left(\sum_{j=1}^{M} h_j R_{jk'} + \sum_{k=1}^{m} \dot{\alpha}_k D_{kk'} \right) \Delta A_{k'}. \tag{3.30}$$

Hence F vanishes if

$$\dot{\alpha} = -(\mathsf{R}'\mathsf{D}^{-1})^T \mathbf{h}, \tag{3.31}$$

where we have denoted by R' the $M \times m$ matrix consisting of the first m columns of R.

The TDRH method yields the exact wave function if the differential equation (3.31) has a solution. The existence of a unique and global solution of (3.31) depends on certain properties of the matrices R and D. Their matrix elements depend on the choice and ordering of the basis. To discriminate not well behaved matrices from well behaved ones we introduce the following definition.

Definition 3.12: Let $G \subset \mathbb{R}^m$ be an open set. The generators A_k, $k = 1,...,m$, are called *properly ordered* on G if

(i) D^{-1} exists for all $\alpha \in G$,

(ii) $\|\mathsf{R}'\mathsf{D}^{-1}\| \leqslant c_1 + c_2\|\alpha\|$,

for some constants c_1, c_2 and for all $\alpha \in G$.

We remark that the choice and ordering of the basis operators not belonging to $\mathscr{D}$, i.e., A_k, $k = m+1,...,M$, is irrelevant. We are now ready to state the central result of this section.

Theorem 3.1: Let $\{A_k\}$, $k = 1,...,m$, denote a set of Hermitian operators which form a basis of some Δ algebra $\mathscr{D}$. Let the generators A_k be properly ordered on some domain $G \subset \mathbb{R}^m$. Let $\{A_k\}$, $k = 1,...,M$, denote a basis of the maximal embedding algebra $\mathscr{L}_{\max}$ and let the Hamiltonian H belong to this algebra. The expansion coefficients of $H = \Sigma_{k=1}^{M} h_k A_k$ are assumed to be uniformly bounded, i.e., $|h_k(t)| \leqslant$ const. Finally let $\alpha(t=0) \in G$. Then the TDRH method as defined by Eqs. (2.17) and (3.31) provides the *exact* wave function within some time interval $0 \leqslant t \leqslant t_0$. The parameters $\alpha(t)$ are uniquely determined. The time t_0, if finite, is given by the condition $\alpha(t_0) \in \partial G$. The solution is *global*, i.e., $t_0 = \infty$, if $\alpha(t) \in G$ for all t. A global solution exists in particular if $G = \mathbb{R}^m$.

To prove the theorem we remark that we only have to show that the differential equation (3.31) has a unique solution on G. The uniqueness of the solution is given if the differential equation satisfies the Lipschitz condition.[26] The matrix elements of D and R can be shown[20,21] to be *analytic* functions of α. The conditions raised in Definition 3.12 now ensure that the matrix elements of $\mathsf{R}'\mathsf{D}^{-1}$ are analytic on G. Hence the Lipschitz condition can be satisfied. The existence of the solution is also guaranteed by the conditions raised in Definition 3.12. The condition (ii) together with the assumption $|h_k(t)| \leqslant$ const ensures that the solution cannot "blow up," i.e., $\|\alpha(t)\| \to \infty$ for finite t.

In closing this section we comment on the case $H \notin \mathscr{L}_{\max}$. An exact solution is no longer guaranteed but one wishes to ensure the *global* existence of the approximate solution. In view of Theorem 3.1 we may conclude that TDRH has a global solution if the approximate Hamiltonian H_{app} has uniformly bounded expansion coefficients $h_k(t)$. However, because B_0 or C_0 may become singular one cannot ensure that $h_k(t)$ is bounded. But if we remove the singularities of B_0 or C_0 as discussed above [cf. (3.23)] then we obtain a global solution. This is stated in the following theorem.

Theorem 3.2: Let

(i) $\|\mathsf{D}^{-1}\| \leqslant c_D$, for all α,

(ii) $\|H(t)\psi(t)\| \leqslant c_H$, for $0 \leqslant t \leqslant t_0$,

(iii) $\|\Delta A_k \Phi\| \leqslant c_A$, for $0 \leqslant t \leqslant t_0$ and $1 \leqslant k \leqslant m$,

hold for some constants c_D, c_H, and c_A and some time t_0. The third condition is only necessary if the LVP version of TDRH is used. The equations of motion (2.27) or (3.10) are replaced by

$$\dot{\alpha} = \mathsf{D}^{-1T}(\mathsf{B}_0^2 + \epsilon \mathsf{1})^{-1} \mathsf{B}_0 \mathbf{b}_0 \tag{3.32a}$$

or

$$\dot{\alpha} = \mathsf{D}^{-1T}(\mathsf{C}_0^T \mathsf{C}_0 + \epsilon \mathsf{1})^{-1} \mathsf{C}_0^T \mathbf{c}_0 \tag{3.32b}$$

for the use of the NVP or LVP form, respectively. The number ϵ is assumed to be positive. The thus modified TDRH method has a unique solution in the time interval $[0,t_0]$. The solution is global if the assumptions (ii) and (iii) hold for all times. [If the NVP is used, then condition (iii) may be ignored.] The time derivatives of the parameters are bounded by

$$\|\dot{\alpha}\| \leqslant \epsilon^{-1/4} c_D c_H m^{1/2} \tag{3.33a}$$

or

$$\|\dot{\alpha}\| \leqslant \epsilon^{-1/2} c_D c_A c_H m \tag{3.33b}$$

for using the NVP or the LVP, respectively.

Before we sketch the proof of the theorem we remark that the assumptions (ii) and (iii) are trivially obeyed if t_0 is finite because $\|H\psi\|$ and $\|\Delta A_k \Phi\|$ are continuous functions of time. Moreover, the condition (ii) merely excludes time-dependent Hamiltonians with a somewhat wildly time dependence. For time-independent Hamiltonians we have $\|H\psi_{\text{ex}}(t)\|^2 = \langle \psi_{\text{ex}} | H^2 | \psi_{\text{ex}} \rangle = \text{const}$ and replacing the exact solution ψ_{ex} by the TDRH approximation ψ one can assume that $\|H\psi\|$ does not change considerably.

Turning to the proof of Theorem 3.2 we remark that D^{-1}, $\mathbf{b}_0$, and $\mathbf{c}_0$ are analytic functions of α ($\mathbf{b}_0$ and $\mathbf{c}_0$ depend on α via $\Delta\tilde{H}$). The Lipschitz condition[26] for the differential equations (3.32) is therefore obeyed and the solution is unique, provided it exists. The existence of the solution follows immediately from Eqs. (3.33). To prove these equations is just a matter of simple algebra. We merely remark that to prove (3.33a) one may use the diagonal representation of B_0.

We finally comment on the convergence of the solutions with respect to the limit $\epsilon \to 0$. For this purpose we introduce two definitions. A detailed analysis of the convergence is then given in Theorem 3.3.

Definition 3.13: Assume that the limits $\psi(\epsilon,t) \to \psi(0,t)$ exist and that

$$\|\psi(\epsilon,t) - \psi(0,t)\| \leqslant a \cdot \epsilon^{\nu} \tag{3.34}$$

holds for some constants a and ν and for all positive ϵ. The

supremum for all ν which satisfy Eq. (3.34) is called the *speed of convergence.*

Definition 3.14: Let d_j denote the eigenvalues of B_0 or C_0 and assume that B_0 (C_0) is singular at $t = t_s$. The singularity is called to be of *bounded by* μ if there exist some positive constants a, τ, and ϵ_0 such that

$$|d_j(\epsilon,t)| \geqslant a \cdot |t - t_s(\epsilon)|^{\mu} \tag{3.35}$$

holds for all $0 < \epsilon < \epsilon_0$ and $|t - t_s(\epsilon)| \leqslant \tau$.

Theorem 3.3: Let the conditions of Theorem 3.2 be satisfied. Let $\psi(\epsilon, t=0) = \psi_0$ be some initial condition. This wave function is propagated over the time interval $[0,t]$ accordingly to the NVP or the LVP, respectively. The final wave function $\psi(\epsilon,t)$ converges as $\epsilon \to 0$, provided one of the following sets of conditions is met.

(1) If a singularity of B_0 or C_0 is never encountered within the time interval $[0,t]$ then $\psi(\epsilon,t)$ converges with the speed $\nu \geqslant 1$ for both of the VP's.

(2a) If the exact solution is accessible then the solution of the NVP version of TDRH converges with the speed $\nu \geqslant \frac{1}{2}$.

(3a) If there is only a finite number of singularities of B_0 within the time interval $[0,t]$ and if these singularities are bounded by μ then the solution of the NVP version of TDRH converges with the speed $\nu \geqslant 1/(2\mu)$.

The statements (2a) and (3a) hold analogously also for the LVP version, provided the singularities of C_0 are common singularities of B_0, i.e., if $\text{rank}(C_0) = \text{rank}(B_0)$ for all times considered. If the matrix C_0 has singularities which are not common singularities of B_0 then one arrives at the following weaker statement.

(2b) If an exact solution is accessible and if those singularities of C_0 which are not common singularities of B_0 are bounded by μ then the solution of the LVP version of TDRH converges with the speed $\nu \geqslant 1/(2\mu)$.

Rather than prove the theorem we shall comment on it. The first point is fairly trivial. If one never encounters a singularity then the limit $\epsilon \to 0$ of course exists. More interesting is the second point. It raises *no* restrictions on the singularities [point (2a), NVP] or on the common singularities [point (2b), LVP], respectively. If an exact solution is accessible then we shall converge to it, even if we "sit" on a singularity of B_0, i.e., if B_0 is singular over a whole period of time. Such a situation occurs, e.g., if one adopts the four-dimensional Δ algebra of example 8 of Sec. IV, assumes that $H \in \mathcal{L}_{\max}$, and chooses an initial wave function as given by Eq. (4.15). As well known a Gaussian remains a Gaussian in a quadratic potential and hence the Hartree product will always assume the form (4.15). One finds that $\text{rank}(B_0) = \text{rank}(C_0) = 2$ for all times, i.e., there are constantly two vanishing eigenvalues; but this does not affect the convergence.

We now turn to point (3a). If an exact solution is not accessible then one can find arguments that the character of the singularities of B_0 is typically $\mu = 2$. In rare cases one may find $\mu = 4$ or $\mu = 6$ but in any event $\mu < \infty$. These arguments together with the point (2a) show that the NVP version of TDRH *always* converges.

Due to the very construction of the NVP [see Eq. (2.38)] it is clear that $\|F\Phi\| \leqslant \|\Delta\tilde{H}\Phi\|$ holds for all ϵ. This is *not* the case if one adopts the LVP. In fact, $\|F\Phi\|$ may diverge like $\epsilon^{-1/2}$ if an exact solution is not accessible and if we are close to a singularity of C_0 which is not a common singularity of B_0. Close to such a singularity the computed wave function changes rapidly and may not converge for $\epsilon \to 0$. In this case one has to live with the fact that one computes an ϵ-dependent approximation to the exact wave function.

Theorem 3.3 demonstrates again that the NVP version of TDRH is numerically better behaved than the LVP version. The numerical calculations we have performed[32] have shown that the dependence of $\psi(\epsilon)$ on ϵ is fortunately rather weak for both of the VP's.

IV. THE TWO-DIMENSIONAL HARMONIC ALGEBRA

In this section we shall explicitly evaluate various quantities defined in the previous section. The examples given here will help to clarify the concepts. We concentrate on systems of two degrees of freedom and in order to keep the notation simple we use x and y for the coordinates and p and q for their conjugate momenta rather than x_1, x_2 and p_1, p_2, respectively.

Any operator on the Hilbert space of square integrable functions of two variables can be considered as a function of the four basic operators x, y, p, and q. The two-dimensional harmonic algebra is spanned by all products of these operators up to the second degree, i.e., by the 15 operators $1, x, y, p, q, x^2, xy, xp, xq, y^2, yp, yq, p^2, pq, q^2$. It is easy to show that this algebra is a Lie algebra. The mixing operators in the above set are xy, xq, yp, and pq. All the other operators are separable because they operate on either the x or the y degree of freedom only. To arrive at a Hermitian basis of the Lie algebra one has, of course, to symmetrize the two operators xp and yq. We found it convenient, however, to introduce the following Hermitian linear combinations of the quadratic operators:

$$\begin{aligned} &I = xy - pq, \quad J = xq + yp, \quad K = xy + pq, \\ &L = xq - yp, \quad A = xp + qy, \quad B = xp - yq, \\ &V = \tfrac{1}{2}(x^2 + y^2), \quad W = \tfrac{1}{2}(x^2 - y^2), \\ &T = \tfrac{1}{2}(p^2 + q^2), \quad U = \tfrac{1}{2}(p^2 - q^2). \end{aligned} \tag{4.1}$$

The first four operators—I,J,K,L—are the mixing ones. The quadratic operators form a sub-Lie algebra. The commutation relations of this algebra are shown in Table I. In order to illustrate the action of the above ten operators we display in Table II the four basic operators rotated by one of the quadratic operators. The transformation generated by the operator I is a skewing transformation in the x-q and y-p planes. The operator J performs a similar transformation in the x-y and p-q planes. The operators K and L generate (true) rotations in the above mentioned planes. Note that J and L generate coordinate transformations; they mix the degrees of freedom but they do not mix coordinates and momenta. The transformations generated by I and K, however, do mix coordinates and momenta. They are hence the more difficult transformations. The transformations generated by the separable operators A and B are totally symmetric and antisymmetric scaling transformations, respectively. The potential

TABLE I. Commutation relations among the quadratic operators of the harmonic algebra. The table entries are $(1/i)[O_1, O_2]$. The entries xy and pq may be replaced by $(I+K)/2$ and $(I-K)/2$, respectively.

O_1 \ O_2	I	J	K	L	A	B	V	W	T	U	xy	pq
I	$\cdots$	$2(V+T)$	$2A$	$2(W-U)$	$2K$	0	J	L	J	$-L$	A	A
J	$-2(V+T)$	$\cdots$	$2(T-V)$	$2B$	0	$2L$	$-2xy$	0	$2pq$	0	$-2V$	$2T$
K	$-2A$	$2(V-T)$	$\cdots$	$2(W+U)$	$2I$	0	$-J$	$-L$	J	$-L$	$-A$	A
L	$2(U-W)$	$-2B$	$-2(W+U)$	$\cdots$	0	$2J$	0	$2xy$	0	$2pq$	$-2W$	$-2U$
A	$-2K$	0	$-2I$	0	$\cdots$	0	$-2V$	$-2W$	$2T$	$2U$	$-2xy$	$2pq$
B	0	$-2L$	0	$-2J$	0	$\cdots$	$-2W$	$-2V$	$2U$	$2T$	0	0
V	$-J$	$2xy$	J	0	$2V$	$2W$	$\cdots$	0	A	B	0	J
W	$-L$	0	L	$-2xy$	$2W$	$2V$	0	$\cdots$	B	A	0	L
T	$-J$	$-2pq$	$-J$	0	$-2T$	$-2U$	$-A$	$-B$	$\cdots$	0	$-J$	0
U	L	0	L	$-2pq$	$-2U$	$-2T$	$-B$	$-A$	0	$\cdots$	L	0
xy	$-A$	$2V$	A	$2W$	$2xy$	0	0	0	J	$-L$	$\cdots$	A
pq	$-A$	$-2T$	$-A$	$2U$	$-2pq$	0	$-J$	$-L$	0	0	$-A$	$\cdots$

operators V and W or the kinetic energy operators T and U mix the coordinates and momenta within each degree of freedom if they rotate the momenta or coordinates, respectively. In Tables I and II we have also included the operators xy and pq because we found it sometimes more convenient to work with these operators rather than with I and K. For the sake of completeness we present in Table III the quadratic operators as rotated by the mixing ones.

It may be useful to indicate the relation of the two-dimensional harmonic algebra with the classical Lie algebras.[30] The harmonic algebra is the semidirect sum of the radical $\{1,x,y,p,q\}$ and the simple subalgebra which is built up by the ten quadratic operators (4.1). This subalgebra is isomorphic to SP(4). The Lie algebraic structure of the harmonic algebra was recently investigated by Wolf and Korsch[33] in connection with the Wei and Norman approach.

The four mixing operators I, J, K, and L (or xy, pq, J, and L) can easily be shown to be Δ complete. Moreover, any subset of these operators is Δ complete! Hence there is a large variety of Δ algebras which are subsets of the two-dimensional harmonic Lie algebra. In the following examples we list some of these Δ algebras together with their maximal and minimal embedding Lie algebras.

Example 1:

$$\mathcal{D} = \{0\}, \quad \mathcal{L}_{\min} = \{0\}, \quad \mathcal{L}_{\max} = \{A \mid \Delta A = 0\}.$$

Example 2:

$$\mathcal{D} = \mathrm{span}\{L\}, \quad \mathcal{L}_{\min} = \mathrm{span}\{L\},$$
$$\mathcal{L}_{\max} = \mathrm{span}\{L,A,V,T,x,y,p,q,1\}.$$

Example 3:

$$\mathcal{D} = \mathrm{span}\{xy,L\}, \quad \mathcal{L}_{\min} = \mathrm{span}\{xy,L,W\},$$
$$\mathcal{L}_{\max} = \mathrm{span}\{xy,L,A,V,W,x,y,p,q,1\}.$$

Example 4:

$$\mathcal{D} = \mathrm{span}\{xy,J\}, \quad \mathcal{L}_{\min} = \mathrm{span}\{xy,J,V\},$$
$$\mathcal{L}_{\max} = \mathrm{span}\{xy,J,A,V,W,x,y,p,q,1\}.$$

Example 5:

$$\mathcal{D} = \mathrm{span}\{xy,J,L\}, \quad \mathcal{L}_{\min} = \mathrm{span}\{xy,J,L,B,V,W\},$$
$$\mathcal{L}_{\max} = \mathrm{span}\{xy,J,L,A,B,V,W,x,y,p,q,1\}.$$

Example 6:

$$\mathcal{D} = \mathrm{span}\{I,K\} \equiv \mathrm{span}\{xy,pq\},$$
$$\mathcal{L}_{\min} = \mathrm{span}\{I,K,A\},$$
$$\mathcal{L}_{\max} = \mathrm{span}\{I,K,A,B,x,y,p,q,1\}.$$

TABLE II. Basic operators rotated by the quadratic ones. The table entries are $\exp(-i\lambda O_1) O_2 \exp(i\lambda O_1)$. The last column of the table repeats the definition of the quadratic operators for convenience.

O_1 \ O_2	x	y	p	q	Definition
I	$x\cdot\cosh\lambda + q\cdot\sinh\lambda$	$y\cdot\cosh\lambda + p\cdot\sinh\lambda$	$p\cdot\cosh\lambda + y\cdot\sinh\lambda$	$q\cdot\cosh\lambda + x\cdot\sinh\lambda$	$xy - pq$
J	$x\cdot\cosh\lambda - y\cdot\sinh\lambda$	$y\cdot\cosh\lambda - x\cdot\sinh\lambda$	$p\cdot\cosh\lambda + q\cdot\sinh\lambda$	$q\cdot\cosh\lambda + p\cdot\sinh\lambda$	$xq + yp$
K	$x\cdot\cos\lambda - q\cdot\sin\lambda$	$y\cdot\cos\lambda - p\cdot\sin\lambda$	$p\cdot\cos\lambda + y\cdot\sin\lambda$	$q\cdot\cos\lambda + x\cdot\sin\lambda$	$xy + pq$
L	$x\cdot\cos\lambda + y\cdot\sin\lambda$	$y\cdot\cos\lambda - x\cdot\sin\lambda$	$p\cdot\cos\lambda + q\cdot\sin\lambda$	$q\cdot\cos\lambda - p\cdot\sin\lambda$	$xq - yp$
A	$e^{-\lambda}x$	$e^{-\lambda}y$	$e^{\lambda}p$	$e^{\lambda}q$	$xp + qy$
B	$e^{-\lambda}x$	$e^{\lambda}y$	$e^{\lambda}p$	$e^{-\lambda}q$	$xp - yq$
V	x	y	$p + \lambda x$	$q + \lambda y$	$\frac{1}{2}(x^2 + y^2)$
W	x	y	$p + \lambda x$	$q - \lambda y$	$\frac{1}{2}(x^2 - y^2)$
T	$x - \lambda p$	$y - \lambda q$	p	q	$\frac{1}{2}(p^2 + q^2)$
U	$x - \lambda p$	$y + \lambda q$	p	q	$\frac{1}{2}(p^2 - q^2)$
xy	x	y	$p + \lambda y$	$q + \lambda x$	xy
pq	$x - \lambda q$	$y - \lambda p$	p	q	pq

TABLE III. Quadratic operators rotated by the mixing ones. The table entries are $\exp(-i\lambda O_1)O_2\exp(i\lambda O_1)$.

O_1 \ O_2	I	J	K	L	A	B	V	W	T	U	xy	pq
I	I	$J\cdot\cosh 2\lambda + (V+T)\sinh 2\lambda$	$K\cdot\cosh 2\lambda + A\cdot\sinh 2\lambda$	$L\cdot\cosh 2\lambda + (W-U)\sinh 2\lambda$	$A\cdot\cosh 2\lambda + K\cdot\sinh 2\lambda$	B	$V\cdot\cosh^2\lambda + T\cdot\sinh^2\lambda + \frac{1}{2}J\cdot\sinh 2\lambda$	$W\cdot\cosh^2\lambda - U\cdot\sinh^2\lambda + \frac{1}{2}L\cdot\sinh 2\lambda$	$T\cdot\cosh^2\lambda + V\cdot\sinh^2\lambda + \frac{1}{2}J\cdot\sinh 2\lambda$	$U\cdot\cosh^2\lambda - W\cdot\sinh^2\lambda - \frac{1}{2}L\cdot\sinh 2\lambda$	$xy\cdot\cosh^2\lambda + pq\cdot\sinh^2\lambda + \frac{1}{2}A\cdot\sinh 2\lambda$	$pq\cdot\cosh^2\lambda + xy\cdot\sinh^2\lambda + \frac{1}{2}A\cdot\sinh 2\lambda$
J	$I\cdot\cosh 2\lambda - (V+T)\sinh 2\lambda$	J	$K\cdot\cosh 2\lambda + (T-V)\sinh 2\lambda$	$L\cdot\cosh 2\lambda + B\cdot\sinh 2\lambda$	A	$B\cdot\cosh 2\lambda + L\cdot\sinh 2\lambda$	$V\cdot\cosh 2\lambda - xy\cdot\sinh\lambda$	W	$T\cdot\cosh 2\lambda + pq\cdot\sinh 2\lambda$	U	$xy\cdot\cosh 2\lambda - V\sinh 2\lambda$	$pq\cdot\cosh 2\lambda + T\cdot\sinh 2\lambda$
K	$I\cdot\cos 2\lambda - A\cdot\sin 2\lambda$	$J\cdot\cos 2\lambda + (V-T)\sin 2\lambda$	K	$L\cdot\cos 2\lambda + (W+U)\sin 2\lambda$	$A\cdot\cos 2\lambda + I\cdot\sin 2\lambda$	B	$V\cdot\cos^2\lambda + T\cdot\sin^2\lambda - \frac{1}{2}J\cdot\sin 2\lambda$	$W\cdot\cos^2\lambda - U\cdot\sin^2\lambda - \frac{1}{2}L\cdot\sin 2\lambda$	$T\cdot\cos^2\lambda + V\cdot\sin^2\lambda + \frac{1}{2}J\cdot\sin 2\lambda$	$U\cdot\cos^2\lambda - W\cdot\sin^2\lambda - \frac{1}{2}L\cdot\sin 2\lambda$	$xy\cdot\cos^2\lambda + pq\cdot\sin^2\lambda - \frac{1}{2}A\cdot\sin 2\lambda$	$pq\cdot\cos^2\lambda + xy\cdot\sin^2\lambda + \frac{1}{2}A\cdot\sin 2\lambda$
L	$I\cdot\cos 2\lambda + (U-W)\sin 2\lambda$	$J\cdot\cos 2\lambda - B\cdot\sin 2\lambda$	$K\cdot\cos 2\lambda - (W+U)\sin 2\lambda$	L	A	$B\cdot\cos 2\lambda + J\cdot\sin 2\lambda$	V	$W\cdot\cos 2\lambda + xy\cdot\sin 2\lambda$	T	$U\cdot\cos 2\lambda + pq\cdot\sin 2\lambda$	$xy\cdot\cos 2\lambda - W\cdot\sin 2\lambda$	$pq\cdot\cos 2\lambda - U\cdot\sin 2\lambda$
xy	$I-\lambda A - \lambda^2 xy$	$J+2\lambda V$	$K+\lambda A + \lambda^2 xy$	$L+2\lambda W$	$A+2\lambda xy$	B	V	W	$T+\lambda J + \lambda^2 V$	$U-\lambda L - \lambda^2 W$	xy	$pq+\lambda A + \lambda^2 xy$
pq	$I-\lambda A + \lambda^2 pq$	$J-2\lambda T$	$K-\lambda A + \lambda^2 pq$	$L+2\lambda U$	$A-2\lambda pq$	B	$V-\lambda J + \lambda^2 T$	$W-\lambda L - \lambda^2 U$	T	U	$xy-\lambda A + \lambda^2 pq$	pq

Example 7:

$$\mathcal{D} = \text{span}\{I,J,K\} \equiv \text{span}\{xy,pq,J\},$$

$$\mathcal{L}_{\min} = \text{span}\{I,J,K,A,V,T\},$$

$$\mathcal{L}_{\max} = \text{span}\{I,J,K,A,V,T,x,y,p,q,1\}.$$

Example 8:

$$\mathcal{D} = \text{span}\{I,J,K,L\},$$

$$\mathcal{L}_{\min} = \text{span}\{I,J,K,L,A,B,V,W,T,U\},$$

$$\mathcal{L}_{\max} = \text{span}\{I,J,K,L,A,B,V,W,T,U,x,y,p,q,1\}.$$

We now discuss the various examples. In the first example one does not use any generator, i.e., one performs a usual Hartree calculation. The Hamiltonian lies in $\mathcal{L}_{\max}$ and the result is exact, if H is a separable operator. If one wants to use only one single generator it is convenient to use the angular momentum operator L as done in example 2. The generalized rotation becomes a true rotation. The structure matrix D reduces to unity because there is only one generator. In the third example we encounter for the first time a nontrivial structure matrix. Assume we order the generators as $A_1 = L$ and $A_2 = xy$. With the aid of Table III the structure matrix is easily calculated and reads

$$\mathsf{D} = \begin{pmatrix} 1 & 0 \\ 0 & \cos 2\alpha \end{pmatrix}. \tag{4.2}$$

Hence the structure matrix becomes singular for $\alpha = \pi/4$. If we use the other ordering $A_1 = xy$ and $A_2 = L$ we find that D is regular,

$$\mathsf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.3}$$

This illustrates the importance of ordering. The matrix R′, which—for this example and ordering—is given by

$$\mathsf{R}' = \begin{pmatrix} \cos 2\beta & 0 & 0 & \sin 2\beta & 2\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T, \tag{4.4}$$

is obviously linearly bounded. Hence the generators are properly ordered on $\mathbb{R}^2$ and a global solution (in the sense of Theorem 3.1) is guaranteed. In Eqs. (4.2) and (4.4) we found it convenient to use α and β rather than α_1 and α_2.

We now skip example 4 and turn to example 5. Using the ordering $A_1 = xy$, $A_2 = J$, and $A_3 = L$, i.e., $U = e^{i\gamma L} e^{i\beta J} e^{i\alpha xy}$, one finds that the structure matrix reads

$$\mathsf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \cosh 2\beta \end{pmatrix}. \tag{4.5}$$

This matrix is obviously regular for all parameters. However, the matrix R′ possess matrix elements containing $\sinh 2\beta$ and $\cosh 2\beta$ such that $\|\mathsf{R}'\mathsf{D}^{-1}\|$ grows like $\cosh 2\beta$ for large β. The generators are hence properly ordered on G_c, where

$$G_c = \{(\alpha,\beta,\gamma)^T \mid |\beta| < c\} \tag{4.6}$$

and where c denotes some positive number. Inspection of Table III shows that the two positive definite operators V and T are multiplied by $\cosh 2\beta$ under a generalized rotation with J. Since $\langle T \rangle \langle V \rangle \geqslant \frac{1}{4}$ due to the uncertainty principle, one finds that conservation of energy requires that $|\beta|$ is bounded. One can choose c larger than this bound which proves that example 5 provides us with a *global* solution in the sense of Theorem 3.1. Using similar arguments one can also prove that examples 4,6, and 7 yield a global solution.

Turning finally to example 8 we find that the ordering $A_1 = I$, $A_2 = J$, $A_3 = K$, and $A_4 = L$ is the most convenient one. Using this ordering the structure matrix reads

$$\mathsf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cosh 2\alpha & 0 & 0 \\ 0 & 0 & \cosh 2\alpha \cosh 2\beta & 0 \\ 0 & 0 & 0 & \cosh 2\alpha \cosh 2\beta \cos 2\gamma \end{pmatrix}, \tag{4.7}$$

which is singular for $\gamma = \pi/4$. The matrix R′ can be bounded by

$$\|\mathsf{R}'\| \leqslant c_0 \cosh 2\alpha \cosh 2\beta$$

with some constant c_0. The generators are hence properly ordered on

$$G_\epsilon = \{(\alpha,\beta,\gamma,\delta)^T \mid |a| < \epsilon^{-1},\ |\beta| < \epsilon^{-1},\ |\gamma| < \pi/4 - \epsilon\}, \tag{4.8}$$

where ϵ denotes some small positive constant. Similar as before one can argue that α and β remain bounded because of energy conservation. We cannot, however, prove that $|\gamma|$ does not approach $\pi/4$. The singularity of the structure matrix D can be removed by adding a separable operator to the set of generators. We choose the separable operator $W = \frac{1}{2}(x^2 - y^2)$ because W rotated by K contains the operator L (compare Table III). The singularity of the structure matrix was caused by the loss of the operator L from the set of partially rotated generators. We therefore define the unitary operator U now by

$$U = e^{i\eta L} e^{i\delta W} e^{i\gamma K} e^{i\beta J} e^{i\alpha I}. \tag{4.9}$$

The structure matrix D is now a rectangular 5×4 matrix relating the partially rotated operators $\Delta\hat{I}$, $\Delta\hat{J}$, $\Delta\hat{K}$, $\Delta\hat{W}$, and $\Delta\hat{L}$ to ΔI, ΔJ, ΔK, and ΔL:

$$\begin{pmatrix} \Delta\hat{I} \\ \Delta\hat{J} \\ \Delta\hat{K} \\ \Delta\hat{W} \\ \Delta\hat{L} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cosh 2\alpha & 0 & 0 \\ 0 & 0 & \cosh 2\alpha \cosh 2\beta & 0 \\ 0 & 0 & 0 & d \\ a & b & c & \cosh 2\alpha \cosh 2\beta \cos 2\gamma \end{pmatrix} \begin{pmatrix} \Delta I \\ \Delta J \\ \Delta K \\ \Delta L \end{pmatrix}, \tag{4.10}$$

where

$$a = -\delta\cdot\cosh 2\beta\cdot\cos 2\gamma,$$
$$b = \delta\cdot\cos 2\gamma\cdot\sinh 2\beta\cdot\sinh 2\alpha,$$
$$c = \delta(\sin 2\gamma\cdot\sinh 2\alpha - \cosh 2\beta\cdot\cosh 2\alpha), \quad (4.11)$$
$$d = \tfrac{1}{2}(\sinh 2\alpha - \cosh 2\alpha\cdot\cosh 2\beta\cdot\sin 2\gamma).$$

The equations of motion for the parameters are still given by [cf. (3.3)]

$$\tilde{\mathbf{h}} = -\mathsf{D}^T\dot{\boldsymbol{\alpha}}, \quad (4.12)$$

where $\tilde{\mathbf{h}}$ denotes the four-component vector of expansion coefficients of $\Delta\tilde{H}$ and where now $\dot{\boldsymbol{\alpha}} = (\dot{\alpha},\dot{\beta},\dot{\gamma},\dot{\delta},\dot{\eta})^T$ is the five-component velocity vector. Equation (4.12) has always a solution since the structure matrix defined by Eq. (4.10) has maximal rank. A unique solution can be obtained by choosing that solution for which $\|\dot{\boldsymbol{\alpha}}\|$ is minimal, i.e., by employing the pseudoinverse. Since $\mathsf{D}^T\mathsf{D}$ is nonsingular we find with the aid of Eqs. (3.17) and (3.18d)

$$(\mathsf{D}^T)^I = \mathsf{D}(\mathsf{D}^T\mathsf{D})^{-1}. \quad (4.13)$$

Using

$$\dot{\boldsymbol{\alpha}} = -(\mathsf{D}^T)^I\mathsf{R}'^T\mathbf{h} \quad (4.14)$$

we can conclude that the thus modified example 8 allows for a *global* solution.

Returning to the discussion of the examples we remark that the examples 2–5 are of greater practical importance than the examples 6 and 7. This is because the examples 6 and 7 contain the operator pq in their Δ algebras. This operator (as well as $I = xy - pq$ and $K = xy + pq$) is much harder to exponentiate than xy, J, or L. The example 8, on the other hand, is of importance because it alone has the full two-dimensional harmonic Lie algebra as embedding algebra. Hence using all four mixing operators as generators TDRH solves every harmonic problem exactly. If the Hamiltonian is not harmonic then TDRH accounts for the harmonic part of the Hamiltonian exactly and treats the anharmonic terms in a mean field approximation.

So far in this section we have implicitly assumed that the Hamiltonian is a member of the maximal embedding algebra. We now assume that $H \notin \mathscr{L}_{\max}$ and that an exact solution is not accessible. This leads to a discussion of the matrices B_0 and C_0 and in particular to an investigation of their singularities. To study the singular properties of B_0 we solve Eq. (3.16). If the Δ algebra is spanned by one of the following sets: $\{I\}$, $\{J\}$, $\{xy\}$, $\{pq\}$, $\{xy,J\}$, $\{pq,J\}$, or $\{I,J\}$ then there exist no square integrable solution of Eq. (3.16) and hence B_0 possesses no singular points. For the other Δ algebras which are subsets of the harmonic algebra we find that the solution of Eq. (3.16) has the form

$$\Phi = a\exp(-(\omega_1/2)(x-x_0)^2 + ip_0(x-x_0))$$
$$\cdot\exp(-(\omega_2/2)(y-y_0)^2 + iq_0(y-y_0)), \quad (4.15)$$

where a, ω_1, and ω_2 are complex and where x_0, y_0, p_0, and q_0 are real. Depending on the Δ algebra there are different conditions on ω_1 and ω_2. These conditions are given in Table IV. One notices that singularities are more likely for larger Δ algebras. However, the restriction (4.15) that the Hartree product is a product of two Gaussians is a condition which is hardly ever met, except, of course, if the initial wave function is chosen to be of this form. Anyhow, as shown in Sec. III the singularities of B_0 are of no physical importance although unpleasant numerically.

TABLE IV. Conditions on ω_1 and ω_2 which lead to a singularity of $\mathbf{B}_0$ (see text).

Δ algebra spanned by	Relation between ω_1 and ω_2
$\{K\}$	$\omega_1\cdot\omega_2 = 4$
$\{L\}$	$\omega_1 = \omega_2$
$\{xy,L\}$	$\mathrm{Re}(\omega_1) = \mathrm{Re}(\omega_2)$
$\{pq,L\}$	$\mathrm{Re}(1/\omega_1) = \mathrm{Re}(1/\omega_2)$
$\{J,L\}$	$\mathrm{Im}(\omega_1/\omega_2) = 0$
$\{xy,pq\} \hat{=} \{I,K\}$	$\mathrm{Im}(\omega_1\cdot\omega_2) = 0$
$\{xy,J,L\}$	no restrictions
$\{pq,J,L\}$	
$\{xy,pq,J\} \hat{=} \{I,K,J\}$	
$\{xy,pq,L\} = \{I,K,L\}$	
$\{I,J,K,L\}$	

The investigation of the singular points of the matrix C_0 is more elaborate because one has to evaluate the determinant of this matrix for each Δ algebra separately. For illustrative purposes we will do so for the Δ algebras of example 3 and 8. The C_0 matrix of example 3 is singular if

$$(x^2)_0 = (y^2)_0 \quad (4.16)$$

holds where the *number* $(O_1O_2)_0$ is defined as

$$(O_1O_2)_0 = \langle O_1O_2\rangle - \langle O_1\rangle\langle O_2\rangle \quad (4.17)$$

with operators O_1 and O_2 taken from the set $\{x,y,p,q\}$. The matrix C_0 is thus singular if the width of $\Phi_1(x)$ equals the width of $\Phi_2(y)$. This is a condition which is much easier to meet than (4.15). Hence, as mentioned before, singular points of C_0 are much more likely than singular points of B_0. We now turn to example 8 and find that the matrix C_0 has a vanishing determinant if

$$4(x^2)_0(p^2)_0 - (xp+px)_0^2 = 4(y^2)_0(q^2)_0 - (yp+qy)_0^2 \quad (4.18)$$

holds. Again, this is much weaker condition than (4.15). For the generalized Gaussian (4.15) both sides of Eq. (4.18) assume unity.

The examples 2, 5, and 7 cannot be treated directly within the LVP form of TDRH because they have an odd number of generators. The eigenvalues of a real antisymmetric matrix, as C_0 is, are imaginary and appear in complex conjugate pairs. A antisymmetric matrix of odd dimension has necessarily one vanishing eigenvalue. However, the replacement of the inverse by the pseudoinverse helps again. To show this, let us concentrate on the discussion of example 5. The matrix C_0 now reads

$$\mathsf{C}_0 = \begin{bmatrix} 0 & -W_0 & -V_0 \\ W_0 & 0 & B_0 \\ V_0 & -B_0 & 0 \end{bmatrix}. \quad (4.19)$$

The eigenvector with vanishing eigenvalues is given by

$$\mathbf{e}_0 = (B_0, V_0, -W_0)^T. \quad (4.20)$$

The matrix $C_0^T C_0 + e_0 e_0^T$ is diagonal and can be inverted analytically. Hence

$$C_0^I = (V_0^2 + W_0^2 + B_0^2)^{-1} C_0^T. \tag{4.21}$$

Since $V_0^2 + W_0^2 + B_0^2 > 0$ we find that this approximation to the inverse of C_0 is free of singularities. We found[32] that the resulting equations of motion are very convenient to use; however, they do *not* compute the exact solution of $H \in \mathcal{L}_{\max}$ of example 5. This is because one effectively uses only two generators rather than three. These two generators can be considered as time-dependent linear combinations of the three original generators xy, J, and L. Hence one has a time-dependent maximal embedding algebra and it is not easy to state for which Hamiltonians we arrive at an exact solution. For general nonharmonic Hamiltonians, however, the method outlined above gives usually better results than the use of the two-dimensional Δ algebra of example 3.[32]

V. CONCLUSION

In this paper we have analyzed the TDRH method in detail. The basic idea of TDRH is simple. One enlarges the space of trial functions of TDH by replacing the Hartree product Φ by $U(\alpha)\Phi$, where $U(\alpha)$ denotes a unitary operator depending on a set of parameters α. The time evolution of the product wave function Φ as well as of the (real) parameters α can be deduced by applying a time-dependent VP to the trial function $U(\alpha)\Phi$. We have used two different VP's, NVP and LVP, and have found that these two VP's lead to identical mean field equations for Φ but to different equations of motion for the parameters α. The differences of the two resulting methods have been discussed.

The ansatz of TDRH is reminiscent to the "optimized coordinate SCF" recently introduced by several authors[34–36] in order to improve the time-*independent* Hartree. There are, however, important differences. The action of the unitary operators of TDRH is *not* restricted to cause coordinate transformations. In fact, the inclusion of operators like $\exp(i\alpha xy)$ is very important in order to compensate the nonseparable potential terms in $\Delta\tilde{H}$ [cf. (2.26)].

The VP approach to TDRH is simple. It is, however, very difficult within this approach to prove the existence of the solution. As an alternative way to the equations of motion we therefore have also investigated the algebraic approach to TDRH. This approach allows for a much deeper analysis of the mathematical structure of TDRH. In particular it allows the formulation of theorems which show under which conditions a unique global solution exist. The algebraic approach introduces the structure matrix D which illuminates the importance of the ordering of the generators. If the generators are properly ordered—i.e., if the structure matrix has no singularities—then TDRH yields an exact global solution for Hamiltonians being members of the maximal embedding algebra $\mathcal{L}_{\max}$. (For a more precise statement see Theorem 3.1.)

If the Hamiltonian does not belong to $\mathcal{L}_{\max}$ then one has to approximate $\Delta\tilde{H}$ by some $\Delta\tilde{H}_{\text{app}}$ which belongs to $\Delta\mathcal{D}$. The two different VP's, NVP and LVP, correspond to two different constructions of $\Delta\tilde{H}_{\text{app}}$ in the algebraic approach. Using NVP (LVP) this construction introduces the matrix B_0 (C_0) which depends on the Hartree product Φ. The equations of motion become singular if B_0 (C_0) has a vanishing determinant. A singularity of B_0 indicates that—according to the construction (3.5)—there is no unique $\Delta\tilde{H}_{\text{app}}$. A singularity of C_0, on the other hand, indicates that there is either no unique or no solution at all which is consistent with Eq. (3.11), i.e., with the construction of $\Delta\tilde{H}_{\text{app}}$ according to the LVP. Singularities of C_0 are much more likely to occur than singularities of B_0. A simple trick can be used [cf. (3.23)] to eliminate the singular behavior of the equations of motion for both VP's. This modification ensures that the TDRH method yields a global approximation to the exact wave function.

ACKNOWLEDGMENT

One of us (J.K.) thanks the Deutschen Akademischen Austauschdienst (DAAD) for financial support.

[1] E. A. McCullough and R. E. Wyatt, J. Chem. Phys. **54**, 3578 (1971).
[2] A. Askar and A. S. Cakmak, J. Chem. Phys. **68**, 2794 (1978).
[3] K. C. Kulander, J. Chem. Phys. **69**, 5064 (1978).
[4] M. D. Feit, J. A. Fleck, and A. Steiger, J. Comput. Phys. **47**, 412 (1982); M. D. Feit and J. R. Fleck, J. Chem. Phys. **78**, 301 (1983).
[5] E. J. Mele and J. Socolar, Int. J. Quantum Chem. **18**, 347 (1984).
[6] R. C. Mowrey and D. J. Kouri, Chem. Phys. Lett. **119**, 285 (1985); J. Chem. Phys. **84**, 6466 (1986).
[7] D. Kosloff and R. Kosloff, J. Comput. Phys. **52**, 35 (1983).
[8] R. Kosloff and D. Kosloff, J. Chem. Phys. **79**, 1823 (1983).
[9] H. Tal-Ezer and R. Kosloff, J. Chem. Phys. **81**, 3967 (1984).
[10] *Time Dependent Hartree Fock and Beyond, Springer Lecture Notes in Physics*, Vol. 171, edited by K. Goeke and P. G. Reinhard (Springer, Berlin, 1982).
[11] R. B. Gerber, M. A. Ratner, and V. Buch, Chem. Phys. Lett. **91**, 173 (1982).
[12] R. B. Gerber, V. Buch, and M. A. Ratner, J. Chem. Phys. **77**, 3022 (1982).
[13] G. C. Schatz, V. Buch, M. A. Ratner, and R. B. Gerber, J. Chem. Phys. **79**, 1808 (1983).
[14] P. Kramer and M. Saraceno, *Geometry of the Time-Dependent Variational Principle in Quantum Mechanics, Springer Lecture Notes in Physics*, Vol. 140 (Springer, Berlin, 1981).
[15] E. J. Heller, J. Chem. Phys. **62**, 1544 (1975).
[16] E. J. Heller, J. Chem. Phys. **64**, 63 (1976).
[17] S. I Sawada, R. Heather, B. Jackson, and H. Metiu, J. Chem. Phys. **83**, 3009 (1985), and references therein.
[18] W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[19] P. Pechukas and J. C. Light, J. Chem. Phys. **44**, 3897 (1966).
[20] J. Wei and E. Norman, J. Math. Phys. **4**, 575 (1963).
[21] J. Wei and E. Norman, Proc. Am. Math. Soc. **15**, 327 (1964).
[22] J. Kucar, H. -D. Meyer, and L. S. Cederbaum, Chem. Phys. Lett. **140**, 525 (1987).
[23] A. D. McLachlan, Mol. Phys. **8**, 39 (1964).
[24] P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930).
[25] J. Frenkel, *Wave Mechanics* (Oxford U. P., Oxford, 1934).
[26] P. Hartman, *Ordinary Differential Equations* (Wiley, New York, 1964).
[27] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications* (Wiley, New York, 1971); A. Bjerhammar, *Theory of Errors and Generalized Matrix Inverses* (Elsevier, Amsterdam, 1973).
[28] R. M. Wilcox, J. Math. Phys. **8**, 962 (1967).
[29] N. Jacobson, *Lie Algebras* (Interscience, New York, 1962).
[30] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).
[31] V. S. Varadaraja, *Lie Groups, Lie Algebras and Their Representations* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[32] J. Kucar, H. -D. Meyer, and L. S. Cederbaum, work in progress.
[33] F. Wolf and H. J. Korsch, Phys. Rev. A **37**, 1934 (1988).
[34] T. C. Thompson and D. G. Truhlar, J. Chem. Phys. **77**, 3031 (1982).
[35] R. Lefebvre, Int. J. Quantum Chem. **23**, 543 (1983).
[36] Z. Bacic, R. G. Gerber, and M. A. Ratner, J. Phys. Chem. **90**, 3606 (1986).

Conditions for zero not to be an eigenvalue of the Schrödinger operator. II

A. G. Ramm
Mathematics Department, Kansas State University, Manhattan, Kansas 66506

(Received 25 September 1987; accepted for publication 3 February 1988)

A condition for zero not to be an eigenvalue of the Shrödinger operator is given.

I. INTRODUCTION

Let $H = -\nabla^2 + q(x)$ in R^3, $q = \bar{q}$, $[q] := (4\pi)^{-1} \times \int_{S^2} q(r,\omega)d\omega$, S^2 be the unit sphere in R^3, $f_+ := \max(f,0)$, and $D_R := \{x: |x| \geqslant R\}$, where R is arbitrarily large. The main result of this paper is the following theorem.

Theorem: If $Hu = 0$ in D_R, $[q]_+ \in L^{3/2}(D_R)$, and $0 < u \in L^m(D_R)$, $m < 3$, then $u = 0$ in D_R.

After publication of Ref. 1, Lieb drew my attention to his work,[2] where the result in the theorem above is given for $m = 2$ and $[q]_+ \in L^{3/2}(D_R)$. The argument in Ref. 2 relies essentially on a result in Ref. 3. It is noted in Ref. 2 that H. Brezis proved the result for $m < 3$ (unpublished). The aim of this work is to give a proof of the theorem that does not rely on the result in Ref. 3 but is based on the method given in Ref. 1 and uses the ideas and results from Ref. 2. Our proof is self-contained and relatively short. It treats the problem as a local one near infinity; the argument uses the assumptions in D_R only. In Ref. 3 the global assumption $q \in L^{3/2}(R^3)$ was used. This assumption is not necessary locally: $q(x)$ may be L^m_{loc} with $m < \frac{3}{2}$. In Ref. 1 the condition (a) $|q(x)| \leqslant c(1 + |x|)^{-a}$, $a > 2$, was used in D_R. This condition implies that (b) $q \in L^{3/2}(D_R)$, and it is very close to (b) in the sense that (b) does not hold if $a \leqslant 2$. It was shown in Ref. 1 that the theorem is not valid if $a \leqslant 2$, so that conditions (a) and (b) are exact conditions on q for zero not to be an eigenvalue of H.

In Sec. II proofs are given.

II. PROOFS

First we reduce the problem to the ODE problem. This reduction, Step 1, is given in Ref. 2, and we reproduce the argument in Ref. 2 for convenience of the reader in order to make the argument self-contained.

Step 1: Suppose

$$Hu = 0 \quad \text{in } D_R, \quad 0 < u \in L^m(D_R), \quad m < 3, \quad [q]_+ := p(r) \in L^{3/2}(D_R). \tag{1}$$

Let $f := \exp([\ln u])$. Then by Lemma 1 (below),

$$-f^{-1}\Delta f + [q] \geqslant [-u^{-1}\Delta u + q] \geqslant 0. \tag{2}$$

By Jensen's inequality, $f^m = \exp[\ln u^m] \leqslant [u^m]$, $m \geqslant 1$. Thus

$$\int_{D_r} f^m \, dx \leqslant \int_{D_R} u^m \, dx, \quad m \geqslant 1.$$

Therefore if $f \notin L^m(D_R)$ then $u \notin L^m(D_R)$. Thus if $p(r) := [q]_+$ and

$$-\Delta f + p(r)f \geqslant 0, \quad f > 0, \quad r \geqslant R, \tag{3}$$

implies that $f \notin L^m(D_R)$, then $u \notin L^m(D_R)$, $m < 3$. Since $f = f(r)$, the problem is reduced to the problem for ODE. To complete the argument let us prove the following.

Lemma 1: (See Ref. 2, p. 632.) If $f = \exp([\ln u])$ then $f^{-1}\Delta f \leqslant [u^{-1}\Delta u]$, $\Delta u := \nabla^2 u$.

Proof: If $g = \ln u$ then $u^{-1}\Delta u = \Delta g + (\nabla g)^2$. One has $[\Delta g] = \Delta[g]$, $(\nabla g)^2 \geqslant (\partial g/\partial r)^2$, and $[(\partial g/\partial r)^2] \geqslant (d[g]/dr)^2$ by Cauchy's inequality, since $[1] = 1$. Thus

$$[u^{-1}\Delta u] \geqslant \Delta[g] + (\nabla[g])^2 = f^{-1}\Delta f.$$

Step 2: Let $f = r^{-1}v$. Then (3) reduces to

$$-v'' + p(r)v := h(r) \geqslant 0, \quad v > 0, \quad r \geqslant R. \tag{4}$$

Note that

$$f \in L^m(D_R) \Leftrightarrow v \in L^m((R,\infty), t^{-(m-2)}) := \mathcal{L}^m.$$

We need to prove the following.

Lemma 2: Inequality (4) implies $v = 0$ provided that $v \in \mathcal{L}^m$, $m < 3$, and $0 \leqslant p(r) \in L^{3/2}(D_R)$, i.e., if $\int_R^\infty t^2 p^{3/2}(t)dt < \infty$.

Write (4) as

$$v = A + Br + \int_r^\infty (t - r)p(t)v(t)dt - \int_R^r (r - t)h \, dt, \tag{5}$$

where $A, B = \text{const}$. Let

$$\alpha(r) := \left(\int_r^\infty t^2 p^{3/2}(t)dt\right)^{2/3}.$$

One has

$$\begin{aligned}
&\int_r^\infty (t - r)pv \, dt \\
&\leqslant \left(\int_r^\infty t^2 p^{3/2}(t)dt\right)^{2/3} \left(\int_r^\infty t^{-1}v^3 \, dt\right)^{1/3} \\
&= \alpha(r)\left(\int_r^\infty t^{-(m-2)}v^m (t^{-1}v)^{3-m} \, dt\right)^{1/3} \\
&\leqslant \alpha(r)\|v\|^{m/3}_{\mathcal{L}^m} \|r^{-1}v\|^{(3-m)/3} \to 0 \quad \text{as } r \to \infty.
\end{aligned} \tag{6}$$

Here $\|v\| := \max_{r > R} |v|$, and the estimate

$$\|r^{-1}v\| < \infty \tag{7}$$

was used. To prove (7), differentiate (5) to get

$$v'(r) = B - \int_r^\infty pv \, dt - \int_R^r h \, dt \leqslant B, \tag{8}$$

where the inequalities $pv \geqslant 0$ and $h \geqslant 0$ were used. Integrate (8) and use the inequality $v > 0$ to get

$$0 < v(r) \leqslant Br + v(R). \tag{9}$$

From (9) inequality (7) follows. Let us prove that

$$\int_R^\infty h(t)dt < \infty. \tag{10}$$

Suppose (*) $\int_R^\infty h \, dt = \infty$. Then choose $R_1 > R$ so that

 0022-2488/88/061431-02$02.50

$\int_R^{R_1} h\,dt > 2B$, fix R_1, and choose r so large that

$$\int_R^r (r-t)h\,dt \geqslant r\int_R^{R_1}(1-r^{-1}t)h\,dt \geqslant 2Br. \tag{11}$$

From (11), (6), and (5) it follows that $v \to -\infty$ if $r \to +\infty$. This contradicts the assumption $v > 0$, $r \geqslant R$. Therefore (*) is false and (10) holds.

Proof of Lemma 2: Write (4) as $v'' \leqslant pv$ and integrate this inequality over (r, r_n), $r_n \to \infty$, and let $n \to \infty$ to get

$$-v'(r) \leqslant \int_r^\infty pv\,dt. \tag{12}$$

Here we used the fact that $v \in \mathcal{L}^m$, $m < 3$, implies existence of a sequence $r_n \to \infty$ such that $v'(r_n) \to 0$. An easy proof of this fact is left to the reader. Integrate (12) over (r, R_n), $R_n \to \infty$, and choose R_n so that $v(R_n) \to 0$ as $n \to \infty$. This is possible since $v \in \mathcal{L}^m$, $m < 3$. The result is

$$0 < v(r) \leqslant \int_r^\infty dt \int_t^\infty pv\,ds = \int_r^\infty (t-r)pv\,dt \to 0, \quad r \to \infty, \tag{13}$$

where the assumption $v > 0$ and the inequality (6) were used. Note that inequality (13) implies $v = 0$ provided that

$$\int_R^\infty tp\,dt < \infty. \tag{14}$$

Indeed (13) implies $\max_{r \geqslant r_0} v < \max_{r \geqslant r_0} v$, where r_0 is chosen so that $\int_{r_0}^\infty tp\,dt < 1$. Thus $v = 0$ for $r \geqslant r_0$. By the uniqueness of the solution to the Cauchy problem, $v = 0$ for $r \geqslant R$.

Let us derive from (13) that $v = 0$ under the assumption $p \in L^{3/2}(D_R)$. One has as in (6)

$$v \leqslant \alpha(r)\left(\int_r^\infty t^{-1}v^3\,dt\right)^{1/3} \leqslant \alpha(r)\|v\|_{\mathcal{L}^m}^{m/3}\,\|v\|^{\epsilon/3} r^{-\epsilon/3}, \tag{15}$$

where $\epsilon := 3 - m$. Define $\|r^{\epsilon/3}v\|_0 := \max_{r \geqslant r_0} |r^{\epsilon/3}v(r)| := N(v)$. Then

$$\int_r^\infty t^{-1}v^3\,dt \leqslant \int_r^\infty t^{-1-\epsilon}(r^{\epsilon/3}v)^3\,dt \leqslant N^3(v)\epsilon^{-1}r^{-\epsilon}, \quad r \geqslant r_0. \tag{16}$$

From (16) and (15) one gets

$$v \leqslant \alpha(r)N(v)\epsilon^{-1/3}r^{-\epsilon/3}. \tag{17}$$

Multiply (17) by $r^{\epsilon/3}$ to obtain

$$N(v) \leqslant \max_{r \geqslant r_0} \alpha(r)\epsilon^{-1/3}N(v). \tag{18}$$

Choose r_0 so that $\epsilon^{-1/3}\max_{r \geqslant r_0}\alpha(r) < 1$. Then (18) implies $N(v) = 0$, $v = 0$ for $r \geqslant r_0$, and, as above, $v = 0$ for $r \geqslant R$. Lemma 2 is proved.

This completes the proof of the theorem.

ACKNOWLEDGMENTS

The author thanks ONR and NSF for support and Professor E. Lieb for drawing his attention to Refs. 2 and 3.

APPENDIX: SKETCH OF BREZIS' ARGUMENT

This appendix is an excerpt from a letter by Professor H. Brezis to the editor of this Journal. In this letter Professor Brezis sketches his unpublished argument as follows.

Let $u > 0$ in D_R and

$$-\Delta u + p(r)u = q_-(x)u \geqslant 0 \quad \text{in } D_R \subset R^3. \tag{A1}$$

Let

$$b(r) := \int_R^r ds\, s^{-2}\int_R^s p(t)t^2\,dt.$$

Then $b(r) \geqslant 0$, $b(R) = b'(R) = 0$, $\Delta b = p(r)$. Multiply (A1) by $\exp[b(r)] - 1$ and integrate over D_R to get

$$\int \{-u\Delta \exp[b(r)] + pu[\exp(b(r)) - 1] - q_-u(\exp[b(r)] - 1)\}dx = 0, \tag{A2}$$

where $\int = \int_{D_R}$. Since

$$\Delta \exp(b(r)) = \exp(b(r))[p + (\nabla b)^2],$$

it follows from (A2) that

$$\int \{pu + u\exp(b(r))|\nabla b|^2 + q_-u[\exp(b(r)) - 1]\}dx = 0.$$

Thus $pu = 0$, and (A1) reduces to

$$-\Delta u \geqslant 0 \quad \text{in } D_R, \quad u \geqslant 0 \quad \text{in } D_R, \quad u \in L^m(D_R), \quad m < 3. \tag{A3}$$

This implies $u = 0$. The argument that leads to (A2) can be justified: multiply (A1) by $\{\exp[b(r)] - 1\}\eta(x/n)$, where $\eta \geqslant 0$, $\eta \in C_0^\infty$, $\eta = 1$ in a neighborhood of the origin, integrate over D_R, integrate by parts, and then let $n \to \infty$.

[1] A. G. Ramm, J. Math. Phys. **28**, 1341 (1987).
[2] E. Lieb, Rev. Mod. Phys. **53**, 603 (1981).
[3] B. Simon, J. Funct. Anal. **40**, 66 (1981).

On the power-series construction of bound states. I. The energies as zeros of the infinite Hill determinants

M. Znojil
Institute of Nuclear Physics, Czechoslovak Academy of Sciences, 250 68 Řež, Czechoslovakia

(Received 30 January 1986; accepted for publication 30 December 1987)

For a broad class of potentials, we show that the nonvariational (so-called Hill-determinant) intuitive identification of binding energies with zeros of certain infinite determinants may be given a rigorous mathematical foundation. The essence of the construction lies in an appropriate restriction of the class of the admissible *Ansätze*. This eliminates the undesirable confluence of the physical and unphysical components of $\psi(r)$ that may take place in the general case.

I. INTRODUCTION

An exceptional simplicity of the harmonic oscillator wave functions

$$\psi_n(r) = \exp\left(-\frac{1}{2}r^2\right)\sum_{m=0}^{n} p_m r^{2m+l+1} \tag{1.1}$$

(cf., e.g., Flügge[1]) forms a natural background for the perturbative[2] as well as variational[3] treatment of anharmonicities. Recently, the direct generalizations of (1.1),

$$\psi_n(r) = \exp(\text{polynomial}) \times \text{power series}, \tag{1.2}$$

were also studied in the numerical context,[4,5] with the intention of deriving the new resummation techniques,[6] etc. Besides the simplicity and semianalytic character of the wave functions, additional merit of (1.2) lies also in the possibility of an approximative replacement of the power series by a polynomial, and in the related tractability of the approximate energies as zeros of the so-called Hill determinants.[7]

In a purely formal manner, the "generalized harmonic oscillator" wave functions (1.2) may be interpreted as an *Ansatz*, converting, say, the radial differential Schrödinger equation

$$\left[-\frac{d^2}{dr^2}+\frac{l(l+1)}{r^2}+V(r)\right]\psi(r) = E\psi(r) \tag{1.3}$$

into a purely algebraic problem.[4,7,8] Unfortunately, in contrast to the more standard numerical and variational methods,[3] it is not always easy to treat properly also the physical boundary conditions in the new algebraic language.[4] Even the simplest example with the special sextic anharmonic oscillator potential,

$$V(r) = ar^2 + br^4 + cr^6, \quad c>0, \quad b<0,$$

was not treated properly from this point of view.[6,9] This inspired an extensive discussion in the literature (cf., e.g., Ref. 10 for the review).

At present, the latter $b<0$ sextic oscillator puzzle seems resolved. Singh's "completely WKB" choice of the polynomial exponent $P(r)$ in (1.2) has been proved incompatible[11] with the intuitive identification of the physical bound-state energies with zeros of the related "Hill determinants."[7] Conversely, rigorous validity of the method has been confirmed for $b>0$,[12] or for certain incompletely WKB[13] or completely non-WKB[5] modified exponents $P(r)$ in (1.2).

For a broader class of potentials, the situation will be clarified in the present paper. In Sec. II we specify the forces in more detail, in Sec. III we characterize them formally as a natural generalization of the harmonic oscillator, and, separating their class into a pair of its natural subclasses, we deliver the corresponding construction and proofs in Secs. IV and V. Section VI is a summary.

II. THE CLASS OF FORCES

An interest in simple potentials stems from the phenomenological needs of the atomic and molecular physics as well as from the methodical considerations of the perturbative field theory.[14] In our preceding paper,[15] our interest has been concentrated upon the general superpositions of the rational powers of the coordinate

$$V_0(r) = \sum_{i=1}^{N} g_i^{(0)} r^{m_i/n_i}, \quad m_N/n_N > \cdots > m_1/n_1 > -2. \tag{2.1}$$

These forces were shown to be tractable by the Hill-determinant technique, provided only that a certain "superconfinement restriction" (schematically, $g_{N-1}^{(0)}>0$) was satisfied by the coupling constants. Now we intend to get rid of it. Indeed, a formulation of a procedure independent of the particular coupling values is important methodically. Moreover, our choice and study of the class of forces (2.1) may also be shown useful from the following points of view.[15]

(a) Their flexibility forms a good background for approximations of the various realistic forces. If needed, a component $g_0^{(0)}r^{-2}$ may also be used and incorporated into the centrifugal term, $l(l+1)+g_0^{(0)} = l'(l'+1)$. The condition of regularity $l' > -\frac{1}{2}$ implies that the coupling $g_0^{(0)} > -(l+\frac{1}{2})^2$ cannot be arbitrary.

(b) The restrictions imposed upon the exponents m_i/n_i and the singular coupling $g_0^{(0)}$ are minimal. They merely reflect an acceptable nonregularity of the differential equation and of its general solution near the origin,

$$\psi(r) \sim c_1 r^{l'+1} + c_2 r^{-l}, \quad r \approx 0.$$

The latter, irregular component of $\psi(r)$ may immediately be omitted as unphysical.[16]

(c) Asymptotically, formula (2.1) incorporates both the finite wells ($m_N/n_N \leqslant 0$) and the confining potentials

 0022-2488/88/061433-07$02.50

$(g_N^{(0)} > 0,\ m_N/n_N > 0$, and $E \geqslant E_{\min})$. In the wave functions, an asymptotic appearance of a node, $\psi(\infty) = 0$, is in a one-to-one correspondence to an emergence of a new bound state. This is valid within the discrete spectrum domain (for $E < g_N^{(0)}$ if $m_N/n_N = 0$) and follows from the standard Sturm–Liouville oscillation theorems.[17] In particular, we shall have also $|\psi(r)| \sim \exp(\text{polynomial}) \gg 1$, for large $r \gg 1$ and for $E \neq E_{\text{phys}}$.

(d) An elementary change of variables,

$$\psi \to \psi_i, \quad r \to r_i, \quad l \to l_i, \quad \{E, g_n\} \to \{g_n^{(i)}, E_i\}, \quad 1 \leqslant i \leqslant i_{\max},$$

in Eq. (1.3) may be used to preserve its Schrödinger equation form and modify the potential[18] $V_0 \to V_i$. In particular, we may obtain the even polynomials

$$V_1(r^2) = \sum_{m=0}^{s} g_m^{(1)} r^{2m}, \quad g_s > 0, \tag{2.2a}$$

ordinary polynomials

$$V_2(r) = \sum_{n=-1}^{t} g_n^{(2)} r^n, \quad g_t > 0 \text{ for } t \geqslant 1, \tag{2.2b}$$

and all the fractional-power superpositions up to $i = i_{\max} = I$ and $V_I(\infty) = \text{const} < \infty$,

$$V_I(r) = \sum_{k=0}^{T} g_k^{(I)} r^{-2k/T}. \tag{2.3}$$

All these potentials are equivalent, formally, to the original force $V_0 = V_{i_0}$, and we may study any of them without loss of generality. In the paper, we shall use only the polynomial forms (2.2), recalling the special case of their mutual transformation,[18]

$$\begin{aligned} &\psi_2(r_2) = \sqrt{r_1}\,\psi_1(r_1), \quad r_2 = r_1^2, \quad 2l_2 + 1 = l_1 + \tfrac{1}{2}, \\ &g_{-1}^{(2)} = \tfrac{1}{4}(g_0^{(1)} - E_1), \quad g_0^{(2)} - E_2 = \tfrac{1}{4} g_1^{(1)}, \\ &g_m^{(2)} = \tfrac{1}{4} g_{m+1}^{(1)}, \quad m = 1,2,\ldots, \quad s - 1 \equiv t, \end{aligned} \tag{2.4}$$

as an illustration of the above implicit formula.

An analysis of the polynomial potentials (2.2) leads to an important formal difference between the even and odd degree t. Indeed, the general asymptotic solution of Eq. (1.3) reads

$$\begin{aligned} &\psi(r) \approx d_1 \exp\left(\frac{a r^{t/2+1}}{t/2+1}\right) + d_2 \exp\left(\frac{-a r^{t/2+1}}{t/2+1}\right), \quad r \gg 1, \\ &a^2 = g_t^{(2)}, \quad t \geqslant 1, \\ &a^2 = g_0^{(2)} - E, \quad t < 1, \end{aligned} \tag{2.5}$$

and may be used as a factor in Eq. (1.2) for $t = 2q$ only. In Ref. 15, an elimination of the square-root variable $r^{1/2}$ has been achieved by means of a transition (2.4) from odd t and $V = V_2$ to its even-degree representation V_1 with $t^{\text{new}} = 4q$, $t^{\text{old}} = 2q - 1$. Here, we shall treat both parities of t separately.

III. HARMONIC OSCILLATOR AS A METHODICAL GUIDE

The simplest form of *Ansatz* (1.2) reads

$$\psi(r) = e^{-\beta r^2/2} \sum_{n=0}^{N} p_n r^{2n+l+1}, \quad N \to \infty, \tag{3.1}$$

and may be inserted in our Schrödinger equation (1.3) with an arbitrary potential in principle. In order to illustrate this procedure, let us contemplate the simplest potential $V(r) = a^2 r^2$. Obviously, the recurrences obtained from (1.3),

$$\begin{aligned} &B_n p_{n+1} = (A_n - E) p_n + C_n p_{n-1}, \quad n = 0,1,\ldots, \\ &B_n = (2n+2)(2n+2l+3), \\ &A_n = (4n+2l+3)\beta, \quad C_n = a^2 - \beta^2, \end{aligned} \tag{3.2}$$

may immediately be solved by means of the explicit determinantal formula

$$p_{n+1} = \left(\prod_{k=1}^{n} B_k\right)^{-1} \cdot p_0 \det \begin{pmatrix} A_0 - E, & -B_0, & 0,\ldots & \\ C_k, & A_k - E, & -B_k, & 0,\ldots \\ 0, & C_n, & A_n - E, & \end{pmatrix}, \quad n = 0,1,\ldots, \tag{3.3}$$

for an arbitrary choice of the parameter β in (3.1).

Of course, in the particular example in question, a "clever" choice of the parameter $\beta = a$ implies that $C_k = 0$ for all k. This makes the regular infinite series (1.2) coincide with the confluent hypergeometric function. It will terminate at the exceptional values $E = E_n = (4n + 2l + 3)a$, $n = 0,1,\ldots$, of energies. Then the overall exponential factor makes the norm of $\psi(r)$ finite—the exceptional energies form a complete spectrum. Formally, we may write a determinantal equation that defines exactly the first $N - 1$ binding energies,[7]

$$p_N = 0, \quad N \gg 1. \tag{3.4}$$

The oversimplified harmonic oscillator example may simulate the more complicated cases, provided that we pick up a "wrong" value of the parameter $\beta (\neq a)$ in (3.1). Then, in a purely heuristic manner, we may still simplify Eq. (3.2) (three-term recurrences) by a change of variables:

$$p_n = \beta^n q_n / n!, \quad n \geqslant 0. \tag{3.5}$$

In an asymptotic domain of indices, a simplified form of (3.2),

$$q_{n+1} - q_n = \tfrac{1}{4}(a^2/\beta^2 - 1) q_{n-1}, \quad n \gg 1, \tag{3.6}$$

may be used in place of the original difference equation.

The new equation has constant coefficients and may be solved by the *Ansatz*

$q_n = \lambda^n$.

After a determination of $\lambda = \lambda_\pm = (1 \pm a/\beta)/2$ from (3.6), the general solution of Eq. (3.2) becomes available,

$$p_n = \frac{c_1}{n!}\left(\frac{\beta+a}{2}\right)^n + \frac{c_2}{n!}\left(\frac{\beta-a}{2}\right)^n, \quad n \gg 1, \tag{3.7}$$

and leads also to the $\beta = a$ termination condition. Of course, its relation to the normalizability physical requirement becomes less obvious.

In the detailed analysis, we may recall (2.5), i.e.,

$$\psi(r) = d_1 e^{ar^2/2} + d_2 e^{-ar^2/2}, \quad r \gg 1, \tag{3.8}$$

where $d_1 = 0$ if and only if $E = E_{\text{phys}}$.

At the same time, we may insert (3.7) in (3.1) and write

$$\psi(r)\exp(\tfrac{1}{2}\beta r^2) \approx \text{a polynomial} + \sum_{n>N} \cdots$$

$$\approx c_1 \exp\left(\frac{\beta+a}{2} r^2\right) + c_2 \exp\left(\frac{\beta-a}{2} r^2\right)$$

$$+ O(r^{2N}), \quad r \gg 1, \quad N \gg 1. \tag{3.9}$$

Thus, provided that $\beta > 0$, the asymptotics of $\psi(r)$, $r \to \infty$, remain unphysical ($d_1 \neq 0$) whenever $c_1 \neq 0$ in Eq. (3.7) or (3.9). Conversely, both the coefficients d_1 and c_1 change sign precisely at $E = E_{\text{phys}}$ when treated as functions of the variable E. Obviously, we may write

$$N! p_N = \left(\frac{\beta+a}{2}\right)^N \left(c_1 + c_2\left(\frac{\beta-a}{\beta+a}\right)^N\right)$$

$$= c_1((\beta+a)/2)^N(1 + 0(\epsilon^N)),$$

$$0 < \epsilon = (\beta-a)/(\beta+a) < 1, \tag{3.10}$$

and see that the roots of p_N coincide with the physical binding energies E_n in the limit $N \to \infty$. In this way, a rigorous foundation of the Hill-determinant method is obtained—in accord with (3.3), we may employ the approximate secular equation (3.4),

$$\det\left[\begin{pmatrix} A_0 & -B_0 & & \\ C_1 & A_1 & -B_1 & \\ & \ddots & \ddots & \\ & & C_N & A_N \end{pmatrix} - EI\right] = 0, \quad N \gg 1, \tag{3.11}$$

and determine the binding energies by the standard computation algorithms.[3]

IV. POTENTIALS EQUIVALENT TO POLYNOMIALS OF AN ODD DEGREE

For $t = 2q - 1$ in the definition of the potentials $V = V_2(r)$, we may use the change of variables (2.4) and consider an equivalent potential

$$V(r) = g_0 + g_1 r^2 + \cdots + g_{2q} r^{4q}, \quad g_{2q} = a^2 > 0, \tag{4.1}$$

giving the quartic anharmonic oscillator in the simplest nontrivial $q = 1$ case. In accord with the standard Hill-determinant computations,[7] we may choose the exponential (subdominant) factor in an almost arbitrary way in (1.2),

$$\psi(r) = r^{l+1}\exp(-g(r)) \sum_{n=0}^{\infty} p_n r^{2n},$$
$$g(r) = \sum_{j=1}^{q} \beta_j r^{2j}/2j. \tag{4.2}$$

Then, an insertion in Eqs. (1.3) and (4.1) leads to the basic $(2q+2)$-term recurrences of the type (3.2),

$$B_n p_{n+1} = (A_n - E)p_n + \sum_{j=1}^{q-1} p_{n-j} C_n^{(j)}$$
$$+ \sum_{j=q}^{2q} p_{n-j} D^{(j+1-q)},$$
$$n = 0,1,\ldots, \quad p_{-1} = p_{-2} = \cdots = 0. \tag{4.3}$$

Of course, the coefficients are functions of the couplings and other parameters,

$$B_n = (2n+2)(2n+2l+3),$$
$$A_n = (4n+2l+3)\beta_1 + g_0 \equiv C_n^{(0)} + E,$$
$$C_n^{(j)} = (4n+2l+3-2j)\beta_{j+1} + g_j$$
$$+ \sum_{i=0}^{j-1} \beta_{i+1}\beta_{j-i},$$
$$j = 1,2,\ldots,q-1, \quad n = 0,1,\ldots, \tag{4.4}$$
$$D^{(j)} = g_{j+q-1} + \sum_{i=j-1}^{q-1} \beta_{i+1}\beta_{q+j-i-1},$$
$$j = 1,2,\ldots,q,$$
$$D^{(q+1)} = g_{2q} = a^2.$$

A discussion of the physical normalization requirement remains an open question here.

In general, we may expect an asymptotic growth of $\psi(r)$:

$$\psi(r) \approx d_1 \exp\left(\frac{ar^{2q+1}}{2q+1}\right) + d_2 \exp\left(\frac{-ar^{2q+1}}{2q+1}\right), \quad r \gg 1. \tag{4.5}$$

At the same time, each row of Eq. (4.3) defines the nonzero coefficient p_{n+1} at each energy E,

$$p_{n+1} = \left(\prod_{k=0}^{n} B_k\right)^{-1} p_0 \det\begin{pmatrix} A_0 - E, & -B_0, & 0,\ldots & \\ C_1^{(1)}, & A_1 - E, & -B_1, & 0,\ldots \\ & & \ddots & \\ 0,\ldots,0, & D^{(q+1)},\ldots,D^{(1)}, & C_n^{(q-1)},\ldots,C_n^{(1)}, & A_n - E \end{pmatrix}. \tag{4.6}$$

Hence the standard finite-dimensional analogies make no sense in the infinite-dimensional system in question.[10] A detailed analysis of relations between the physical asymptotics of $\psi(r)$, $r \to \infty$, and p_n, $n \to \infty$, necessitates a more careful argumentation as indicated schematically in the preceding section.

In the first step, we shall employ a change of variables inspired by Eq. (3.5),

$$p_{n+1} = \frac{\lambda^n h_n}{\Gamma(1+(n+1)/(2q+1))\Gamma(1+(n+l+\frac{3}{2})/(2q+1))}. \tag{4.7}$$

This converts our basic recurrences (4.3) into a set of the alternative difference equations

$$h_n - h_{n-2q-1} = \sum_{m=0}^{q-1} (w_{q-m} h_{n+m-q} + w_{2q-m} h_{n+m-2q}), \tag{4.8}$$

each of which may be numbered by the parameter k such that

$$\lambda = \lambda_{(k)} = \left[\frac{a^2}{(4q+2)^2}\right]^{1/(2q+1)} \exp\left(i\frac{2\pi k}{2q+1}\right), \quad k = 0, \pm 1, \pm 2, \ldots, \pm q, \tag{4.9}$$

in the *Ansatz* (4.7). The new values of coefficients in (4.8),

$$w_{q-m} = \frac{C_n^{(q-1-m)}\Gamma(1+(n+1)/(2q+1))\Gamma(1+(n+l+\frac{3}{2}/(2q+1))}{\lambda^{q-m} B_n \Gamma(1+(n+1+m-q)/(2q+1))\Gamma(1+(n+l+\frac{3}{2}+m-q)/(2q+1))},$$

$$w_{2q-m} = \frac{D^{(q-m)}\Gamma(1+(n+1)/(2q+1))\Gamma(1+(n+l+\frac{3}{2})/(2q+1))}{\lambda^{2q-m} B_n \Gamma(1+(1+n+m-2q)/(2q+1))\Gamma(1+(n+l+\frac{3}{2}+m-2q)/(2q+1))},$$

$$m = 0,1,\ldots,q-1, \tag{4.10}$$

are ordered with respect to their asymptotic decrease [cf. Fig. 1(a)],

$$w_{q-m} = O(n^{-(2m+1)/(2q+1)}),$$
$$w_{2q-m} = O(n^{-(2m+2)/(2q+1)}). \tag{4.11}$$

In comparison with the harmonic oscillator methodical example (3.6), our equations (4.8) have an asymptotically negligible right-hand-side expression—one of their solutions should obviously be almost constant for $n \gg 1$.

The almost-constant solution may easily be represented by its Taylor series truncated after a few terms. In this case, we may replace also each of the difference equations (4.8) by the simplest nontrivial differential equation approximation,

$$(2q+1)\frac{d}{dn} h_n = w_q h_n + \text{corrections}, \quad n \gg 1, \tag{4.12}$$

and derive immediately the leading-order form of the almost-constant solution. Assuming for simplicity that $\beta_q \neq 0$, we get

$$w_q = (2q+1)^{-1}\beta_q [n/(2q+1)]^{-1/(2q+1)}/\lambda_{(k)}^q + \text{corrections},$$

i.e.,

$$h_n = h_n^{(k)} = \exp\left\{\frac{\beta_q}{2q\lambda_{(k)}^q}[(2q+1)^{-1}n]^{2q/(2q+1)} + \text{corrections}\right\}, \quad n \gg 1, \tag{4.13}$$

for each particular parameter k. This is similar to the construction of the harmonic oscillator solutions for each $e = e_{\pm}$. The general solution of the present difference equation (4.8) may be written as a superposition of the type (3.7),

$$\Gamma\left(1+\frac{n+1}{2q+1}\right)\Gamma\left(1+\frac{n+l+\frac{3}{2}}{2q+1}\right)p_{n+1} = \sum_{k=-q}^{q} c_k \lambda_{(k)}^n h_n^{(k)}, \quad n \gg 1. \tag{4.14}$$

This completes the first part of our analysis—an insertion of (4.13) will enable us to derive the present analog of the harmonic oscillator formula (3.10).

The proof of the previous statement is not complicated. First, we renumber the roots (4.9) in such a way that $\lambda_{(k)}^{-q} = \lambda_{(k_0)}/\lambda_0^{q+1}$, i.e., we put $k_0 = j - q$ for $k = 2j+1$, $k_0 = j$

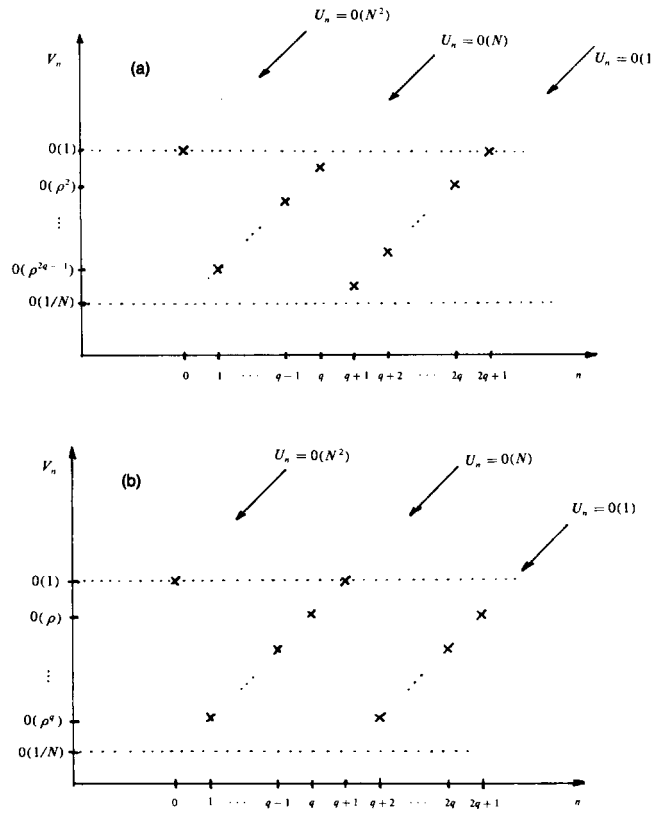


FIG. 1. Order of magnitude of coefficients in our recurrences $\sum_{j=0}^{2q+1} U_j p_{N+1-j} = 0$ and $\sum_{j=0}^{2q+1} V_j h_{N-j} = 0$, for large $N \gg 1$: (a) $t = 2q-1$, $\rho = N^{-1/(2q+1)}$; (b) $t = 2q$, $\rho = N^{-1/(q+1)}$.

for $k=2j$, and $k_0=0$ if and only if $k=0$. Next, we notice that $\mathrm{Re}\,\lambda_{(k_0)}<\lambda_{(0)}$ for all $k_0\neq 0$. Thus due to (4.13) we have $|h_n^{(k)}|\ll|h_n^{(0)}|$, for all $k\neq 0$ and $n\gg 1$. This means that in full analogy to (3.10), we may neglect all the exponentially small corrections and write, for sufficiently large indices N,

$$\Gamma\left(1+\frac{N+1}{2q+1}\right)\Gamma\left(1+\frac{N+l+\frac{3}{2}}{2q+1}\right)p_{N+1}$$
$$\approx c_0\lambda_{(0)}^{N}h_N^{(0)}$$
$$\approx c_0\lambda_{(0)}^{N}\exp\left\{\frac{\beta_q}{2q\lambda_{(0)}^{q}}\left[\frac{N}{2q+1}\right]^{2q/(2q+1)}\right\},\quad N\gg 1. \tag{4.15}$$

The asymptotic zeros of p_N and c_0 will coincide.

In the next step of our considerations, we may try to extend the asymptotic estimate (3.9) beyond the harmonic oscillator trivial case. Thus combining (4.2) with (4.15) we arrive at an estimate

$$\psi(r)=r^{l+1}\exp(-g(r))\left(O(r^{2N})+\sum_{h=N}^{\infty}p_nr^{2n}\right), \tag{4.16}$$

where the summation may be approximated by an integration,

$$\phi_N(r)=\sum_{n=N}^{\infty}p_nr^{2n}\approx c_0\int_N^{\infty}dn\,\frac{r^{2n}\exp(n\ln\lambda_{(0)}+\text{corrections})}{\Gamma(1+(n+1)/(2q+1))\Gamma(1+(n+l+\frac{3}{2})/(2q+1))},\quad N\gg 1. \tag{4.17}$$

Then a change of variables leads to the formula

$$\phi_N(r)\approx\frac{2q+1}{2}c_0\int_{N/(q+1/2)}^{\infty}dm\,\frac{r^{(2q+1)m}\exp[(q+\frac{1}{2})m\ln\lambda_{(0)}+\text{corrections}]}{\Gamma(1+m/2+1/(2q+1))\Gamma(1+m/2+(l+\frac{3}{2})/(2q+1))}$$
$$\approx c_0\sum_{m=[N/(q+\frac{1}{2})]}^{\infty}\frac{r^{m(2q+1)}}{m!}\exp\left[m\left(\ln 2+\frac{2q+1}{2}\ln\lambda_0\right)(1+O(m^{-1/(2q+1)}))\right]$$
$$\approx c_0\exp[ar^{2q+1}/(2q+1)+\text{corrections}],\quad r\gg 1, \tag{4.18}$$

compatible, within the achieved precision, with the asymptotic $d_1\neq 0$ estimate (4.5).

A common zero in c_0, $\phi_N(r)$, or $\psi(r)$ is in a one-to-one correspondence to the energy, crossing its bound-state value $E=E_{\text{phys}}$. We may conclude that the relation

$$\det\left[\begin{pmatrix} A_0, & -B_0, & 0,\dots & \\ C_1^{(1)}, & A_1, & -B_1, & 0,\dots \\ & \ddots & \ddots & \\ 0,\dots,0, & D^{(q+1)},D^{(q)},\dots, & C_N^{(1)}, & A_N \end{pmatrix}-EI\right]=0,\quad N\to\infty, \tag{4.19}$$

defines the bound-state energies for the potential (4.1).

V. POTENTIALS EQUIVALENT TO POLYNOMIALS OF AN EVEN DEGREE

The pair of forces (4.1) and

$$V(r)=g_0+g_1r^2+\cdots+g_{2q+1}r^{4q+2},\quad g_{2q+1}=a^2>0, \tag{5.1}$$

represent the complete class of the general "fractionally anharmonic" oscillators (2.1) [cf. Sec. II (d)]. From the purely formal point of view the new interaction (5.1) contains one more coupling, but we may postulate

$$\psi(r)=r^{l+1}\exp(-g(r))\sum_{n=0}^{\infty}p_nr^{2n},$$
$$g(r)=\sum_{j=1}^{q+1}\frac{\beta_jr^{2j}}{2j},\quad \beta_{q+1}=a>0, \tag{5.2}$$

and arrive again at the $(2q+2)$-term recurrences

$$B_np_{n+1}=(A_n-E)p_n+\sum_{j=1}^{q}p_{n-j}C_n^{(j)}+\sum_{j=q+1}^{2q}p_{n-j}D^{(j-q)},$$
$$n=0,1,\dots,\quad p_{-1}=p_{-2}=\cdots=0, \tag{5.3}$$

analogous to Eq. (4.3).

Before going into details, let us recall a close analogy between (5.1) and the harmonic oscillator [$\equiv$ a special case of (5.1) with $q=0$], and reparametrize (5.1):

$$V(r)=W_1(r)+W_2^2(r)r^2,$$
$$W_1(r)=c_0+c_1r^2+\cdots+c_qr^{2q}, \tag{5.4}$$
$$W_2(r)=\gamma_1+\gamma_2r^2+\cdots+\gamma_{q+1}r^{2q},\quad \gamma_{q+1}=a.$$

In terms of the new couplings, we may express then also the explicit definitions of coefficients in (5.3),

$$B_n = (2n+2)(2n+2l+3),$$
$$A_n = (4n+2l+3)\beta_1 + c_0 \equiv C_n^{(0)} + E,$$
$$C_n^{(j)} = (4n+2l+3-2j)\beta_{j+1} + c_j + d_j,$$
$$d_j = \sum_{i=0}^{j-1} (\gamma_{j-i} + \beta_{j-i})(\gamma_{i+1} - \beta_{i+1}),$$
$$D^{(j)} = \sum_{i=j-1}^{q-1} (\gamma_{q+j-i} + \beta_{q+j-i})(\gamma_{i+1} - \beta_{i+1}),$$
$$j = 1,2,...,q, \tag{5.5}$$

with the possible explicit transition between the definitions (5.1) and (5.4) of the potential,

$$\gamma_{q-m} = \frac{1}{2a}\left(g_{2q-m} - \sum_{i=1}^{m} \gamma_{q+m-i}\gamma_{q+1-i}\right),$$
$$m = 0,1,...,q-1, \tag{5.6}$$

etc.

In comparison with the preceding section, the change of variables (4.7) finds its present counterpart in the relations

$$p_{n+1} = \frac{\lambda^n h_n \Gamma(1 + (n+v)/(q+1))}{\Gamma(1 + (n+1)/(q+1))\Gamma(1 + (n+l+\frac{3}{2})/(q+1))},$$
$$4v = 2l + 3 - 2q + (1/a)(c_q + d_q), \quad n = 0,1,..., \quad \lambda^{q+1} = a/(q+1). \tag{5.7}$$

This transformation converts our basic difference equation (5.3) into the $q+1$ different equivalents

$$h_n - h_{n-q-1} = \sum_{m=0}^{q-1} (w_{q-m} h_{n+m-q} + w_{2q+1-m} h_{n+m-2q-1}), \tag{5.8}$$

where the auxiliary parameter $\lambda = \lambda_{(k)}$ has a form similar to Eq. (4.9),

$$\lambda = \lambda_{(k)} = \left|\left(\frac{a}{q+1}\right)^{1/(q+1)}\right| \exp\left(i\frac{2\pi k}{q+1}\right), \quad k = 0,1,...,q. \tag{5.9}$$

The coefficients

$$w_{q-m} = \frac{C_n^{(q-m-1)} \Gamma(1 + (n+m-q+v)/(q+1))\Gamma(1 + (n+1)/(q+1))\Gamma(1 + (n+l+\frac{3}{2})/(q+1))}{B_n \lambda^{q-m} \Gamma(1 + (n+v)/(q+1))\Gamma((n+m+2)/(q+1))\Gamma((n+m+l+\frac{3}{2})/(q+1))},$$

$$w_{2q+1-m} = \frac{D^{(q-m)}(q+1)\Gamma((n+m-q+v)/(q+1))\Gamma(1 + (n+1)/(q+1))\Gamma(1 + (n+l+\frac{3}{2})/(q+1))}{aB_n \lambda^{q-m} \Gamma(1 + (n+v)/(q+1))\Gamma((n+m+1-q)/(q+1))\Gamma((n+m+l+\frac{3}{2}-q)/(q+1))},$$
$$m = 0,1,...,q-1, \tag{5.10}$$

exhibit again the $n \gg 1$ asymptotic decrease

$$w_{q-m} = (q+1)^{-1}\beta_{q-m}\lambda_{(k)}^{m-q}[n/(q+1)]^{-(m+1)/(q+1)} + \text{corrections},$$
$$w_{2q+1-m} = (q+1)^{-1}D^{(q-m)}[n/(q+1)]^{-(m+1)/(q+1)}/4a\lambda_{(k)}^{q-m} + \text{corrections}, \tag{5.11}$$

displayed also in Fig. 1(b).

In full analogy with the preceding section, the simple $n \gg 1$ differential equation approximations to Eq. (5.8),

$$(q+1)^2 \frac{d}{dn} \ln h_n = \frac{\lambda_{(k)}^{-q}}{2}\left(\frac{n}{q+1}\right)^{-1/(q+1)} [\gamma_q + \beta_q + \text{corrections}], \tag{5.12}$$

give $q+1$ independent solutions

$$h_n = h_n^{(k)} = \exp\left[\frac{\beta_q + \gamma_q}{2q\lambda_{(k)}^q}\left(\frac{n}{q+1}\right)^{q/(q+1)} + \text{corrections}\right],$$
$$n \gg 1, \quad k = 0,1,...,q, \tag{5.13}$$

numbered by the subscript k used in (5.9).

The remaining q independent solutions of Eq. (5.8) with $D^{(q)} \neq 0$ may be found in the same way as above—only the suitable change of variables

$$p_{n-q} = \mu^n \tilde{h}_n / \Gamma(1 + (n+r)/q), \quad \mu^q = -(1/4aq)D^{(q)},$$
$$r = (1/4a)[(2l+3-2q)a + c_q + d_q], \tag{5.14}$$

differs from Eq. (5.7) and replaces Eq. (5.3) by the q relations

$$\tilde{h}_n - \tilde{h}_{n-q} = \frac{\mu}{(n/q)^{1/q}}\left(-\frac{D^{(q)}}{4a^2}\tilde{h}_{n+q+1} - \frac{\beta_q}{a}\tilde{h}_{n+1} + \frac{D^{(q-1)}}{D^{(q)}}\tilde{h}_{n-q+1}\right) + \text{corrections}, \quad n \gg 1. \tag{5.15}$$

We arrive at the q missing and independent quasiconstant solutions

$$\tilde{h}_n^{(k)} = \exp(\text{const} \cdot \mu_{(k)} n^{(q-1)/q} + \text{corrections}),$$
$$n \gg 1, \quad k = 1,2,...,q, \tag{5.16}$$

immediately. A similar procedure works for $D^{(q)} = 0$ as well.

The detailed form of constants in (5.16) is not needed.

When we compare (5.7) with (5.14), we see that the latter q components of p_{n+1} are suppressed by a huge overall factor,

$$\Gamma(\text{const}+n/q)/\Gamma(\text{const}+n/(q+1))$$
$$\approx\Gamma(\text{const}+n/[q(q+1)]),\quad n\gg 1.$$

We may omit them completely from the general solution of Eq. (5.3) and write the latter, within the present error bounds, in the form

$$\frac{\Gamma(1+(n+1)/(q+1))\Gamma(1+(n+l+\frac{3}{2})/(q+1))}{\Gamma(1+(n+v)/(q+1))}p_{n+1}$$
$$=\sum_{k=0}^{q}c_k\lambda_{(k)}^{\,n}h_n^{(k)},\quad n\gg 1, \tag{5.17}$$

fully analogous to our former Eq. (4.14).

In what follows we shall assume a choice of β_q such that

$$\beta_q+\gamma_q>0. \tag{5.18}$$

As a consequence, the right-hand-side sum in (5.17) becomes dominated by the $k=0$ component [cf. Eq. (4.15)]. Indeed, for $k\neq 0$, we have always a nonzero imaginary part in the factor

$$1/\lambda_{(k)}^{\,q}=\lambda_{(0)}^{-q}\exp(-2\pi ikq/(q+1))$$
$$=\lambda_{(0)}^{-q}\exp(-2\pi i(q+1-k)/(q+1)).$$

An insertion in (5.2),

$$\psi(r)=r^{l+1}\exp(-g(r))[O(r^{2N})+\phi_N(r)],\quad r\gg 1,$$
$$\phi_N(r)=\sum_{n=N}^{\infty}p_nr^{2n},$$

enables us to write

$$\phi_N(r)\approx c_0\int_N^{\infty}dn\,r^{2n}\frac{\lambda_{(0)}^{\,n}}{\Gamma(1+(n+1)/(q+1))}$$
$$\times\exp O(n^{q/(q+1)})$$
$$\approx c_0\sum_{m>N/(q+1)}\frac{r^{2m(q+1)}}{m!}$$
$$\times\exp\left(m\ln\frac{a}{q+1}+O(m^{q/(q+1)})\right)$$
$$\approx c_0e^{ar^{2q+2/(q+1)}},\quad r\gg 1, \tag{5.19}$$

in full agreement with the unphysical asymptotic growth [(2.5) with $d_1\neq 0$] whenever $c_0=c_0(E)\neq 0$. Conversely, the zeros of $c_0(E)$ will coincide with the zeros of $\psi(r)$, $r\gg 1$, as well as with the zeros of the Hill determinant,

$$\det\left[\begin{pmatrix} A_0, & -B_0, & 0,\ldots & \\ C_1^{(1)}, & A_1, & -B_1, & 0,\ldots \\ & \ddots & \ddots & \\ 0,\ldots,0, & D^{(q)},\ldots,D^{(1)}, & C_N^{(q)},\ldots,C_N^{(1)}, & A_N \end{pmatrix}-EI\right]$$
$$=0,\quad N\gg 1. \tag{5.20}$$

The proof of the validity of the method [Eq. (3.4)] for the potential (5.1) is completed.

VI. SUMMARY

In this paper we have considered the complete class of the fractional potentials (1.4). We have transformed (i) the potentials into their two "canonical" forms, (4.1) and (5.1), in accord with Ref. 15; (ii) the regular wave functions $\psi(r)$, $r\in(0,\infty)$, into their appropriate Taylor coefficients p_n, $n=1,2,\ldots$ [cf. Eqs. (4.2) and (5.2), respectively]; (iii) the radial (ordinary differential) Schrödinger equation into its difference equation equivalents [Eqs. (4.3) and (5.3), respectively]; and (iv) the standard physical boundary condition

$$\psi(r)=0,\quad r\to\infty, \tag{6.1}$$

into its "Hill-determinant" equivalent

$$p_N=0,\quad N\to\infty. \tag{6.2}$$

The core of the paper lies in a complete asymptotic solution of the difference equations, showing that the asymptotically dominant component of p_n is in a one-to-one correspondence to the $r\gg 1$ asymptotically dominant component of the wave function $\psi(r)$ itself. Our main result [the equivalence between the two boundary conditions (6.1) and (6.2)] is a property of our particular choice of the transformation $\psi(r)\to p_n$, and, in light of the existing counterexamples,[11] it need not be valid in general, of course.

[1] S. Flügge, *Practical Quantum Mechanics I* (Springer, Berlin, 1971), pp. 68 and 160.
[2] See Ref. 1, p. 80.
[3] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
[4] C. A. Ginsburg, Phys. Rev. Lett. **48**, 839 (1982).
[5] A. Hautot, Phys. Rev. D **33**, 437 (1986).
[6] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978); Lett. Math. Phys. **3**, 73 (1979).
[7] See, for example, S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, Phys. Rev. D **4**, 3617 (1971); J. Math. Phys. **14**, 1190 (1973); D. P. Datta and S. Mukherjee, J. Phys. A **13**, 3161 (1980); D. Masson, J. Math. Phys. **24**, 2074 (1983).
[8] M. Znojil, Lett. Math. Phys. **5**, 405 (1981).
[9] A. H. Wilson, Proc. R. Soc. London Ser. A **118**, 617 (1928).
[10] M. Tater, J. Phys. A **20**, 2483 (1987).
[11] R. N. Chaudhuri, Phys. Rev. D **31**, 2687 (1985).
[12] M. Znojil, Phys. Rev. D **26**, 3750 (1982).
[13] M. Znojil, Phys. Rev. D **34**, 1224 (1986).
[14] C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).
[15] M. Znojil, J. Math. Phys. **24**, 1136 (1983).
[16] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1982), Chap. 12.
[17] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. 10.
[18] M. Znojil, J. Phys. A **16**, 213 (1983).

A new proof of the generalized Birkhoff theorem

C. Bona
Department de Fisica, Universitat Illes Balears, 07071 Palma de Mallorca, Spain

(Received 16 July 1987; accepted for publication 9 September 1987)

A new proof of the generalized Birkhoff theorem in general relativity is presented. The partial results contained in previous proofs are recovered in a unified treatment of the different kinds of space-times to which the theorem applies. The proof is based on the fact that these space-times are conformal to the direct product of two two-dimensional manifolds (almost-product structure).

I. INTRODUCTION

The original formulation of the well-known Birkhoff theorem[1] applied only to spherically symmetric vacuum space-times. Its standard generalization applies to space-times with an energy-momentum tensor posessing two double eigenvalues either in the case of the spherical[2,3] or associated symmetries,[4,5] even when the orbits of the isometry group are timelike.[6]

In this work, a new proof of the generalized theorem is given. The partial results contained in Refs. 2–6 are recovered in a unified treatment of the different kinds of space-times to which the generalized Birkhoff theorem applies. The proof is based on the fact that all these space-times are conformally reducible, in the sense of Petrov.[7]

This conformal reduction was used by Petrov[8] in connection with the original version of the Birkhoff theorem, working in specific coordinate systems. The proof of the generalized theorem that is presented here is intrinsic, for the class of adapted coordinate systems in which the almost-product structure of the space-time metric is manifest.

The paper is organized as follows. In Sec. II, the conformal reduction of the space-times considered is performed and the generalized Birkhoff theorem is properly stated. In Sec. III, the components of the Ricci tensor are computed and an important lemma is proved. The proof of the theorem is given in Sec. IV and, finally, Sec. V contains comments on previous formulations on the same theorem.

II. SPACE-TIMES WITH SPHERICAL OR ASSOCIATED SYMMETRIES

Let us consider a pseudo-Riemannian manifold $(V_4,\hat{g})$ admitting a three-dimensional isometry group G_3 acting on two-dimensional non-null orbits[9] O_2. The space-time is then conformally reducible,[7] and the metric $\hat{g}$ has the following structure:

$$\hat{g} = Y^2\mathbf{g}, \tag{1}$$

where g is reducible[7] and it can be thought of as the metric of a direct product space-time $(V_4,\mathbf{g})$.

Let us construct a local coordinate system by taking local coordinates $\{y^A\}$ $(A=1,2)$ in O_2 and $\{x^a\}$ $(a=3,4)$ in the surfaces V_2 orthogonal to O_2. In this adapted coordinate system, the conformal factor Y and the metric $\mathbf{g}$ can be written as follows:

$$Y = Y(x), \quad \mathbf{g} = h_{AB}(y)dy^A dy^B + g_{ab}(x)dx^a dx^b, \tag{2}$$

where h and g are the two-dimensional metrics induced by $\mathbf{g}$ on O_2 and V_2, respectively.

The two-dimensional manifold V_2 is of constant curvature, so that the conformal factor Y can be normalized in order to have

$$\mathrm{Ric}(h) = kh, \tag{3}$$

where k can be either $+1$ (spherical symmetry), 0 (plane symmetry), or -1 (hyperbolic symmetry). The specific form of the metric h depends both on the sign of k and on the causal character of the (non-null) orbits O_2.

We are now in the position to state the following theorem.

Theorem 1 (Generalized Birkhoff theorem): Every space-time admitting a three-dimensional isometry group G_3 acting on two-dimensional orbits O_2 and with Ricci tensor pertaining to the algebraic types [(11)(1,1)] or [(111,1)] (Segré notation) admits at least a four-dimensional isometry group G_4, provided that

$$d_a Y \neq 0. \tag{4}$$

III. COMPUTATION OF THE RICCI TENSOR OF $(V_4,\hat{g})$

The Ricci tensor of $(V_4,\hat{g})$ can be computed in two steps. First of all, one can obtain the Ricci tensor of $(V_4,\mathbf{g})$ in terms of k and the Gaussian curvature R of the two-dimensional manifold (V_2,g),

$$\mathbf{R}_{AB} = kh_{AB}, \quad \mathbf{R}_{ab} = Rg_{ab}, \quad \mathbf{R}_{Ab} = \mathbf{R}_{aB} = 0, \tag{5}$$

as it follows easily from the direct product structure (2) of $\mathbf{g}$ and from Eq. (3).

The second step consists in applying to $\hat{g}$ and $\mathbf{g}$ the well-known formulas relating the curvature tensors of two metrics that are conformal one to another.[10] A straightforward calculation gives

$$\begin{aligned} R_{AB} &= [k + Y(\nabla^c d_c Y^{-1}) - 3Y^{-2}(\nabla^c Y \nabla_c Y)]h_{AB}, \\ R_{Ab} &= R_{aB} = 0, \\ R_{ab} &= 2Y\nabla_a d_b Y^{-1} + [R + Y(\nabla^c d_c Y^{-1}) \\ &\quad - 3Y^{-2}(\nabla^c Y \nabla_c Y)]g_{ab}, \end{aligned} \tag{6}$$

where ∇ stands for the covariant derivative in (V_2,g) and all contractions are made using the two-dimensional metric g.

Expressions (6) are covariant in (V_2,g) and allow an intrinsic formulation of the Einstein field equations for the space-times verifying (1)–(3). Here we are only interested in the algebraic structure of the Ricci tensor of $(V_4,\hat{g})$. The decomposition (6) leads directly to the following result.

Lemma 1: The algebraic type of the Ricci tensor of

 0022-2488/88/061440-02$02.50

$(V_4,\hat{g})$ is either [(11)(1,1)] or [(111,1)] (Segré notation) if and only if the second covariant derivatives of $1/Y$ [in the sense of (V_2,g)] are proportional to g, that is,

$$\nabla_a d_b(1/Y) \propto g_{ab} . \tag{7}$$

IV. PROOF OF THE THEOREM

The proof consists of three steps. Let us suppose that the hypotheses of Theorem 1 are verified, so that, allowing for Lemma 1, Eq. (7) holds true.

(1) Let us construct the two-dimensional vector u with components

$$u^a(x) = e^{ab} d_b(1/Y) , \tag{8}$$

where e_{ab} stands for the antisymmetric tensor in (V_2,g). The vector u is nonzero, as we are supposing that the function Y is not constant [Eq. (4)]. It is easy to see that u is a Killing vector of (V_2,g),

$$\nabla_b u_a = \nabla_b [e_a^c \nabla_c (1/Y)] \propto e_{ab} . \tag{9}$$

(2) Let us construct now a four-dimensional vector v with components

$$v^a \equiv u^a(x), \quad v^A \equiv 0 , \tag{10}$$

so that it is tangent to the surfaces V_2 orthogonal to O_2. It follows from (9) and the direct product structure of $\mathbf{g}$ (2) that v is a Killing vector of $(V_4,\mathbf{g})$, that is,

$$L_v(\mathbf{g}) = 0 , \tag{11}$$

where $L(\ )$ stands for the Lie derivative operator in V_4.

(3) Equation (11) implies that the vector v defined in (10) must be a conformal vector of the metric $\hat{g}$ conformally related to $\mathbf{g}$ by (1), that is,

$$L_v(\hat{g}) = 2L_v(Y)\hat{g} \tag{12}$$

and the factor $L_v(Y)$ can be easily computed from (10) and the definition (8) of u,

$$L_v(Y) = u^a(x) d_a Y = 0 , \tag{13}$$

so that v is in fact a fourth Killing vector of $(V_4,\hat{g})$ and Theorem 1 is proved.

There are two more results arising from the proof presented above.

Corollary 1: The components of the fourth Killing vector v are

$$v^a = u^a(x), \quad v^A = 0 ,$$

where u is the two-dimensional vector explicitly given in Eq. (8).

Corollary 2: The fourth Killing vector is invariant by the isometry subgroup G_3 acting on O_2. The extension to a G_4 isometry group is then central.

V. COMMENTS

The generalized Birkhoff theorem is sometimes stated without the restriction (4).[11] This is not correct because, in the case in which Y is constant, the metric $\hat{g}$ itself has a direct product structure and its Ricci tensor (6) becomes

$$R_{AB} = kh_{AB}, \quad R_{ab} = Rg_{ab}, \quad R_{Ab} = R_{aB} = 0 , \tag{14}$$

so that it pertains always to the required algebraic types and one has no restriction at all on the two-dimensional metric g_{ab}. For a generic form of g_{ab}, the space-time $(V_4,\hat{g})$ does not admit a fourth Killing vector.[12]

In other versions,[13] the condition (4) is replaced by

$$d^c Y d_c Y \neq 0 , \tag{15}$$

which is more restrictive than (4) in the case in which the orbits O_2 are spacelike. The cases not covered by (15) are discussed by Foyster and McIntosh.[14] Note that, in these cases, Eq. (7) becomes

$$\nabla_a d_b(1/Y) = 0 , \tag{16}$$

so that (V_2,g) must be flat (we are supposing that Y is not constant). It is easy to verify that, in these cases, the two-dimensional vector u defined by (8) is isotropic.

This vector coincides up to a sign with the vector w defined as follows:

$$w^a(x) = g^{ab} d_b(1/Y) , \tag{17}$$

and a straightforward calculation shows that the four-dimensional isotropic vector W defined as

$$W^a = w^a(x), \quad W^A = 0 \tag{18}$$

is covariantly constant in the sense of $(V_4,\hat{g})$. The space-times $(V_4,\hat{g})$ can be interpreted then as plane-fronted gravitational waves with parallel rays (*pp* waves[10]) and the metric forms are given in Ref. 14.

ACKNOWLEDGMENT

The author acknowledges financial support from the Spanish Comisión Asesora para la Investigación Científica y Técnica (CAICYT) under Project No. 1005/84.

[1] G. D. Birkhoff, *Relativity and Modern Physics* (Harvard U.P., Cambridge, MA, 1923).
[2] M. Cahen and R. Deveber, C. R. Acad. Sci. Paris **260**, 815 (1965).
[3] J. Plebanski and J. Stachel, J. Math. Phys. **9**, 269 (1968).
[4] H. Goener and J. Stachel, J. Math. Phys. **11**, 3358 (1970).
[5] H. Goener, Commun. Math. Phys. **16**, 34 (1970).
[6] A. Barnes, Commun. Math. Phys. **33**, 75 (1973).
[7] A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969).
[8] A. Z. Petrov, Sov. Phys. JETP **17**, 1026 (1963).
[9] The null orbits case is not considered in this work. It has been discussed in Ref. 6.
[10] D. Kramer *et al.*, *Exact solutions of Einstein's Field Equations* (VEB, Berlin, 1980).
[11] See, for instance, statement (3) in Ref. 5.
[12] These metrics correspond to the case (iv) in the classification given in Ref. 13.
[13] See, for instance, Ref. 14.
[14] J. M. Foyster and C. B. G. McIntosh, Commun. Math. Phys. **27**, 241 (1972).

The $\bar{S}(a,b,c/m)$ metrics interpreted as colliding wave solutions

Alberto García D.
Departamento de Física, Centro de Investigación y de Estudios Avanzados del I.P.N., Apdo. Postal 14-740, 07000, México D.F., Mexico

(Received 27 October 1987; accepted for publication 3 February 1988)

It is shown that the class of solutions $\bar{S}(a,0,-1/1)$, a real slice of the "complexified" Plebański $S(a,b,c/m)$ metrics, is interpretable as colliding wave solutions and, for $a=-s_2$, coincides precisely with the Ferrari–Ibañez colliding wave metrics.

I. INTRODUCTION

The main purpose of this work is to demonstrate the equivalence of the $\bar{S}(a,0,-1/1)$ metrics and the Ferrari–Ibañez class of colliding wave solutions.

II. PRESENTATION

In 1980 Plebański[1] published a large class of static Weyl solutions, denoted by $S(a,b,c/m)$, which can be given in a real chart $\{q,p,\sigma,\tau\}$ $(1<q<\infty,\ -1<p<1)$ by the metric structure

$$\begin{aligned} m^{-2}g_4 &= f^{-1}[\mathscr{A}g_2+\mathscr{B}\,d\sigma^2]-f\,d\tau^2,\\ \mathscr{A} &= (q+1)^{(a+b)^2}(q-1)^{(a-b)^2}\\ &\quad\times(1+p)^{(a+c)^2}(1-p)^{(a-c)^2},\\ \mathscr{B} &= (q^2-1)(1-p^2),\\ f &= (q+1)^{-(a+b)}(q-1)^{b-a}\\ &\quad\times(1+p)^{-(a+c)}(1-p)^{c-a},\\ g_2 &= (q^2-p^2)(q+p)^{-(b+c)^2}(q-p)^{-(b-c)^2}\\ &\quad\times\left[\frac{dq^2}{q^2-1}+\frac{dp^2}{1-p^2}\right], \end{aligned} \tag{1}$$

where a, b, c, and m are arbitrary constants.

By complex coordinate transformations (scaling transformations of the form $q\to q/q_0, p\to p/p_0, \sigma\to\sigma/\sigma_0, \tau\to\tau/\tau_0$, accompanied by complex scaling of parameters), the metric (1) can be brought to the form

$$\begin{aligned} m^{-2}\tilde{g}_4 &= \tilde{f}^{-1}[\tilde{\mathscr{A}}\tilde{g}_2+\tilde{\mathscr{B}}\,d\sigma^2]+\tilde{f}\,d\tau^2,\\ \tilde{\mathscr{A}} &= (1+q)^{(a+b)^2}(1-q)^{(a-b)^2}\\ &\quad\times(1+p)^{(a+c)^2}(1-p)^{(a-c)^2},\\ \tilde{\mathscr{B}} &= (1-p^2)(1-q^2)=:\rho^2,\\ \tilde{f} &= (1+q)^{-(a+b)}(1-q)^{b-a}\\ &\quad\times(1+p)^{-(a+c)}(1-p)^{c-a},\\ \tilde{g}_2 &= (p^2-q^2)(p+q)^{-(b+c)^2}(p-q)^{-(b-c)^2}\\ &\quad\times\left\{\frac{dp^2}{1-p^2}-\frac{dq^2}{1-q^2}\right\}, \end{aligned} \tag{2}$$

which are real solutions of cylindrically symmetric character; ∂_σ and ∂_τ are now spacelike Killing vectors. This class of solutions is naturally denoted by $\bar{S}(a,b,c/m)$.

According to the theorem on colliding wave solutions belonging to the CW_1 class,[2] it is necessary and sufficient that

$$b+c=\pm 1,\quad b-c=\pm 1. \tag{3}$$

Thus the $\bar{S}(a,\pm 1,0/1)$ and $\bar{S}(a,0,\pm 1/1)$ metrics are interpretable as colliding wave metrics. It is easy to show that these solutions are equivalent one to another.

Let us consider the particular case of $\bar{S}(a,0,-1/1)$. Then the structural functions and the line element $\tilde{g}_2$ of the above metric $\tilde{g}_4$ reduce to

$$\begin{aligned} \tilde{\mathscr{B}} &= (1-p^2)(1-q^2)=:\rho^2,\\ \tilde{f} &= \rho^{-2a}(1+p)/(1-p),\\ \tilde{\mathscr{A}} &= \rho^{2a^2}(1+p)^{1-2a}(1-p)^{1+2a},\\ \tilde{g}_2 &= \frac{dp^2}{1-p^2}-\frac{dq^2}{1-q^2}. \end{aligned} \tag{4}$$

The $\bar{S}(a,0,-1/1)$ is interpretable as a colliding wave solution, since by accomplishing the transformations

$$p=\cos\phi,\quad q=\cos\theta,\quad \sigma=x^1,\quad \tau=x^2,\quad a=-s_2 \tag{5}$$

in $\tilde{g}_4$ from (2) with structural functions and $\tilde{g}_2$ from (4), for $m=1$, one arrives, modulo the change of signature, justly at the colliding wave solutions derived and studied by Ferrari and Ibañez.[3]

We expect that certain cylindrically symmetric real slices of the "complexified" multiexponent Weyl metrics[4] could bear a colliding wave interpretation.

[1] J. F. Plebański, Phys. Rev. Lett. **45**, 303 (1980).
[2] F. J. Ernst, A. García D., and I. Hauser, J. Math. Phys. **28**, 2951 (1987).
[3] V. Ferrari and J. Ibañez, Gen. Relativ. Gravit. **19**, 405 (1987).
[4] J. F. Plebański and A. García D., Phys. Rev. Lett. **48**, 1447 (1982).

 0022-2488/88/061442-01$02.50

On self-similar Tolman models

S. D. Maharaj
Department of Mathematics and Applied Mathematics, University of Natal, King George V Avenue, Durban 4001, South Africa

(Received 1 December 1987; accepted for publication 27 January 1988)

The self-similar spherically symmetric solutions of the Einstein field equation for the case of dust are identified. These form a subclass of the Tolman models. These self-similar models contain the solution recently presented by Chi [J. Math. Phys. **28**, 1539 (1987)], thereby refuting the claim of having found a new solution to the Einstein field equations.

I. INTRODUCTION

The assumption of self-similarity in general relativity has particular appeal because of the simplification that results in the Einstein field equations. The mathematical simplification arises because the metric functions, in a spherically symmetric problem, are now essentially functions of the single variable ct/R, where R is the radial coordinate and t is the coordinate time. Consequently the field equations reduce to a system of ordinary differential equations. The solution of these equations should be useful in astrophysics, for example in studying the asymptotic behavior of a relativistic supernova shock wave created by a source from which energy is released.

Recently Chi[1] found a self-similar spherically symmetric solution to the field equations for the special case of dust. However, the general dust solutions for spherically symmetric space-times, the Tolman models,[2] are known and are listed by Kramer *et al.*[3] Thus the solution of Chi must be a particular Tolman model. We reexpress the solution of Chi to ease comparison with the equations of Kramer *et al.* This solution contains the Einstein–de Sitter model as a particular case.

We extend the solution of Chi by finding two further classes of self-similar solutions for the Tolman metrics. In fact the self-similar solutions presented in this paper are the only possible self-similar spherically symmetric solutions admitted by the field equations formulated by Chi. Throughout we follow the notation and conventions of Chi.[1]

II. FIELD EQUATIONS

The spherically symmetric metric can be put in the form

$$ds^2 = c^2 e^{\sigma(t,R)}\,dt^2 - e^{\omega(t,R)}\,dR^2 - r^2(t,R)(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{1}$$

where R is the comoving radial coordinate. Cahill and Taub[4] define a self-similar spherically symmetric solution of the field equations as one for which the resulting space-time admits a conformal Killing vector. They then show that the metric (1) must be of the form

$$ds^2 = c^2 e^{\sigma(\xi)}\,dt^2 - e^{\omega(\xi)}\,dR^2 - R^2 S(\xi)(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{2}$$

where ξ is the self-similarity variable. In his attempt to obtain self-similar solutions Chi[1] expressed the field equations in terms of the dimensionless functions

$$8\pi G p t^2 = N(\xi), \quad 8\pi G p t^2/c^2 = Q(\xi),$$
$$2Gmt^2/R^3 = M(\xi), \quad r = RS(\xi).$$

(Note that if we replace ct with R in these expressions we then obtain the equations considered by Cahill and Taub.[4]) The field equations can then be written as a system of ordinary differential equations,

$$M = \xi^2 S[1 + e^{-\sigma}S'^2 - e^{-\omega}(S - \xi S')^2], \tag{3a}$$
$$3M - \xi M' = NS^2(S - \xi S'), \tag{3b}$$
$$\xi M' - 2M = -\xi Q S^2 S', \tag{3c}$$
$$\xi\omega' = -2(\xi N' - 2N)/(Q + N) - 4\xi S'/S, \tag{3d}$$
$$\sigma' = -2Q'/(Q + N), \tag{3e}$$

where a prime denotes differentiation with respect to ξ. In the case of dust the pressure vanishes so that $Q = 0$ and the field equations (3) can be easily integrated. The metric (2) now has the particular form

$$ds^2 = c^2\,dt^2 - \xi^4 S^{-4} N^{-2}\,dR^2 - R^2 S^2(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{4}$$

where

$$SS'^2 + S(1 - M_0^2) - M_0 = 0, \tag{5a}$$
$$NS^2(S - \xi S') - M_0\xi^2 = 0, \tag{5b}$$

and M_0 is a constant of integration. It remains to obtain the function $S(\xi)$.

III. THE CASE $M_0 = 1$

This special case was considered by Chi.[1] Equations (5) yield

$$S = [\tfrac{3}{2}(\xi_0 \pm \xi)]^{2/3}, \tag{6a}$$
$$N = \tfrac{4}{9}\xi^2(\xi_0 \pm \xi)^{-1}(\xi_0 \pm \xi/3)^{-1}. \tag{6b}$$

On using Eq. (5b) we obtain the relationship

$$\xi^4 N^{-2} S^{-4} = (S + RS')^2 = \left(\frac{\partial r}{\partial R}\right)^2.$$

Then the metric (4) can be written as

$$ds^2 = c^2\,dt^2 - \left(\frac{\partial r}{\partial R}\right)^2 dR^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2). \tag{7}$$

The solution (7) is a special case of the general dust equation (13.39) (on setting $\epsilon = 0$) of Kramer *et al.*[3] We use (6a) to obtain an explicit form for r,

 0022-2488/88/061443-02$02.50

$$t - t_0(R) = \pm [m(R)]^{1/2} r^{3/2}, \tag{8}$$

where $t_0(R) = \mp \xi_0 R/c$ and $m(R) = [2/(3cR^2)]^{1/2}$. The result (8) is equivalent to Eq. (13.38a) of Kramer *et al.* (In the general dust solutions the quantities t_0 and m are arbitrary functions of R.) Thus we have shown that the self-similar dust solution of Chi is a particular Tolman model.

With the aid of Eq. (6) the metric (4) can be written as

$$ds^2 = c^2\,dt^2 - (\tfrac{3}{2})^{4/3}(\xi_0 \pm \xi)^{-2/3}(\xi_0 \pm \xi/3)^2\,dR^2 - (\tfrac{3}{2})^{4/3} R^2 (\xi_0 \pm \xi)^{4/3}(d\theta^2 + \sin^2\theta\,d\phi^2). \tag{9}$$

This self-similar solution contains essentially only one arbitrary constant, namely ξ_0. It is interesting to observe that when we set $\xi_0 = 0$ in (9) we get

$$ds^2 = c^2 dt^2 - t^{4/3}\,dR^2 - t^{4/3}(d\theta^2 + \sin^2\theta\,d\phi^2) \tag{10}$$

after a rescaling of coordinates. The space-time (10) is the familiar Einstein–de Sitter model.

IV. $M_0 \neq 1$

This case was not considered by Chi.[1] However, the field equations (5) can also be integrated for $M_0 \neq 1$ and we obtain two further classes of self-similar Tolman models. Upon integration (5a) yields

$M_0^2 > 1$:

$$S = M_0(M_0^2 - 1)^{-1}(\cosh\eta - 1),$$

$$\xi - \xi_0 = \pm M_0(M_0^2 - 1)^{-3/2}(\sinh\eta - \eta),$$

$M_0^2 < 1$:

$$S = M_0(1 - M_0^2)^{-1}(1 - \cos\eta),$$

$$\xi - \xi_0 = \pm M_0(1 - M_0^2)^{-3/2}(\eta - \sin\eta).$$

We can write this solution in the standard form

$$t - t_0(R) = \pm h(\eta) m(R) f^{-3}, \tag{11}$$

where

$$t_0(R) = \xi_0 R/c, \quad m(R) = M_0 R/c,$$

$$f = \begin{cases} (M_0^2 - 1)^{1/2}, & \text{for } M_0^2 > 1, \\ (1 - M_0^2)^{1/2}, & \text{for } M_0^2 < 1, \end{cases}$$

$$h(\eta) = \begin{cases} \sinh\eta - \eta, & \text{for } M_0^2 > 1, \\ \eta - \sin\eta, & \text{for } M_0^2 < 1. \end{cases}$$

Of course Eq. (11) is contained in the general dust equation (13.38b) of Kramer *et al.*

These self-similar solutions contain essentially two arbitrary constants, namely ξ_0 and M_0. Note that self-similarity seems to isolate those solutions for which $m(R) \sim R$. For $M_0 = 1$ we have $m(R) \sim 1/R$. For $M_0 \neq 1$ the metric (4) becomes

$$ds^2 = c^2\,dt^2 - \left(\frac{1}{M_0}\frac{\partial r}{\partial R}\right)^2 dR^2 - r^2(d\theta^2 + \sin^2\theta\,d\phi^2), \tag{12}$$

where r must satisfy

$$r = h'(\eta) m(R) f^{-2}.$$

The metrics (7) and (12) comprise the entire set of self-similar Tolman models.

V. ENERGY DENSITY

Here we briefly study the behavior of the energy density ρ in the self-similar Tolman models. The energy density is the only nonvanishing dynamical quantity in the Tolman models. For the case $M_0 = 1$ we have

$$\rho = (18\pi G)^{-1} t^{-2} \xi^2 (\xi_0 \pm \xi)^{-1}(\xi_0 \pm \xi/3)^{-1}. \tag{13}$$

Note that the expression corresponding to (13) given by Chi incorrectly contains an additional factor of ξ^2. For small values of ξ, we obtain the behavior

$$\rho \sim \xi^2/t^2 = c^2/R^2,$$

and ρ has the form of an inverse square law. For large values of ξ we obtain the behavior

$$\rho \sim t^{-2}.$$

Also, in the classes of solution for which $M_0 \neq 1$ the energy density ρ can be written as

$$\rho = M_0(8\pi G)^{-1} t^{-2} \xi^2 S^{-2}\left(\frac{\partial r}{\partial R}\right)^{-1}.$$

If $\partial r/\partial R \sim S^{-2}$ then clearly $\rho \sim \xi^2/t^2$, and again ρ has the form of an inverse square law.

ACKNOWLEDGMENTS

I am grateful to Professor W. B. Bonnor for enlightening discussions on the Tolman models.

[1] L. K. Chi, J. Math. Phys. **28**, 1539 (1987).
[2] R. C. Tolman, Proc. Natl. Acad. Sci. (USA) **20**, 69 (1934).
[3] D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., London, 1980).
[4] M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).

Hauser–Malhiot spaces admitting a perfect fluid energy-momentum tensor

Taxiarchis Papacostas
Department of Physics, University of Crete, 711-10 Iraklion, Greece

(Received 30 December 1986; accepted for publication 28 October 1987)

In this paper a special case of Hauser–Malliot (HM) space-times is examined in the presence of a perfect fluid source. The obtained solutions are all known except for a generalization of a stationary axisymmetric solution found by Kramer [Class. Quantum Gravit. **1**, L3 (1984)].

I. INTRODUCTION

In a 1968 paper Carter[1] examined spaces with a two-parameter Abelian isometry group, G_2 (condition I in Ref. 1), invertible with non-null surfaces of transitivity (condition II in Ref. 1), in which the Hamilton–Jacobi equation (HJ) is separable for non-null geodesics (condition IV in Ref. 1). The metric satisfying these four conditions is metric (77) in Ref. 1. Carter, in order to simplify the field equations, imposed a supplementary condition (III S in Ref. 1), the separability of the Schrödinger equation. The necessary and sufficient condition for the separability of the Schrödinger equation is relation (79) in Ref. 1. The resulting spaces split into four families $[\tilde{A}]$, $[\tilde{B}(+)]$, $[\tilde{B}(-)]$, and $[\tilde{D}]$ in the presence of a nonsingular electromagnetic field.

The separability of the HJ equation, which has been shown independently by Matravers[2] and Carter,[1] gives rise to a fourth constant of the motion for particle orbits (the three other constants being the rest mass, the energy, and the angular momentum about the symmetry axis). This implies the existence of a second-rank Killing tensor in these solutions. Matravers has also shown that the separability of the HJ equation for null geodesics gives rise to a nonzero quadratic first integral for the null geodesic equations, which implies the existence of a conformal Killing tensor. Walker and Penrose[3] found that all the vacuum type-D solutions admit an irreducible second-rank conformal Killing tensor for null geodesics, while an irreducible second-rank Killing tensor and its corresponding first integral for all geodesics exists only in a subclass of these solutions.

The results of Walker and Penrose have been extended to type-D electovac solutions with an aligned nonsingular electromagnetic field by Hughston *et al.*[4] and by Hughston and Sommers.[5] The latter authors have shown that the C metric and the C-NUT metric and their electrovac generalizations are the only metrics in the class that do not admit the Killing tensor. Sufficient conditions for the separability of the HJ equation have been given by Woodhouse.[6] He relates the separability of the HJ equation with the existence of second-rank Killing tensors (Theorem 4.2).

In the same spirit with Carter, Bonanos[7] has studied spaces with a two-parameter, invertible Abelian isometry group in which the HJ equation for null geodesic separates in the presence of perfect fluid sources. The separability of the HJ equation for null geodesics is equivalent to the existence of a conformal Killing tensor (the conformal Killing tensor which corresponds to a [(11) (11)] Killing tensor). This is the only work that has been done on spaces containing, as special cases, metrics admitting [(11) (11)] Killing tensors in the presence of a perfect fluid energy-momentum tensor. A generalization of Carter's space-times has been made by Hauser and Malhiot (HM), in two successive papers.[8,9] In the first (1976) they presented the set of all space-times that admit a second-rank Killing tensor whose Segre characteristics are [(11) (11)], or, equivalently, a second-rank Killing tensor with two double nonconstant eigenvalues, λ_1 and λ_2. In a special coordinate system the resulting metric has the same general form as Carter's HJ separable metric [(77) in Ref. 1] except that there is no group of symmetry. However, under an additional assumption on the Ricci tensor (the R_{12} component of the Ricci tensor vanishes) the metric of HM admits a two-parameter Abelian isometry group and, in this case, the space-time coincides with Carter's metric (77) under the restriction of relation (79). The vanishing of the R_{12} component of the Ricci tensor is the sufficient and necessary condition for the separability of the Schrödinger equation and it is completely equivalent to relation (79) of Carter.

In a second paper (1978) HM completed their previous results and obtained a larger class of metrics, namely, they found all space-times that admit, or are conformal to, those that admit nonsingular [(11) (11)] Killing tensors, with nonsingular meaning with the nonconstant eigenvalues λ_1 and λ_2. Their general metric splits into four subfamilies: (1,1) when $\delta_1\delta_2\neq 0$; (0,1) when $\delta_1 = 0$, $\delta_2\neq 0$; (1,0) when $\delta_1\neq 0$, $\delta_2 = 0$; and (0,0) when $\delta_1 = \delta_2 = 0$, where $\delta_1 = \rho - \bar{\rho}$ and $\delta_2 = \tau + \bar{\tau}$ are expressions of the Newman–Penrose (NP) spin coefficients. In this second paper they also proposed two possible directions of research: the first is to find spaces with $R_{12}\neq 0$ (non-Schrödinger separable) and the second is the search for physically plausible matter tensors, which permit the hidden symmetry characterized by a [(11) (11)] Killing tensor.

In the present paper we examine only the (1,1) subfamily of HM spaces in the presence of a perfect fluid source and under the assumption of the separability of the Schrödinger equation.

In the (1,1) subfamily of HM spaces, the existence of a [(11) (11)] Killing tensor in the presence of a perfect fluid energy-momentum tensor implies the existence of a G_2 (at least), invertible with a non-null surface of transivity. The orbits of the group are timelike (one timelike Killing field and one spacelike), or spacelike (two spacelike Killing fields).

Our work generalizes, in a certain way, Bonanos' results, because (a) we also examine the spacelike case (Bonanos examined only the timelike case, namely, the stationary

 0022-2488/88/061445-06$02.50

axisymmetric spaces), and (b) we find all solutions for the timelike case (Bonanos has not found Kramer's solution and its generalization). We say "in a certain way" because Bonanos' spaces are more general than the spaces examined in this paper since they admit the conformal Killing tensor corresponding to the [(11) (11)] Killing tensor, which characterizes the (1,1) subfamily of HM spaces.

In Sec. II we present the (1,1) subfamily of HM spaces and we make a classification of this subfamily based on the eigenvalues λ_1 and λ_2 of the [(11) (11)] Killing tensor.

In Sec. III we formulate the consequences of the existence of a perfect fluid source for the (1,1) case in NP formalism.

In Secs. IV and V we solve the field equations and we present the obtained solutions.

We shall perform our calculations using the NP formalism[10] and the complex vectorial formalism of Cahen, Debever, and Defrise.[11–13]

II. SPACES ADMITTING [(11) (11)] KILLING TENSORS

The Killing tensor has the following form in a local coordinate system (x^i) (see Refs. 8, 9, and 14):

$$K_{ij} = \lambda_1(n_i l_j + l_i n_j) + \lambda_2(m_i \bar{m}_j + \bar{m}_i m_j)\,, \tag{1}$$

where the covariant vectors l_i and n_i are real and null and the complex null vectors m_i and $\bar{m}_j$ are complex conjugate. The functions λ_1 and λ_2 are real.

The Killing tensor equation,

$$\nabla_{(k} K_{ij)} = 0\,, \tag{2}$$

can be written in the NP notation:

$$k = \sigma = \lambda = \nu = 0, \tag{3}$$

$$d\lambda_1 = (\lambda_1 + \lambda_2)\{(\bar{\pi} - \tau)\vartheta^3 + (\pi - \bar{\tau})\vartheta^4\}, \tag{4}$$

$$d\lambda_2 = (\lambda_1 + \lambda_2)\{-(\rho + \bar{\rho})\vartheta^1 + (\mu + \bar{\mu})\vartheta^2\}, \tag{5}$$

where ϑ^a $(a = 1,2,3,4)$ are one-forms forming a covariant null tetrad in which the metric has the form

$$ds^2 = 2(\vartheta^1\vartheta^2 - \vartheta^3\vartheta^4), \tag{6}$$

and in a local coordinate system (x^i),

$$\begin{aligned} &\vartheta^1 = n_i\, dx^i \quad \vartheta^2 = l_i\, dx^i \\ &\vartheta^3 = -\bar{m}_i\, dx^i \quad \vartheta^4 = -m_i\, dx^i\,. \end{aligned} \tag{7}$$

The basis dual to ϑ^a is denoted by $\{X_a\}$ and the correspondence with the NP operators is given by

$$\begin{aligned} &X_1 = D = l^i \frac{\partial}{\partial x^i}, \quad X_2 = \Delta = \eta^i \frac{\partial}{\partial x^i}, \\ &X_3 = \delta = m^i \frac{\partial}{\partial x^i}, \quad X_4 = \bar{\delta} = \bar{m}^i \frac{\partial}{\partial x^i}. \end{aligned} \tag{8}$$

The differential of a scalar function f is, in this notation,

$$df = (Df)\vartheta^1 + (\Delta f)\vartheta^2 + (\delta f)\vartheta^3 + (\bar{\delta} f)\vartheta^4\,. \tag{9}$$

The integrability conditions of (4) and (5) are simply

$$d^2\lambda_1 = d^2\lambda_2 \equiv 0\,. \tag{10}$$

In the NP formalism, conditions (10) can be written as[8,14]

$$\pi\bar{\pi} = \tau\bar{\tau}\,, \tag{11a}$$

$$\rho\bar{\mu} = \bar{\rho}\mu\,, \tag{11b}$$

$$\begin{aligned} D(\bar{\pi} - \tau) &= (\rho + \bar{\rho})(\bar{\pi} - \tau) \\ &\quad + (\epsilon - \bar{\epsilon} + \bar{\rho})(\bar{\pi} - \tau), \end{aligned} \tag{11c}$$

$$\begin{aligned} \Delta(\bar{\pi} - \tau) &= -(\mu + \bar{\mu})(\bar{\pi} - \tau) \\ &\quad + (\bar{\gamma} - \gamma + \mu)(\bar{\pi} - \tau), \end{aligned} \tag{11d}$$

$$\begin{aligned} \delta(\rho + \bar{\rho}) &= (\rho + \bar{\rho})(\bar{\alpha} + \beta - \bar{\pi}) \\ &\quad - (\rho + \bar{\rho})(\bar{\pi} - \tau)\,, \end{aligned} \tag{11e}$$

$$\begin{aligned} \delta(\mu + \bar{\mu}) &= -(\mu + \bar{\mu})(\bar{\alpha} + \beta - \tau) \\ &\quad - (\mu - \bar{\mu})(\bar{\pi} - \tau)\,, \end{aligned} \tag{11f}$$

$$\begin{aligned} \delta(\pi - \bar{\tau}) - \bar{\delta}(\bar{\pi} - \tau) &= (\bar{\alpha} - \beta)(\bar{\pi} - \tau) \\ &\quad - (\alpha - \bar{\beta})(\bar{\pi} - \tau), \end{aligned} \tag{11g}$$

$$\begin{aligned} D(\mu + \bar{\mu}) + \Delta(\rho + \bar{\rho}) &= (\rho + \bar{\rho})(\gamma + \bar{\gamma}) \\ &\quad - (\mu + \bar{\mu})(\epsilon + \bar{\epsilon})\,. \end{aligned} \tag{11h}$$

The canonical form of the [(11) (11)] Killing tensor is conserved under the transformation

$$l' = e^a l\,, \tag{12a}$$

$$n' = e^{-a} n, \tag{12b}$$

$$m' = e^{ib} m, \tag{12c}$$

where n, l, m, and $\bar{m}$ are the four null vectors forming the covariant null tetrad (6). Relations (11a) and (11b) suggest that we choose our tetrad, imposing the conditions

$$\mu = f\rho, \quad \pi = \tau, \tag{13}$$

where $f = \pm 1$ (we can also make $\pi = -\tau$ without any change in the results). Imposing $\tau = \pi$, we have that $\tau = 0$ implies $\pi = 0$ which is in agreement with (11a), but $\mu = 0$ does not necessarily imply $\rho = 0$ as we can see from (11b). Thus the condition $\mu = f\rho$ does not cover all cases and we have to consider also $\mu = 0, \rho \neq 0$:

$$\mu = f\rho \neq 0, \quad \pi = \tau, \tag{14}$$

$$\mu = 0, \quad \rho \neq 0, \quad \pi = \tau. \tag{15}$$

Before we present the metrical forms for the (1,1) subfamily of HM spaces, we have to mention an important theorem of HM.[8]

Theorem: The null vectors l and n are shear-free geodesic and are therefore principal null vectors of the Weyl tensor.

The proof of this theorem is based on relations (3).

Equations (3)–(5) and (11), combined with the NP equations (14), permit us to obtain the following forms for the metric.[14]

(ia) λ_1 and λ_2 are not constants, $\mu = f\rho \neq 0$, $\pi = \tau$, $(\rho - \bar{\rho})\cdot(\tau + \bar{\tau}) \neq 0$,

$$\begin{aligned} ds^2 = \Omega^2 \Big\{ & f\frac{E^2}{(B-A)^2}(dt + A\, dz)^2 \\ & - \frac{H^2}{(B-A)^2}(dt + B\, dz)^2 \\ & - f\frac{(\psi_y\, dy)^2}{4G^2} - \frac{(\phi_x\, dx)^2}{4F^2} \Big\}. \end{aligned} \tag{16}$$

(ib) λ_1 and λ_2 are not constants, $\mu=0$, $\rho\neq 0$, $\pi=\tau$, $(\rho-\bar{\rho})(\tau+\bar{\tau})\neq 0$,

$$ds^2=\Omega^2\left\{\frac{2\psi_y}{G}\frac{E}{(B-A)}\,dy(dt+A\,dz)-\frac{H^2}{(B-A)^2}(dt+B\,dz)^2-\frac{(\phi_x\,dx)^2}{4F^2}\right\},\quad(17)$$

where

$$\Omega^2=\lambda_1+\lambda_2,\quad \lambda_1=\phi(x),\quad \lambda_2=\psi(y),$$
$$A=A(x,z,t),\quad H=H(x,z,t),\quad B=B(y,z,t),$$
$$E=E(y,z,t),\quad G=G(y),\quad F=F(x),\quad f=\pm 1.$$

The unknown functions A, H, B, and E of both metrics have to also satisfy the following two differential equations on t and z[8,9,14]:

$$\frac{A}{E}\cdot\frac{\partial E}{\partial t}+\frac{\partial A}{\partial t}-\frac{A}{B-A}\frac{\partial(B-A)}{\partial t}=\frac{1}{E}\frac{\partial E}{\partial z}-\frac{1}{B-A}\frac{\partial(B-A)}{\partial z},\quad(18)$$

$$\frac{B}{H}\cdot\frac{\partial H}{\partial t}+\frac{\partial B}{\partial t}-\frac{B}{B-A}\frac{\partial(B-A)}{\partial t}=\frac{1}{B-A}\frac{\partial(B-A)}{\partial z}.\quad(19)$$

These differential equations have been integrated by HM[9] and in a different way by the present author[15] with the same results. Metrics (16) and (17) belong to the (1,1) subfamily of HM spaces.

Metric (16) does not admit, in general, any isometry group. Hauser and Malhiot have proved[8] that a sufficient condition for the existence of a two-parameter, invertible Abelian group with the generators $\partial/\partial t$ and $\partial/\partial z$ is

$$R^{ij}(\nabla_i\lambda_1)(\nabla_j\lambda_2)=0.\quad(20)$$

Condition (20) can be written in the NP formalism as follows:

$$R^{ij}(\nabla_i\lambda_1)(\nabla_j\lambda_2)=-12f(\lambda_1+\lambda_2)^2(\rho+\bar{\rho})\times[f(\phi_{01}-\phi_{10})+\phi_{21}-\phi_{12}]=0,\quad(21)$$

and if we replace the Ricci traceless tensor components by their expressions obtained by the NP equations we have the equivalent relation,

$$R^{ij}(\nabla_i\lambda_1)(\nabla_j\lambda_2)=-12f(\lambda_1+\lambda_2)^2(\rho+\bar{\rho})(\bar{\tau}-\tau)(\bar{\rho}\tau-\rho\bar{\tau})=0.\quad(22)$$

Then, condition (20) reduces to

$$f(\phi_{01}-\phi_{10})+\phi_{21}-\phi_{12}=6f(\bar{\rho}\tau-\rho\bar{\tau})=0,\quad(23)$$

where the assumption that λ_1 and λ_2 are not constants, $(\rho+\bar{\rho})(\tau-\bar{\tau})\neq 0$, has been made.

Relation (23) is the necessary and sufficient condition for the separability of the Schrödinger equation.[8,14]

We have to remark that, by imposing condition (20), Hauser and Malhiot exclude a class of spaces, namely, those spaces which admit the two-parameter invertible Abelian group and are not Schrödinger separable [Carter's metric (77) without condition (79)]. For this reason we have searched and found the sufficient and necessary condition for the existence of the above-mentioned group,[14] a condition that does not imply the separability of the Schrödinger equation,

$$f\phi_{01}=\phi_{21}.\quad(24)$$

The group's orbits may be timelike ($f=1$) or spacelike ($f=-1$) and in the case of electrovac field equations, metric (16) reduces to Carter's metric [$\tilde{A}$].

Metric (17) also does not admit, in general, any group of motion and it belongs to the (1,1) subfamily of HM spaces.[9] The necessary and sufficient condition for the existence of a two-parameter Abelian group is[14]

$$\phi_{21}=0,\quad(25)$$
$$\phi_{01}-\phi_{10}=3(\bar{\rho}\tau-\rho\bar{\tau}).\quad(26)$$

In this case the group is noninvertible but orthogonally transitive with null orbits and the generators $\partial/\partial t$ and $\partial/\partial z$. In the presence of a nonsingular electromagnetic field, metric (17) reduces to a metric found by Leroy[16] and Debever and McLenhaghan.[17]

(ii) $\lambda_1=\text{const}$, $\lambda_2\neq\text{const}$, $\mu=f\rho\neq 0$, $\pi=\tau$, $(\rho-\bar{\rho})\times(\tau+\bar{\tau})\neq 0$,

$$ds^2=f\frac{E^2}{(B-A)^2}(dt+A\,dz)^2-\Omega^2\frac{H^2}{(B-A)^2}\times(dt+B\,dz)^2-fR^2\,dy^2-\Omega^2\Gamma^2\,dx^2,\quad(27)$$

where

$$\Omega^2=\lambda_1+\lambda_2,\quad \lambda_1=\text{const},\quad \lambda_2=\psi(y),$$
$$A=A(x,z,t),\quad H=H(x,z,t),\quad B=B(y,z,t),$$
$$E=E(y,z,t),\quad R=R(y),\quad \Gamma=\Gamma(x),\quad f=\pm 1.$$

(iii) Here $\lambda_1\neq\text{const}$, $\lambda_2=\text{const}$, $\mu=f\rho\neq 0$, $\pi=\tau$, $(\rho-\bar{\rho})(\tau+\bar{\tau})\neq 0$,

$$ds^2=f\Omega^2\frac{E^2}{(B-A)^2}(dt+A\,dz)^2-\frac{H^2}{(B-A)^2}\times(dt+B\,dz)^2-f\Omega^2\Delta^z\,dy^2-S^2\,dx^2,\quad(28)$$

where

$$\Omega^2=\lambda_1+\lambda_2,\quad \lambda_1=\phi(x),\quad \lambda_2=\text{const},$$
$$A=A(x,z,t),\quad H=H(x,z,t),\quad B=B(y,z,t),$$
$$E=E(y,z,t),\quad \Delta=\Delta(y),\quad S=S(x),\quad f=\pm 1.$$

Functions A, H, B, and E have to satisfy the differential equations [Eqs. (18) and (19)] for both cases. These metrics do not admit, in general, any group of motion, but, under certain conditions on the Weyl and the traceless Ricci tensors, they admit a two-parameter (at least) invertible Abelian isometry group, with the generators $\partial/\partial t$ and $\partial/\partial z$ whose orbits are timelike ($f=1$) or spacelike ($f=-1$).[18] In the case of electovac field equations, metrics (27) and (28) reduce to [$\tilde{B}(-)$] and [$\tilde{B}(+)$] Carter spaces, respectively.

In Secs. III–V we are going to consider metric (16) when $f=1$, and metrics (27) and (28) when $f=\pm 1$ in the presence of a perfect fluid source. We have neglected to consider metric (17) because it does not admit a perfect fluid

energy-momentum tensor.[18] Also we have not examined metric (16) when $f=-1$ because the equations are very complicated and we hope to present a solution for $f=-1$ as well as for the rest of the HM subfamilies in the presence of a perfect fluid source.

III. METRICS (16), (27), AND (28) IN THE PRESENCE OF A PERFECT FLUID SOURCE

We consider the Einstein equations with a perfect fluid energy-momentum tensor:

$$R_{ij}-\tfrac{1}{2}Rg_{ij}+\lambda g_{ij}=(e+p)u_iu_j-pg_{ij}, \tag{29}$$

where λ is the cosmological constant, e is the rest energy density, p is the fluid pressure, u is the velocity field of the fluid, and

$$u_i=k_1n_i+k_2l_i-k_3\bar{m}_i-k_4m_i\,. \tag{30}$$

Here, k_1,k_2 are real and k_3,k_4 complex functions of the coordinates

$$k_3=\bar{k}_4=\varphi_1+i\varphi_2; \tag{31}$$

also,

$$u_iu^i=1\Leftrightarrow k_1k_2-k_3k_4=\tfrac{1}{2}\,.$$

We assume that the energy conditions are satisfied:

$$e>0,\quad e+p>0. \tag{32}$$

The Einstein equations (29) give the following expressions for the components of the traceless Ricci tensor in the NP notation:

$$\begin{aligned}
&\phi_{00}=\tfrac{1}{2}(e+p)k_1^2,\quad \phi_{22}=\tfrac{1}{2}(e+p)k_2^2\,,\\
&\phi_{02}=\tfrac{1}{2}(e+p)k_3^2,\quad \phi_{20}=\tfrac{1}{2}(e+p)k_4^2\,,\\
&\phi_{01}=\tfrac{1}{2}(e+p)k_1k_3,\quad \phi_{21}=\tfrac{1}{2}(e+p)k_2k_4\,,\\
&\phi_{10}=\tfrac{1}{2}(e+p)k_1k_4,\quad \phi_{12}=\tfrac{1}{2}(e+p)k_2k_3\,,\\
&\phi_{11}=\tfrac{1}{4}(e+p)(k_1k_2+k_3k_4),\quad \lambda-\tfrac{1}{4}R=\tfrac{1}{4}(e-3p)\,.
\end{aligned} \tag{33}$$

The NP equations and the relations between spin coefficients permit us to distinguish two cases for metrics (16), (27), and (28), which are dependent on whether $f=1$ or $f=-1$.[14,18]

(a) $f=1$, one timelike Killing field and one spacelike,

$$\begin{aligned}
&k_1=k_2,\quad k_3=k_4=\varphi_1,\quad k_1^2-\varphi_1^2=\tfrac{1}{2},\\
&\phi_{00}=\phi_{22}=\tfrac{1}{2}(e+p)k_1^2\,,\\
&\phi_{02}=\phi_{20}=\tfrac{1}{2}(e+p)\varphi_1^2\,,\\
&\phi_{01}=\phi_{21}=\tfrac{1}{2}(e+p)k_1\varphi_1\,,\\
&2\phi_{11}=\phi_{00}+\phi_{02}=\tfrac{1}{2}(e+p)(k_1^2+\varphi_1^2)\,,\\
&\lambda-\tfrac{1}{4}R=\tfrac{1}{4}(e-3p)\,,\\
&\phi_{00}-\phi_{02}>0,\quad \phi_{00}>0,\quad \phi_{02}>0\,.
\end{aligned} \tag{34}$$

(b) $f=-1$, both Killing vectors are spacelike,

$$\begin{aligned}
&k_1=k_2,\quad k_3=-k_4=i\varphi_2,\quad k_1^2-\varphi_2^2=\tfrac{1}{2}\,,\\
&\phi_{00}=\phi_{22}=\tfrac{1}{2}(e+p)k_1^2\,,\\
&\phi_{02}=\phi_{20}=-\tfrac{1}{2}(e+p)\varphi_1^2\,,\\
&\phi_{01}=-\phi_{21}=i\tfrac{1}{2}(e+p)k_1\varphi_1,\\
&2\phi_{11}=\phi_{00}-\phi_{02}=\tfrac{1}{2}(e+p)(k_1^2+\varphi_2^2)\,,\\
&\lambda-R/4=\tfrac{1}{4}(e-3p)\,,\\
&\phi_{00}+\phi_{02}>0,\quad \phi_{00}>0,\quad \phi_{02}<0\,.
\end{aligned} \tag{35}$$

We note that in case (a) the HJ and Schrödinger equations are separable while in case (b) only HJ is separable.

If $f=-1$ and $\phi_{01}\neq 0$ then the Schrödinger equation is not separable for metrics (16), (27), and (28).[14,18] We do not consider here non-Schrödinger separable spaces.

Now we can state the following theorem.[14,18]

Theorem I: The (1,1) subfamily of HM spaces in the presence of a perfect fluid source always admits a two-parameter (at least) Abelian invertible isometry group with $\partial/\partial t$ and $\partial/\partial z$ as generators, and whose orbits may be timelike ($f=1$) or spacelike ($f=-1$).

The proof of this theorem is based on the remark that relations (34) and (35), which express the perfect fluid source presence for metrics (16), (27) and (28), satisfy the necessary and sufficient conditions of the existence of the G_2 group. As an example for metric (16) we have, from (34) and (35),

$$\phi_{01}=\phi_{21},\quad f=1;\quad \phi_{01}=-\phi_{21},\quad f=-1\,.$$

Clearly these relations are the necessary and sufficient conditions for the existence of the group for the same metric [relation (24)].

IV. INTEGRATION OF FIELD EQUATIONS FOR METRICS (16) AND (27)

The fact that l and n are principal null vectors of the Weyl tensor leads us to distinguish two cases for the four-velocity field of the perfect fluid (see Wainwright[19]),

$$u_{[i}\eta_jl_{k]}\neq 0\,, \tag{36}$$

$$u_{[i}\eta_jl_{k]}=0\,. \tag{37}$$

Condition (36) implies that u_i does not lie on the space spanned by l and n while condition (37) implies that it does lie on the above mentioned two-space.

For metric (16) we can state a theorem.

Theorem II: The only perfect fluid solution of Einstein's equations for the (1,1) subfamily of HM spaces, when λ_1 and λ_2 are not constants and $f=1$, is the Wahlquist solution.[14,20]

The metric takes the following form[20]:

$$\begin{aligned}
ds^2=&\frac{E^2}{x^2+y^2}\left[dt+r_0\delta\,(x_A^2-x^2)dz\right]^2\\
&-\frac{H^2}{x^2+y^2}\left[dt+r_0\delta(x_A^2+y^2)dz\right]^2\\
&-(x^2+y^2)\left[\frac{dx^2}{H^2(1+k^2x^2)}+\frac{dy^2}{E^2(1+k^2y^2)}\right],
\end{aligned} \tag{38}$$

where

$$\begin{aligned}
E^2(y)&=1+y^2\\
&\quad+y[y-k^{-1}(1-k^2y^2)^{1/2}\sin^{-1}(ky)]g^{-1}\,,\\
H^2(x)&=1-x^2\\
&\quad-x[x-k^{-1}(1+k^2x^2)^{1/2}\sinh^{-1}(kx)]g^{-1}\,,
\end{aligned}$$

r_0, δ, x_A, k, and g are constants, and

$$u_i=[(2)^{1/2}/2](E^2-H^2)\{E(n_i+l_i)+H(m_i+\bar{m}_i)\}\,.$$

Here, u is the fluid four-velocity and belongs to the family obeying relation (36). Also u is nonexpanding with zero acceleration and shear tensor but with a nonzero two-form.[20] There is no solution with λ_1 and λ_2 nonconstant and $f=1$ and (37) satisfied.

For metric (27) the obtained solutions belong to the (37) family and we have

$$ds^2 = E^2(y)[dt + 2bx\,dz]^2 - \Omega^2\left[\frac{dx^2}{B^2(x)} + B^2(x)dz^2\right] - R^2(x)dy^2 \tag{39}$$

when $f=1$, and

$$ds^2 = dt^2 - E^2(t)[dy + 2bx\,dz]^2 - \Omega^2\left[\frac{dx^2}{B^2(x)} + B^2(x)dz^2\right] \tag{40}$$

when $f=-1$.

In metric (40) we put $y=\hat{t}$, $t=\hat{y}$ and we have dropped the caret, the Killing vectors being in this case $\partial/\partial y$ and $\partial/\partial z$ (spacelike).

For metrics (39) and (40),

$$\phi_{01}=\phi_{21}=0, \quad B^2(x)=c_2X^2+c_1X+c_0.$$

The field equations are solved for special cases:

$$E=R=1. \tag{39'}$$

These are the locally rotationally symmetric (LRS) solutions of type I found by Ellis.[21] Between them we mention the Gödel universe corresponding to $\Omega=1$:

$$b=0, \quad e=\text{const.} \tag{39''}$$

This is the Schwarzschild interior solution.

Metric (41) is the LRS solution of type III[18,21] and the field equations have been solved for the special cases

$$L=\Omega, \tag{40'}$$

which are the Friedman solutions,[21] and

$$b=0, \tag{40''}$$

which are the Kantowski–Sachs solutions.[22]

V. INTEGRATION OF FIELD EQUATIONS FOR METRIC (28)

For this case, the expressions for the traceless Ricci tensor and the energy conditions imply that the metric takes the following form:

$$ds^2 = f\Omega^2E^2(y)[A(x)dt - dz]^2 - H^2(x)dt^2 - f\frac{\Omega^2}{E^2(y)}dy^2 - \frac{dx^2}{H^2(x)}. \tag{41}$$

The only possible cases are

$$u_{[i}n_jl_{k]}\neq 0, \quad f=1, \tag{42}$$

$$u_{[i}n_jl_{k]}=0, \quad f=\pm 1. \tag{43}$$

When (42) holds, then,

$$u_i = k_1(n_i+l_i) - \varphi_1(m_i+\bar{m}_i). \tag{44}$$

The integration of the field equations yields

$$E^2(y) = a\sin(\pm a_0y + a_3) + a_1/a_0^2, \tag{45}$$

where $a=(a_1^2-2a_2^2a_0^2)/a_0$, a_0, a_1, and a_2 are constants;

$$\Omega^2(x) = \exp[b_0x+b_1], \tag{46}$$

$$H^2(x) = -(a_1/b_0^2)\exp[-(b_0x+b_1)] + b_2 + b_3, \tag{47}$$

$$A(x) = \mp(a_0/b_0^2)\exp[-(b_0x+b_1)] + C_0, \tag{48}$$

where b_0, b_1, b_2, b_3, c_0 are constants; and

$$k_1 = \frac{(2)^{1/2}}{2}\frac{4a_0E}{[16E^2a_0^2 - H^2b_0\Omega^2]^{1/2}}, \tag{49}$$

$$\varphi_1 = \frac{(2)^{1/2}}{2}\frac{b_0HE}{[16E^2a_0^2 - H^2b_0^2\Omega^2]^{1/2}}, \tag{50}$$

$$e+3p = \tfrac{1}{4}[-4\lambda + 2b_0b_2], \tag{51}$$

where λ = cosmological constant.

This solution is a generalization of Kramer's solution,[23] which is obtained by making $a_1=0$, $\lambda=0$, and adjusting the remaining parameters.

This same metric has also been found by Kramer independently under different assumptions[24] (the vanishing of the Simon tensor).

The NP coefficients obey the following relations:

$$\mu=\rho, \quad \varepsilon=\gamma, \quad \pi=\tau, \quad \alpha=\beta, \tag{52}$$

$$k=\nu=\sigma=\lambda=\rho+\bar{\rho}=\tau+\bar{\tau}=\alpha+\bar{\alpha}=0. \tag{53}$$

The fluid four-velocity is nonexpanding and is shear-free, its acceleration and the rotation vector are orthogonal to the group orbits. The solution is of Petrov type D.

The Abelian group generated by $\partial/\partial t$ and $\partial/\partial z$ is the maximal group of motion.

The corresponding vacuum case ($e=p=0$) is a subcase of $[\tilde{B}(+)]$ spaces of Carter. Obviously then, this solution cannot be matched to the Kerr solution, which belongs to the $[A]$ family of Carter's spaces, but it could be matched to the $[\tilde{B}(+)]$ spaces.

When (43) holds, we obtain, for $f=1$,

$$ds^2 = x^2[dt - (a-bx^{-3})dz]^2 - H^2(x)dz^2 - x^2\,dy^2 - [dx^2/H^2(x)], \tag{54}$$

where

$$H^2(x) = c_2x_2 + c_1x^{-1} + (b^2/2)x^{-4}. \tag{55}$$

The pressure and the rest energy density are

$$p = \tfrac{3}{4}b^2x^{-6} + 3c_2 - \lambda, \tag{56}$$

$$e = \tfrac{15}{4}b^2x^{-6} - 3c_2 + \lambda, \tag{57}$$

where a, b, c_1, and c_2 are constants.

The equation of state is

$$e-5p=\text{const.} \tag{58}$$

The metric (54) admits a three-parameter Abelian group generated by the vectors $\partial/\partial t$, $\partial/\partial y$, and $\partial/\partial z$.

The corresponding vacuum metric ($e=p=0$) also belongs to the $[\tilde{B}(+)]$ Carter spaces.

If we put $a=0$, and make some variables changes, we obtain a special case of the stationary cylindrically symmetric perfect fluid solution with rigid rotation found by Krasinski.[25]

For spacelike orbits ($f=-1$) we obtain

$$ds^2 = x^2\,dt^2 - x^2[dy + (a - bx^{-3})dz]^2 - H^2(x)dz^2 - dx^2/H^2(x)\,, \tag{59}$$

where we have put $y = \hat{t}$, $t = \hat{y}$, and have dropped the caret in (41). We have

$$H^2(x) = cx + cx - b^2x^{-4}\,, \tag{60}$$

$$p = \tfrac{3}{4}b^2x^{-6} + 3c_2 - \lambda\,, \tag{61}$$

$$e = \tfrac{15}{4}b^2x^{-6} - 3c_2 + \lambda\,. \tag{62}$$

The fluid's four-velocity is nonexpanding, shear-free, and nonrotating but it has a nonzero acceleration. Metrics (54) and (59) are of Petrov type I and the Killing tensor is, for both of them,

$$K_{ij} = x^2(n_il_j + l_in_j) - k^2g_{ij}\,.$$

Metric (59) could represent a cosmological model but it does not depend on the time coordinate and, consequently, it is without physical interest.

This paper makes a generalization of Bonanos' results[7] by finding metric (41), which, although it belongs to the family of spaces examined by him, he did not find since he considered only metrics with $u_{[i}n_jl_{k]} = 0$ (for this classification see also Wainwright[19]). Another kind of generalization of Bonanos' results is coming from the fact that we have also found spaces with spacelike group orbits (with both Killing vectors spacelike). In order to complete this last kind of generalization, we have to examine the Schrödinger nonseparable spaces, an assumption that in the case of a perfect fluid imposes $f = -1$ (space-like orbits) and $\phi_{01} \neq 0$. It will be interesting then to search for spaces that admit the two-parameter invertible Abelian group as the maximal group of motion because it could be used in the study of colliding waves.[26,27]

A possible direction can also be the study of spaces with $(\rho - \bar{\rho})(\tau + \bar{\tau}) = 0$ in the presence of a perfect fluid or with another matter tensor in order to answer the problem posed by Hauser and Malhiot, that being if there is any physically plausible matter tensor permitted by the existence of a [(11) (11)] Killing tensor without any other symmetry.

[1]B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[2]D. Matravers, Util. Math. **2**, 55 (1972).
[3]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[4]L. P. Hughston, P. Sommers, R. Penrose, and M. Walker, Commun. Math. Phys. **27**, 303 (1972).
[5]L. P. Hughston and P. Sommers, Commun. Math. Phys. **32**, 147 (1973).
[6]N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[7]S. Bonanos, Commun. Math. Phys. **49**, 53 (1976).
[8]I. Hauser and R. J. Malhiot, J. Math. Phys. **17**, 1306 (1976).
[9]I. Hauser and R. J. Malhiot, J. Math. Phys. **19**, 187 (1978).
[10]E. Newman and R. Penrose, J. Math. Phys **3**, 566 (1962).
[11]R. Debever, Cah. Phys. **18**, 303 (1964).
[12]R. Debever and R. G. McLenaghan, and N. Tarig, Gen. Relativ. Gravit. **10**, 853 (1979).
[13]R. Debever, R. G. McLenaghan, J. Math. Phys. **22**, 1711 (1981).
[14]T. Papacostas, Bull. Cl. Sci. Acad. R. Belg. **LXIX**, 495 (1983).
[15]T. Papacostas, Bull. Cl. Sci. Acad. R. Belg. **LXIX**, 641 (1983).
[16]J. Leroy, Bull. Cl. Sci. Acad. R. Belg. **KXIV**, 130 (1978).
[17]R. Debever and R. G. McLenaghan, Bull. Cl. Sci. Acad. R. Belg. **LXVI**, 204 (1980).
[18]T. Papacostas, Ph. D. thesis, Universite Libre Bruxelles (1981).
[19]J. Wainwright, Gen. Relativ. Gravit. **8**, 797 (1977).
[20]H. D. Wahlquist, Phys. Rev. **1972**, 1291 (1968).
[21]G. F. Ellis, J. Math. Phys. **8**, 1171 (1966).
[22]R. Kantowski and R. K. Sachs, J. Math. Phys. **7**, 443 (1966).
[23]D. Kramer, Class. Quantum Gravit. **1**, L3 (1984).
[24]D. Kramer, preprint, FSU Jena, Sektion Phys., 1985.
[25]D. Kramer, H. Stephani, M. MacCallum, and E. Hertl, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980), p. 223.
[26]S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London A **402**, 37, 1970 (1985).
[27]P. Szekeres, Nature (London) **228**, 1183 (1962).

Conformally Ricci-flat perfect fluids. II

N. Van den Bergh
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada

(Received 8 December 1987; accepted for publication 10 February 1988)

A review is given of the perfect fluid solutions which can be constructed by conformal transformations of nonflat vacuum solutions of Einstein's field equations. In addition a proof is given that (a) under physically reasonable restrictions all solutions for which the gradient of the scalar field is orthogonal to the fluid velocity and (b) all solutions sharing with the corresponding vacuum solution a symmetry group of dimension $n \geqslant 2$ are necessarily shear- or vorticity-free and hence are explicitly known.

I. INTRODUCTION AND REVIEW OF KNOWN RESULTS

Exact solutions for inhomogeneous perfect fluid solutions of Einstein's field equations are important for a better understanding of several aspects of cosmology and astrophysics. Only relatively few such models are known, and very often restrictions of a purely geometrical nature are used to construct new classes of solutions. In view of the successful construction of all conformally flat perfect fluids,[1] one can wonder whether one could take, more generally, one of the vacuum models as a "seed solution" and conformally transform it to a perfect fluid solution. Some simple questions one would like to see answered are then, for example, the following.

(a) Can the Schwarzschild solution be transformed by $\phi(r,t)$ to yield an expanding fluid sphere?

(b) Can the stationary axisymmetric vacuum models be transformed by $\phi(\rho,z)$ to rigidly or differentially rotating perfect fluids?

(c) Can spatially homogeneous cosmological vacuum models be transformed by $\phi(t)$ to spatially homogeneous perfect fluids?

(d) Can *arbitrary* vacuum solutions always be transformed into perfect fluid solutions by a suitable choice of ϕ?

So far, the investigation of these—and related—questions has resulted in the following conclusions[2–4]: A necessary and sufficient condition in order that the metric g_{ab} of a perfect fluid model [with energy-momentum tensor $(w+p)u_a u_b + p g_{ab}$] should be conformally related to a vacuum metric $\hat{g}_{ab} = e^{2\phi} g_{ab}$ is the existence of a solution ϕ of

$$\phi_{a;b} - \phi_{,a}\phi_{,b} + \tfrac{1}{2} g_{ab}\phi_{,e}\phi^{,e} = [(w+p)/2]u_a u_b + \tfrac{1}{6} w g_{ab} \,. \tag{1.1}$$

One may suppose $w + p \neq 0$, as otherwise[5] g_{ab} is the de Sitter or anti-de Sitter metric or a vacuum *pp* wave.

When $\nabla\phi$ is aligned with the fluid velocity $\mathbf{u}$, the resulting space-times are conformally flat, and are just the FRW models.[4] When the fluid is shear-free and $\nabla\phi$ is orthogonal to $\mathbf{u}$, the resulting models are locally rotationally symmetric. They are either rotating class I or nonrotating class IIc solutions in Stewart and Ellis' classification[6] and have a (non-γ-law) equation of state $p = p(w)$. When the fluid is shear-free and $\nabla\phi$ is not orthogonal to $\mathbf{u}$, the models are conformally flat. They are either the expanding Stephani universes or generalizations of the interior Schwarzschild solution.[1]

When none of the previous conditions hold and when the fluid is vorticity-free with $\nabla\phi$ orthogonal to $\mathbf{u}$, the only solution[4] is a nonstatic model admitting a three-dimensional isometry group of Bianchi type VI_0 acting on timelike hypersurfaces. The fluid has an equation of state $p = p(w)$. When $\nabla\phi$ is not orthogonal to $\mathbf{u}$, solutions[4] are (pseudo-) spherically or plane symmetric [and hence the answer to question (a) above is affirmative], but an equation of state does not exist. It also has been shown that the answer to question (d) is negative: Vacuum solutions of Petrov type *N cannot* be conformally transformed to perfect fluid solutions.

Questions (b) and (c), and related ones, in which one assumes that space-times $(M,\hat{g})$ and (M,g) share some isometry group G_n $(n \geqslant 2)$, have remained unanswered so far. In Sec. II it will be shown that, when $\nabla\phi$ is not orthogonal to $\mathbf{u}$, the resulting perfect fluid models are always vorticity-free. When $\nabla\phi$ is orthogonal to $\mathbf{u}$ no immediate answer can be given. However, when one assumes the existence of an equation of state, the orthogonal case can be completely solved *without* making any assumptions about the existence of an isometry group: In Sec. III, a proof is given that all such solutions are shear- or vorticity-free and hence belong to the known cases discussed above. As a consequence the answer to questions (b) and (c) is negative, too.

In contrast with the investigation of the general cases, where the Newman–Penrose formalism was used, the situations where $\nabla\phi$ is orthogonal to $\mathbf{u}$, or where Killing vectors are present, are most easily discussed within the orthonormal tetrad formalism,[7] by choosing a frame such that $\mathbf{e}_0 = \mathbf{u}$ and $\partial_2\phi = \partial_3\phi = 0$. With $F = \partial_1\phi$ and $G = \partial_0\phi$ (and hence $F \neq 0$, as a consequence of the first result mentioned above) one obtains for the frame components of (1.1) the following:

$$2\,\partial_0 G - 2F\dot{u}_1 - F^2 - G^2 = \tfrac{1}{3}(2w + 3p) \,, \tag{1.2}$$

$$\partial_1 G = F(\theta_1 + G) \,, \tag{1.3}$$

$$\partial_2 G = F(\sigma_{12} + \omega_3) \,, \tag{1.4}$$

$$\partial_3 G = F(\sigma_{13} - \omega_2) \,, \tag{1.5}$$

$$2\,\partial_1 F - 2G\theta_1 - F^2 - G^2 = \tfrac{1}{3}w \,, \tag{1.6}$$

$$\partial_0 F = G(\dot{u}_1 + F) \,, \tag{1.7}$$

$$\partial_2 F = G(\sigma_{12} + \omega_3) \,, \tag{1.8}$$

$$\partial_3 F = G(\sigma_{13} - \omega_2) \,, \tag{1.9}$$

and

 0022-2488/88/061451-04$02.50

$$-2G\theta_2 - 2F(n_{23}+a_1) + F^2 - G^2 = \tfrac{1}{3}w\,, \tag{1.10}$$

$$-2G\theta_3 + 2F(n_{23}-a_1) + F^2 - G^2 = \tfrac{1}{3}w\,, \tag{1.11}$$

$$G\dot{u}_2 - F\Omega_3 = 0\,, \tag{1.12}$$

$$G\dot{u}_3 + F\Omega_2 = 0\,, \tag{1.13}$$

$$G(\sigma_{12}-\omega_3) + F(n_{13}-a_2) = 0\,, \tag{1.14}$$

$$G(\sigma_{13}+\omega_2) - F(n_{12}+a_3) = 0\,, \tag{1.15}$$

$$2G\omega_1 - Fn_{11} = 0\,, \tag{1.16}$$

$$2G\sigma_{23} + F(n_{33}-n_{22}) = 0\,. \tag{1.17}$$

II. $(M,\hat{g})$ AND (M,g) SHARING A GROUP OF ISOMETRIES $G_{n>2}$

When one supposes that space-times $(M,\hat{g})$ and (M,g) admit at least a two-dimensional group G_2 of isometries, with generators **K** and **L**, such that

$$\pounds_{\mathbf{K}}\hat{g} = \pounds_{\mathbf{K}}g = \pounds_{\mathbf{L}}\hat{g} = \pounds_{\mathbf{L}}g = 0\,, \tag{2.1}$$

then obviously also

$$\pounds_{\mathbf{K}}\phi = \pounds_{\mathbf{L}}\phi = 0\,, \tag{2.2}$$

i.e., the conformal transformation "preserves" the symmetries. Taking now the directional derivative of (2.2) along $\nabla\phi$, one obtains

$$(K^a\phi_{,a})_{;b}\phi^{,b} = 0\,, \tag{2.3}$$

or, with the aid of Killing's equations,

$$K^a\phi^{,b}\phi_{,a;b} = 0\,. \tag{2.4}$$

Substitution of (1.1) in the latter equation results in

$$(K^a u_a)(\phi_{,b}u^b) = 0\,, \tag{2.5}$$

whereas for **L** one obtains similarly

$$(L^a u_a)(\phi_{,b}u^b) = 0\,. \tag{2.6}$$

The case $\phi_{,b}u^b = 0$ will be dealt with in Sec. III in its full generality. So now assume $\phi_{,b}u^b \neq 0$ (hence F and $G \neq 0$), such that, by (2.5) and (2.6), **K** and **L** are orthogonal to **u**:

$$K^a u_a = L^a u_a = 0\,. \tag{2.7}$$

From Killing's equations one also obtains then

$$K^a\dot{u}_a = L^a\dot{u}_a = 0\,, \tag{2.8}$$

whereas (2.2) implies that **K** and **L** lie in the $(\mathbf{e}_2,\mathbf{e}_3)$ plane.

First notice that $G \neq 0$ always implies $\omega_1 = 0$: Indeed (1.4), (1.5) and (1.8), (1.9) show that $\partial_2(F^2-G^2) = \partial_3(F^2-G^2) = 0$ and hence that $[\partial_2,\partial_3](F^2-G^2) = 0$. With the aid of (1.16) this can be written as $G\omega_1(w+p) = 0$ or $\omega_1 = 0$. Unless the fluid is vorticity-free, the frame can then be *invariantly* fixed by requiring $\omega_2 = 0$. By (2.2) and (2.7) the operators ∂_2 and ∂_3 will then yield identically 0 when applied to any invariantly defined scalar.

In particular one will have $\partial_2 F = \partial_3 F = \partial_2 G = \partial_3 G = 0$, as the scalar-field ϕ for a nonflat CRF solution is uniquely defined up to a constant factor.[8] By (1.4) one then has

$$\sigma_{12} = -\omega_3\,, \tag{2.9}$$

such that (1.14) yields

$$-2\omega_3 G + (n_{13}-a_2)F = 0\,. \tag{2.10}$$

Acting now with $[\partial_1,\partial_2]$ on F^2-G^2 gives

$$-2\omega_3\,\partial_0(F^2-G^2) + (n_{13}-a_2)\partial_1(F^2-G^2) = 0\,, \tag{2.11}$$

which, together with (2.10), shows that $\omega_3 G(w+p) = 0$ and hence that $\omega_3 = 0$, a contradiction. As a consequence we have shown that all CRF perfect fluids sharing at least a group G_2 with their corresponding vacuum solution, are vorticity-free provided $\nabla\phi$ is not orthogonal to **u**. Together with the result of Sec. III, one obtains that they are *precisely* given by the classes of shear-free or vorticity-free solutions discussed in the Introduction, provided the fluids have an equation of state $p = p(w)$.

III. CRF PERFECT FLUIDS WITH $\nabla\phi$ ORTHOGONAL TO u

When $\phi_{,a}u^a = 0$ (i.e., $G = 0$) strong restrictions on the rotation coefficients can be obtained, provided one assumes the existence of an equation of state $p = p(w)$. From (1.2)–(1.17) one immediately obtains the following relations:

$$\partial_1 F = \tfrac{1}{2}F^2 - w/6\,, \tag{3.1}$$

$$\partial_0 F = \partial_2 F = \partial_3 F = 0\,, \tag{3.2}$$

and

$$\theta_1 = n_{11} = n_{23} = \Omega_2 = \Omega_3 = 0, \quad \sigma_{12} = -\omega_3\,, \quad \sigma_{13} = \omega_2, \quad n_{12} = -a_3, \quad n_{13} = a_2, \quad n_{33} = n_{22}, \tag{3.3}$$

with p and w determined by

$$w = 3F^2 - 6Fa_1\,, \tag{3.4}$$

$$p = -3F^2 + 4Fa_1 - 2F\dot{u}_1\,. \tag{3.5}$$

The relevant Jacobi equations and field equations are given in the Appendix.

Acting now with the commutators $[\partial_0,\partial_1]$, $[\partial_1,\partial_2]$, and $[\partial_3,\partial_1]$ on F yields with the aid of (3.1) and (3.2):

$$\partial_0 w = \partial_2 w = \partial_3 w = 0\,. \tag{3.6}$$

When $w+p \neq 0$ this implies that the resulting fluids are nonexpanding:

$$\theta_3 = -\theta_2\,. \tag{3.7}$$

Also, when $p = p(w)$ one has

$$\dot{u}_2 = \dot{u}_3 = 0\,, \tag{3.8}$$

whereas (3.4) and (3.5) show that $\dot{u}_1$ and a_1 satisfy relations similar to (3.2) or (3.6).

Consider first the case $\omega_2 = \omega_3 = 0$: Under a rotation in the $(\mathbf{e}_2,\mathbf{e}_3)$ plane, $\mathbf{e}_2' = \cos\alpha\mathbf{e}_2 - \sin\alpha\mathbf{e}_3$, $\mathbf{e}_3' = \sin\alpha\mathbf{e}_2 + \cos\alpha\mathbf{e}_3$, one has

$$\sigma_{23}' = \sin 2\alpha\theta_2 + \cos 2\alpha\sigma_{23}\,, \tag{3.9}$$

and hence $(\mathbf{e}_\alpha)$ can be chosen to be a shear eigenframe. From the Jacobi equations (A2) and (A3) one then obtains $n_{22}\theta_2 = 0$, which implies that solutions are either shear-free (and hence are explicitly known[2]), or that $n_{22} = 0$ and $\theta_2 \neq 0$. In the latter case, however, the field equation (A5) implies $\Omega_1 = 0$, which, after substitution in (A2), shows that $\omega_1(\dot{u}_1 + a_1) = 0$. As $\dot{u}_1 + a_1 = 0$ would imply with

(3.4) and (3.5) that $w+p=0$, one has $\omega_1=0$, i.e., the only possible solution is the temporally homogeneous vorticity-free solution discussed earlier.[4]

Assuming henceforth that $\omega_2^2+\omega_3^2\neq 0$, the frame can be invariantly fixed by choosing

$$\omega_2=0, \quad \omega_3\neq 0 . \tag{3.10}$$

Then (A1) implies $s_{23}=\omega_1-\Omega_1$, and, for the remaining variables F, $\dot{u}_1$, a_1, a_2, a_3, ω_1, ω_3, n_{22}, θ_2, and Ω_1, Jacobi equations and field equations can be combined to yield the following:

$$\partial_0 F=\partial_2 F=\partial_3 F=0 , \tag{3.11}$$

$$\partial_1 F=F(F-a_1) , \tag{3.12}$$

$$\partial_0\dot{u}_1=\partial_2\dot{u}_1=\partial_3\dot{u}_1=0 , \tag{3.13}$$

$$\partial_1\dot{u}_1=-\dot{u}_1^2+2a_1\dot{u}_1+3F(a_1-F-\dot{u}_1)+2(\theta_2^2+\Omega_1^2-2\Omega_1\omega_1) , \tag{3.14}$$

$$\partial_0 a_1=\partial_2 a_1=\partial_3 a_1=0 , \tag{3.15}$$

$$\partial_1 a_1=(a_1-F)(a_1+\dot{u}_1)+\theta_2^2+\Omega_1^2-2\Omega_1\omega_1 , \tag{3.16}$$

$$\partial_0\omega_1=0 , \tag{3.17}$$

$$\partial_0\omega_3=-\theta_2\omega_3 , \tag{3.18}$$

$$\partial_2\omega_3=2\omega_3 a_2 , \tag{3.19}$$

$$\partial_0 n_{22}=4\omega_1(\dot{u}_1+a_1)+2n_{22}\theta_2+8\omega_3 a_3 , \tag{3.20}$$

$$\partial_0\theta_2=2\Omega_1(\omega_1-\Omega_1) , \tag{3.21}$$

$$\partial_1\theta_2=2n_{22}(\omega_1-\Omega_1)-\dot{u}_1\theta_2-4\omega_3 a_2 , \tag{3.22}$$

$$\partial_0\Omega_1=2\theta_2\Omega_1 , \tag{3.23}$$

$$\partial_1\Omega_1=2(\dot{u}_1+a_1)\omega_1+2n_{22}\theta_2+4\omega_3 a_3-\dot{u}_1\Omega_1 , \tag{3.24}$$

and

$$\partial_1\omega_1+\partial_3\omega_3=\omega_1(\dot{u}_1+2a_1)+2\omega_3 a_3 , \tag{3.25}$$

$$\partial_2\Omega_1-2\partial_0 a_3-2\theta_2 a_3-4\Omega_1 a_2=0 , \tag{3.26}$$

$$2\partial_0 a_2+\partial_2\theta_2=4a_3(2\omega_1-\Omega_1)-4(\dot{u}+a_1)\omega_3+6\theta_2 a_2 . \tag{3.27}$$

The essential step is now the careful calculation of some of the resulting integrability conditions. Acting first with $[\partial_0,\partial_1]$ on Ω_1 one obtains with the aid of (3.11)–(3.24) and (A6):

$$2\theta_2[\omega_1(a_1+\dot{u}_1)+a_3\omega_3]-\omega_3[\partial_2\Omega_1-2\partial_0 a_3-4\Omega_1 a_2]=0 , \tag{3.28}$$

which, with (3.26), results in $\theta_2\omega_1(a_1+\dot{u}_1)=0$ or, as $a_1+\dot{u}_1=0$ would imply $w+p=0$,

$$\theta_2\omega_1=0 . \tag{3.29}$$

Two more integrability conditions are required: acting with $[\partial_0,\partial_2]$ on ω_3 and with $[\partial_0,\partial_1]$ on θ_2 yields, by (A6), (A7), (3.26), and (3.27):

$$\Omega_1\partial_3\omega_3=2\omega_3[(2\omega_1-\Omega_1)a_3+2\theta_2 a_2-\omega_3(\dot{u}_1+a_1)] \tag{3.30}$$

and

$$\Omega_1\partial_1\omega_1=4\omega_3[(\Omega_1-\omega_1)a_3+(\dot{u}_1+a_1)\omega_3]+2\omega_1^2(a_1+\dot{u}_1)-\Omega_1\omega_1\dot{u}_1-4\theta_2\omega_3 a_2 . \tag{3.31}$$

Together with (3.28) the latter two equations result in the algebraic restriction

$$[\omega_3^2+\omega_1(\omega_1-\Omega_1)](\dot{u}_1+a_1)=0 , \tag{3.32}$$

or, as $\dot{u}_1+a_1=0$ would imply $w+p=0$,

$$\omega_3^2+\omega_1(\omega_1-\Omega_1)=0 . \tag{3.33}$$

It is clear now that the case $\theta_2\neq 0$ admits no solutions: then, in fact, by (3.29), $\omega_1=0$ and hence, by (3.33), $\omega_3=0$. On the other hand, when $\theta_2=0$, (3.21) implies that $\Omega_1(\omega_1-\Omega_1)=0$ and hence $\Omega_1=0$ [$\omega_1-\Omega_1=0$ would give us by (3.33) $\omega_3=0$]. Then, however, (3.33) leads to $\omega_3^2+\omega_1^2=0$, which is in contradiction with our assumption $\omega_3\neq 0$.

Herewith we have obtained the result that all CRF perfect fluids having an equation of state $p=p(w)$ and having $\nabla\phi$ orthogonal to the fluid velocity $\mathbf{u}$ are either shear-free or vorticity-free, and hence belong to explicitly known classes of solutions.[2,4]

ACKNOWLEDGMENTS

All calculations were done with the aid of the orthonormal tetrad package ORTHOFRAME[9] within the symbolic calculation system MAPLE,[10] developed by the Symbolic Computation Group of the University of Waterloo. I thank H. Warren for her excellent typing work.

This work was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada under Grant No. A3978.

APPENDIX: EQUATIONS USED IN SEC. III

The Jacobi equations relevant for the first part of Sec. III are the following:

$$\partial_0\omega_2-\omega_3(\Omega_1-\omega_1+\sigma_{23})-\theta_2\omega_2=0 , \tag{A1}$$

$$\partial_0 n_{22}-\partial_1(\sigma_{23}+\Omega_1)+\partial_2\omega_2-\partial_3\omega_3-\dot{u}_1(\sigma_{23}+\Omega_1+2\omega_1)-2\omega_1 a_1-2\omega_2 a_2-6\omega_3 a_3-2n_{22}\theta_2=0 , \tag{A2}$$

$$\partial_0 n_{22}+\partial_1(\sigma_{23}-\Omega_1)-\partial_2\omega_2+\partial_3\omega_3+\dot{u}_1(\sigma_{23}-\Omega_1-2\omega_1)-2\omega_1 a_1-2\omega_3 a_3-6\omega_2 a_2+2n_{22}\theta_2=0 , \tag{A3}$$

$$\partial_2\omega_3-\partial_3\omega_2-2\omega_3 a_2+2\omega_2 a_3=0 , \tag{A4}$$

together with the (23) field equation

$$\partial_0\sigma_{23}+2\theta_2\Omega_1=0 . \tag{A5}$$

The commutator relations relevant for the system (3.13)–(3.31) are

$$[\partial_0,\partial_1]=\dot{u}_1\partial_0+2\omega_3\partial_2 \tag{A6}$$

and

$$[\partial_0,\partial_2]=-\theta_2\partial_2+2\Omega_1\partial_3 . \tag{A7}$$

[1]D. Kramer, H. Stephani, M. A. H. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[2]N. Van den Bergh, J. Math. Phys. **27**, 1076 (1986).
[3]N. Van den Bergh, Lett. Math. Phys. **11**, 141 (1986).
[4]N. Van den Bergh, Gen. Relativ. Gravit. **18**, 649 (1986).
[5]H. W. Brinkman, Math. Ann. **91**, 269 (1924).
[6]J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).
[7]M. A. H. MacCallum, *Cosmological Models from a Geometric Point of View in Cargèse Lectures in Physics*, Vol. 6, edited by E. Schatzman (Gordon and Breach, New York, 1973).
[8]C. D. Collinson and E. G. L. R. Vaz, Gen. Relativ. Gravit. **14**, 5 (1982).
[9]N. Van den Bergh, "ORTHOFRAME, a MAPLE package for performing symbolic calculations in the Ellis–MacCallum formalism," University of Waterloo preprint, 1987.
[10]B. W. Char, K. O. Geddes, G. H. Gonnet, and S. M. Watt, *MAPLE User's Guide* (WATCOM, Waterloo, Ontario, 1985).

Real anomalies

J. Lott
Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109

(Received 7 January 1988; accepted for publication 17 February 1988)

The relationship between global anomalies of quantum theory and the topology of spaces of real Fredholm operators is shown. The spectral properties of such operators and how they are seen in examples of global anomalies on both compact and noncompact space-times are discussed.

I. INTRODUCTION

It has become clear that many of the anomalies of quantum field theory are due to the nontrivial topology of various moduli spaces, such as the space of connections on a fixed vector bundle modulo the group of gauge transformations.[1] More abstractly, one can view the moduli space as parametrizing a family of Dirac-type operators, and so one is using the particular family of operators in order to view the space $\mathscr{F}$ of all complex Fredholm operators.

If one captures a nontrivial cohomology class of $\mathscr{F}$ by means of a family of operators then this may prevent one from defining the renormalized determinant of the operators in a nice way. To be more precise, in free fermionic path integrals there are two types of determinants which arise. In the Lagrangian the relevant differential operator (the inverse of the covariance) may either map one function space to itself, or to another.

In the second case the determinant can be complex and its anomalous symmetry properties reflect the topology of $\mathscr{F}$. In the first case the determinant is always real. One can view the underlying function space as a real vector space and because fermion fields anticommute, the differential operator must be real and skew adjoint. It turns out that the space $\mathscr{F}_1 R$ of real skew-adjoint Fredholm operators has a very rich topology[2] and we wish to show that much of this topology can be seen in quantum field theories (QFT's). This is manifested both in the existence of zero eigenvalues for Dirac-type operators and in the occurrence of global anomalies, the original example of which is Witten's SU(2) anomaly.[3]

When one rotates fermions from Minkowski space to Euclidean space, one may seem to lose special properties, such as the existence of Majorana or Weyl representations. In Euclidean space, these special Minkowski properties are seen in the existence of operators which anticommute with the Euclidean Dirac operator. In general, one can consider the spaces $\mathscr{F}_k R$ which consist of the elements of $\mathscr{F}_1 R$ which anticommute with a Clifford algebra of operators. These spaces have a topology which is different but related to that of $\mathscr{F}_1 R$. We also give examples of how this refined structure is seen in QFT's.

The structure of this paper is as follows. In Sec. I, we review the topology of some spaces of Fredholm operators. In Sec. III, we discuss how this topology is seen in terms of the spectra of such operators. In Sec. IV, we give examples of QFT's on compact space-times which see the topology of the $\mathscr{F}_k R$'s. These examples are more-or-less known, but we hope that it may help to see them in a unified way, and that the derivations of the indices may be new. In Sec. V, we give some new examples of global anomalies on noncompact space-times. These examples are analogs of the Gell-Mann–Lévy σ model[4] and show that the existence of a global anomaly does not necessarily ruin consistency of a QFT. In Sec. VI, we sketch how the anomalies involving real operators can be understood in terms of analogs of the determinant line bundle of Quillen.[5]

Notation: $\{\sigma^j\}_{j=1}^3$ will denote the Pauli matrices:

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

$\{\gamma_M^j\}_{j=0}^3$ will denote the $(-+++)$ real Dirac matrices: $\gamma_M^0 = I \otimes i\tau^2$, $\gamma_M^1 = I \otimes \tau^3$, $\gamma_M^2 = \sigma^1 \otimes \tau^1$, and $\gamma_M^3 = \sigma^3 \otimes \tau^1$, and γ_M^5 will denote $\gamma_M^0 \gamma_M^1 \gamma_M^2 \gamma_M^3$, satisfying $(\gamma_M^5)^2 = -I$, $(\gamma_M^5)^T = -\gamma_M^5$. $\{\gamma_E^j\}_{j=0}^3$ will denote $(++++)$ complex Dirac matrices satisfying $\gamma_E^i \gamma_E^j + \gamma_E^j \gamma_E^i = 2\delta^{ij}$, and γ_E^5 will denote $\gamma_E^0 \gamma_E^1 \gamma_E^2 \gamma_E^3$, satisfying $(\gamma_E^5)^2 = I$, $(\gamma_E^5)^\dagger = \gamma_E^5$. A handy reference for Clifford algebra structures is Ref. 6.

II. REVIEW OF TOPOLOGY OF OPERATOR SPACES

Let H be a complex Hilbert space and consider the space of Fredholm operators

$$\mathscr{F} = \{T \in B(\mathscr{H}) : \dim \ker T < \infty \text{ and } \dim \ker T^* < \infty\}.$$

(If one wishes to consider unbounded Fredholm operators one can generally modify the function spaces to reduce the bounded case.) Put $\mathscr{F}_1 = \{T \in \mathscr{F} : T^* = -T$ and the essential spectrum of iT intersects both components of $\mathbf{R} - \{0\}\}$. One has that $\mathscr{F}$ is a classifying space for complex K theory, i.e., for all compact topological spaces X, the Grothendieck group $K(X)$ of virtual vector bundles over X satisfies $K(X) \cong [X, \mathscr{F}]$, where $[X, \mathscr{F}]$ denotes the homotopy classes of maps from X to $\mathscr{F}$.[7] The relationship is as follows: over $\mathscr{F}$ one has the virtual vector bundle Index with fiber $[\text{Ker } T] - [\text{Coker } T]$ over an element $T \in \mathscr{F}$. Then any element of $K(X)$ can be written as ϕ^* Index for some $\phi \in [X, \mathscr{F}]$. As a consequence, $\mathscr{F}$ has the homotopy type of $\mathbf{Z} \times BU(\infty)$ where the $\mathbf{Z}$ factor refers to the ordinary index of an operator and $BU(\infty)$ is the classifying space for the group $U(\infty)$. By Bott periodicity, $\pi_{k+2}(\mathscr{F}) = \pi_k(\mathscr{F})$ and these homotopy groups $\pi_l(\mathscr{F}) \cong K(S^l)$ are listed in Table I.

Similarly, $\mathscr{F}_1$ is a classifying space for K^{-1}, i.e., $K^{-1}(X) \cong [X, \mathscr{F}_1]$. Then $\pi_{k+2}(\mathscr{F}_1) = \pi_k(\mathscr{F}_1)$ and the

 0022-2488/88/061455-10$02.50

TABLE I. Homotopy groups of complex operator spaces.

	$\mathbb{F}$	$\mathbb{F}_1$
π_0	$\mathbf{Z}$	0
π_1	0	$\mathbf{Z}$

homotopy group are listed in Table I. The relationship between $\mathscr{F}$ and $\mathscr{F}_1$ can be seen as a suspension.[8] Let $\Omega\mathscr{F}$ denote the paths in $\mathscr{F}$ from I to $-I$. Then there is a map ϕ: $\mathscr{F}_1 \to \Omega\mathscr{F}$ given by $\phi(T) = \{\cos \pi t + T \sin \pi t\colon 0 \leqslant t \leqslant 1\}$, which can be shown to a homotopy equivalence. Similarly, let J be an operator unitarily equivalent to $\left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$ and put

$$\mathscr{F}_2 = \{T \in \mathscr{F}_1\colon TJ + JT = 0\}.$$

Let $\Omega\mathscr{F}_1$ denote the paths in $\mathscr{F}_1$ from J to $-J$. Then there is a homotopy equivalence ϕ_1: $\mathscr{F}_2 \to \Omega\mathscr{F}$, given by $\phi_1(T) = \{J \cos \pi t + T \sin \pi t, 0 \leqslant t \leqslant 1\}$. Because $\mathscr{F}_2$ is isomorphic to $\mathscr{F}$, one has $\mathscr{F} \sim \mathscr{F}_2 \sim \Omega\mathscr{F}_1 \sim \Omega^2\mathscr{F}$, which shows the Bott periodicity.

It is now easy to state the relationship between the axial anomaly of QFT and the topology of $\mathscr{F}$. Consider, for example, the space $\mathscr{A}$ of connections on $S^4 \times \mathrm{SU}(N)$, $N > 2$, and the group $\mathscr{G}$ of gauge transformations which are the identity at a point ∞ on S^4. Then the determinant line bundle $\Lambda^{\max}$ Index has first Chern class which is a nontrivial element of $H^2(\mathscr{F}, \mathbf{R}) = \pi_2(\mathscr{F}) \otimes \mathbf{R} = \mathbf{R}$, and which is pulled back via the Dirac operator to give a nontrivial element of $H^2(\mathscr{A}/\mathscr{G}, \mathbf{R}) = H^2(\Omega^3(\mathrm{SU}(N)), \mathbf{R}) = \mathbf{R}$.[1] (More precisely, under $\not{\partial}\colon \mathscr{A} \to \mathscr{F}$, the pullback $\not{\partial}^* \Lambda^{\max}$ Index is a $\mathscr{G}$-equivariant line bundle over $\mathscr{A}$ which pushes forward to a line bundle on $\mathscr{A}/\mathscr{G}$.) To see this another way, fix $A_0 \in \mathscr{A}$ such that $\not{\partial}_{A_0}$ is invertible. Put

$$\mathscr{K} = \{T \in B(H)\colon T - I \text{ is compact and } T \text{ is invertible}\}.$$

Then there is a map ρ: $\mathscr{G} \to \mathscr{K}$ given by $\rho(g) = \not{\partial}_{A_0}^{-1} \not{\partial}_{g \cdot A_0}$. Now $\mathscr{K}$ is homotopically equivalent to $\mathrm{U}(\infty)$, or $\mathscr{F}_1$, and $\rho^*\colon H^1(\mathscr{K}, \mathbf{R}) \to H^1(\mathscr{G}, \mathbf{R})$ is nontrivial from $\mathbf{R}$ to $\mathbf{R}$. This is a precise form of the intuitive idea that the phase of the chiral determinant changes by a nontrivial multiple of 2π when going around a nontrivial loop in $\mathscr{G}$. Finally, from the Hamiltonian viewpoint consider the analogous spaces for $S^3 \times \mathrm{SU}(N)$. The Dirac Hamiltonian $\not{\partial}_A$: $\Gamma(S) \to \Gamma(S)$ is skew adjoint and gives a map σ: $\mathscr{A}/\mathscr{G} \to \mathscr{F}_1$. The nontriviality of $\sigma^* H^3(\mathscr{F}_1, \mathbf{R}) \in H^3(\mathscr{A}/\mathscr{G}, \mathbf{R})$ leads to a Hamiltonian interpretation of the axial anomaly.[9]

Let us now consider the space $\mathscr{F}_0\mathbf{R}$ of real Fredholm operators on a real Hilbert space $\mathscr{H}_{\mathbf{R}}$. For a compact topological space X with involution τ, let $KR(X)$ denote the Grothendieck group of virtual complex vector bundles over X with an antilinear involution covering τ.[10] [If τ is the identity then $KR(X) = KO(X)$.] One has $KR(X) \cong [X, \mathscr{F}_0 R]$. It follows that $\mathscr{F}_0 R$ is homotopically equivalent to $\mathbf{Z} \times BO(\infty)$ and $\pi_{k+8}(\mathscr{F}_0 R) = \pi_k(\mathscr{F}_0 R)$. The homotopy groups are listed in Table II.

In order to get the higher KR functors, let C_k denote the real Clifford algebra generated by $\{e_i\}_{i=1}^k$ with relations $e_i e_j + e_j e_i = -2\delta_{ij}$, $e_i^* = -e_i$. Let ρ: $C_k \to B(H_{\mathbf{R}})$ be a

TABLE II. Homotopy groups of real operator spaces.

	$\mathbb{F}_0 R$	$\mathbb{F}_1 R$	$\mathbb{F}_2$	$\mathbb{F}_3 R$	$\mathbb{F}_4 R$	$\mathbb{F}_5 R$	$\mathbb{F}_6 R$	$\mathbb{F}_7 R$
π_0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0	0	0
π_1	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0	0	0	$\mathbf{Z}$
π_2	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$
π_3	0	$\mathbf{Z}$	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$
π_4	$\mathbf{Z}$	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0
π_5	0	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0	$\mathbf{Z}$
π_6	0	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0
π_7	0	$\mathbf{Z}$	$\mathbf{Z}_2$	$\mathbf{Z}_2$	0	$\mathbf{Z}$	0	0

faithful * representation of C_k. Put

$$\mathscr{G}_k = \{T \in \mathscr{F}_0 R\colon T = -T^*,\ T\rho(e_i) + \rho(e_i)T = 0 \text{ for } 1 \leqslant i \leqslant k-1\}.$$

For $k \equiv -1 \pmod 4$ put $\mathscr{F}_k R = \mathscr{G}_k$. For $k \equiv -1 \pmod 4$ there is a slight subtlety: put $\mathscr{F}_k R = \{T \in \mathscr{G}_k$: the essential spectrum of $\rho(e_1) \cdots \rho(e_{k-1})T$ intersects both components of $\mathbf{R} - 0\}$. Then $KR^{-k}(X) \cong [X, \mathscr{F}_k R]$.[2] As a consequence, the homotopy groups are those listed in Table II. The various $\mathscr{F}_k R$'s are again linked by suspension maps: put $\Omega\mathscr{F}_{k-1} R = \{$paths in $\mathscr{F}_{k-1} R$ from $\rho(e_{k-1})$ to $-\rho(e_{k-1})\}$. Then ϕ: $\mathscr{F}_k R \to \Omega\mathscr{F}_{k-1} R$ is given by

$$\phi(T) = \{\rho(e_{k-1}) \cos \pi t + T \sin \pi t,\ 0 \leqslant t \leqslant 1\}.$$

The various spaces $\mathscr{F}_k R$ have simple interpretations.

$k=1$: $\mathscr{F}_1 R$ is the space of real skew-adjoint Fredholm operators on $H_{\mathbf{R}}$.

$k=2$: Because $\mathscr{F}_2 R$ consists of the elements of $\mathscr{F}_1 R$ which anticommute with $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, they all have the form $\left(\begin{smallmatrix} A & B \\ B & -A \end{smallmatrix}\right)$ with A and B real and skew adjoint. Then $(v, w) \to (Av + Bw, Bv - Aw)$ and $v + iw \to (A + iB)(v - iw)$, showing that $\mathscr{F}_2 R$ is the space of skew-adjoint antilinear Fredholm operators on a complex Hilbert space. Note that

$$\begin{pmatrix} V \\ W \end{pmatrix}^T \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \begin{pmatrix} V \\ W \end{pmatrix} = \mathrm{Re}(v - iw)^\dagger (A + iB)(v + iw),$$

showing that $\mathscr{F}_2 R$ can also be thought of as skew-symmetric Fredholm operators from a complex Hilbert space to its complex conjugate. Finally, because

$$\begin{pmatrix} V \\ W \end{pmatrix}^T \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \begin{pmatrix} V \\ W \end{pmatrix} = \mathrm{Re}(v + iw)^T (A + iB)(v + iw),$$

these operators arise when writing complex Berezin integrals (i.e., Berezin integrals whose total integral is a complex Pfaffian).

$k=3$: $\mathscr{F}_3 R$ consists of the underlying real Fredholm operators coming from skew-adjoint operators on $H_{\mathbf{R}} \otimes \mathbf{C}^2$ of the form $i\alpha_0 + \sigma^1\alpha_1 + \sigma^2\alpha_2 + \sigma^3\alpha_3$, with $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in B(H_{\mathbf{R}})$, which satisfy the essential spectrum property. This anticommutes with the operators $\rho(e_1)$ and $\rho(e_2)$ given by $\rho(e_1)x = \sigma_2 \bar{x}$, $\rho(e_2)x = i\sigma_2 \bar{x}$: the complex structure comes from $\rho(e_1)\rho(e_2)$.

$k=4$: $\mathscr{F}_4 R$ consists of Fredholm operators of the form $\left(\begin{smallmatrix} 0 & B \\ -B^T & 0 \end{smallmatrix}\right)$ acting on $(H_{\mathbf{R}} \otimes \mathscr{H}) \oplus (H_{\mathbf{R}} \otimes \mathscr{H})$, where

$B \in B(H_{\mathbf{R}} \otimes \mathscr{H})$ commutes with the quaternions $\mathscr{H}$. The operators $\rho(e_1)$, $\rho(e_2)$, and $\rho(e_3)$ are $i \otimes \sigma_3$, $j \otimes \sigma_3$, and $k \otimes \sigma_3$.

$k=5$: $\mathscr{F}_5 R$ consists of Fredholm operators of the form $\left(\begin{smallmatrix} 0 & B \\ B0 & \end{smallmatrix}\right)$ acting on $H_{\mathbf{R}} \otimes (\mathscr{H} \oplus \mathscr{H})$ where $B \in B(H_{\mathbf{R}} \otimes \mathscr{H})$ is skew and commutes with $\mathscr{H}$. The operators $\rho(e_1)$, $\rho(e_2)$, $\rho(e_3)$, and $\rho(e_4)$ are $1 \otimes i\sigma_2$, $i \otimes \sigma_3$, $j \otimes \sigma_3$, and $k \otimes \sigma_3$.

$k=6$: For $M \in M(2, \mathscr{H})$, let $L(M)$ denote left multiplication by M on $\mathscr{H} \oplus \mathscr{H}$ and for $q \in \mathscr{H}$, let $R(q)$ denote right multiplication by q on $\mathscr{H} \oplus \mathscr{H}$. Then $\mathscr{F}_6 R$ consists of Fredholm operators of the form $B_1 R(j) + B_2 R(k)$ acting on $(H_{\mathbf{R}} \otimes \mathscr{H}) \oplus (H_{\mathbf{R}} \otimes \mathscr{H})$, with B_1 and B_2 being self-adjoint operators in $B(\mathscr{H}_{\mathbf{R}})$. The operators $\rho(e_1), \ldots, \rho(e_5)$ are $L\left(\begin{smallmatrix} 0 & i \\ -i & 0 \end{smallmatrix}\right) R(i)$, $L\left(\begin{smallmatrix} 0 & j \\ -j & 0 \end{smallmatrix}\right) R(i)$, $L\left(\begin{smallmatrix} 0 & k \\ -k & 0 \end{smallmatrix}\right) R(i)$, $L\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) R(i)$, and $L\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) R(i)$.

$k=7$: Because $C_6 = M(8, \mathbf{R})$, $\mathscr{F}_1 R$ consists of $\{\rho(e_1) \cdots \rho(e_6) T: T \in B(\mathscr{H}_{\mathbf{R}}), T^* = T, T$ is Fredholm and the essential spectrum of T lies on both sides of $\mathbf{R} - \{0\}\}$.

$k=8$: Let $J_1, \ldots, J_6$ denote a representation of the generators of C_6 on $\mathbf{R}^8$ and put $\epsilon = J_1, \ldots, J_6$. Then $\mathscr{F}_8 R$ consists of Fredholm operators of the form $\left(\begin{smallmatrix} A & \epsilon B \\ \epsilon B & -A \end{smallmatrix}\right)$ acting on $H_{\mathbf{R}} \otimes (\mathbf{R}^8 \oplus \mathbf{R}^8)$ with $A \in B(H_{\mathbf{R}})$ skew and $B \in B(H_{\mathbf{R}})$ self-adjoint. The operators $\rho(e_1), \ldots, \rho(e_7)$ are $J_1 \otimes \sigma_1, \ldots, J_6 \otimes \sigma_1$ and $I \otimes i\sigma_2$. The Bott periodicity is seen in the fact that $\mathscr{F}_8 R$ is isomorphic to $\mathscr{F}_0 R$.

III. SPECTRAL PROPERTIES OF REAL INDEX THEORY

We will be interested in the π_0 and π_1 homotopy groups of operator spaces. First, for the complex Fredholm operators $\pi_0(\mathscr{F}) = \mathbf{Z}$ shows that $\mathscr{F}$ breaks into connected components labeled by the index of an operator. That $\pi_1(\mathscr{F}_1)$ equals $\mathbf{Z}$ can be seen using spectral flow. Given a smooth map: $S^1 \to \mathscr{F}_1$, we have that the spectrum of $i\Phi(e^{2\pi i\epsilon})$ is uniformly bounded away from zero as ϵ varies in $[0,1]$, with the possible exception of a finite number of eigenvalues. Because the spectrum for $\epsilon = 1$ is the same as that for $\epsilon = 0$, the generic circle of operators will have a finite number of eigenvalues which flow from negative to positive when going from $\epsilon = 0$ to $\epsilon = 1$; this number defines the spectral flow F: $[S^1, \mathscr{F}^1] \to \mathbf{Z}$. If the operators $i\Phi(e^{2\pi i\epsilon})$ are actually self-adjoint first-order elliptic differential operators acting on cross sections of a vector bundle E over a compact manifold M, one can compute $F(\Phi)$ by means of the eta invariant.[11] Given such an operator H, define

$$\eta(H) = \lim_{s \to 0} \sum_{\lambda_i \neq 0} \lambda_i |\lambda_i|^{-s-1}.$$

If $H(\epsilon)$ is a one-parameter family of such operators then $\eta(H(\epsilon))$ can have integer jumps as eigenvalues cross the origin, but $(d/d\epsilon)\eta(\mathscr{H}(\epsilon))$ has a smooth extension which can be computed in terms of local quantities. Then

$$0 = 2F(\Phi) + \int_0^1 \frac{d}{d\epsilon} \eta(i\Phi(e^{2\pi i\epsilon}))\, d\epsilon \qquad (1)$$

gives an effective way to compute $F(\Phi)$. One can also compute F as an index by means of a "desuspension." Consider the operator $D = \partial/\partial\epsilon + i\Phi(e^{2\pi i\epsilon})$ acting on cross sections of the pullback of E to $S^1 \times M$. Then $F(\Phi) = \text{Index } D$.[11] Similarly, for some Fredholm operators on a noncompact complete manifold, one can define a generalized eta invariant[12] and the spectral flow is again given by (1); however, the expression $(d/d\epsilon)\eta(H(\epsilon))$ then depends both on local quantities and on the behavior of $H(\epsilon)$ at infinity.

For an operator $T \in \mathscr{F}_k R$, one has that $\ker T$ is a C_{k-1} module.[2] If it is not a C_k module then T represents a nontrivial element of $\pi_0(\mathscr{F}_k R)$. Thus the connected components of $\mathscr{F}_k R$ are labeled as follows:

$k = 0$: Index T, $k = 1$: $\dim \ker T \pmod 2$,

$k = 2$: $\frac{1}{2} \dim \ker T \pmod 2$,

$k = 4$: $\frac{1}{4} \dim \ker T = \text{Index}_{\mathscr{H}} B$.

In order to see $\pi_1(\mathscr{F}_k R)$ spectrally, consider first the case $k = 1$. Then one has a one-parameter family $iT(\epsilon)$ of self-adjoint operators, each of which has spectrum symmetric around the origin. As ϵ ranges from 0 to 1 the spectral flow of $iT(\epsilon)$ is zero because of the symmetry, but a finite number of pairs of eigenvalues can be switched. This number (mod 2) then labels the class of $\pi_1(\mathscr{F}_1 R) = \mathbf{Z}_2$ in which $T(\epsilon)$ lies. [Because the switching can be seen by watching what happens near the origin of the spectrum, the definition makes sense even if the operators $iT(\epsilon)$ have continuous spectrum.]

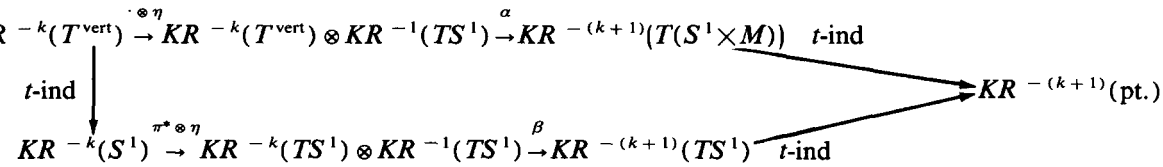
For the case $k = 3$, let $T(\epsilon)$ be a one-parameter family in $\mathscr{F}_3 R$. Viewing $T(\epsilon)$ as a complex operator as in Sec. II, one sees that if x is an eigenvector of $iT(\epsilon)$ with real eigenvalue λ then $\sigma_2 \bar{x}$ is also an eigenvector with an eigenvalue λ. Thus there is an action of the complex Clifford algebra $C_1^{\mathbf{C}}$ on the discrete eigenspaces of $iT(\epsilon)$ given by $x \to \sigma_2 \bar{x}$ and so the eigenspaces have even complex dimension (one cannot solve $\sigma_2 \bar{x} = \alpha x$ with $\alpha \in \mathbf{C}$). The class of $T(\epsilon)$ in $\pi_1(\mathscr{F}_3 R) = \mathbf{Z}$ is labeled by $\frac{1}{2}$ of the spectral flow of $T(\epsilon)$.

Finally, for $k = 7$ the operators are self-adjoint and $\pi_1(\mathscr{F}_7 R)$ is labeled by the spectral flow.

For real first-order differential operators there is a desuspension which maps $\pi_1(\mathscr{F}_k R)$ to $\pi_0(\mathscr{F}_{k+1} R)$. If $T(\epsilon)$ is a one-parameter family of operators in $\mathscr{F}_k R$ then formally $\left(\begin{smallmatrix} \partial_\epsilon & T(\epsilon) \\ T(\epsilon) & -\partial_\epsilon \end{smallmatrix}\right)$ is in $\mathscr{F}_{k+1} R$, as a differential operator on $S^1 \times M$. However, there is a slight subtlety, since to obtain the isomorphism between $\pi_1(\mathscr{F}_k R)$ and $\pi_0(\mathscr{F}_{k+1} R)$ one must also twist the bundles over S^1 by the Hopf bundle $\mathbf{H}$, the flat $\mathbf{R}$ bundle over S^1 with holonomy -1. To be more precise, we state the following.

Proposition 1: Let $T(\epsilon)$ be a circle of elliptic first-order real differential operators acting on $\Gamma(E)$, with $\begin{smallmatrix} E \\ \downarrow \\ M \end{smallmatrix}$ being a real vector bundle over a compact manifold M. Suppose that each $T(\epsilon)$ is in $\mathscr{F}_k R$. Let Φ_1: $S^1 \times M \to S^1$ and Φ_2: $S^1 \times M \to M$ be the projection maps. Consider the first-order operator D acting on $\Gamma(\Phi_1^* \mathbf{H} \otimes (\Phi_2^* E \oplus \Phi_2^* E))$ given locally by $D = \left(\begin{smallmatrix} \partial_\epsilon & T(\epsilon) \\ T(\epsilon) & -\partial_\epsilon \end{smallmatrix}\right)$. Then under the isomorphism $KR^{-k}(S^1) \to KR^{-(k+1)}(pt)$, the topological index of the family $T(\epsilon)$ is mapped to the topological index of D.

Proof: Let T^{vert} denote the vertical directions in $T(S^1 \times M)$, i.e., $T^{\text{vert}} = S^1 \times TM$. Let η be a fixed element of $KR^{-1}(TS^1)$. Consider the diagram



where the maps are as follows: t-ind is the topological index[13] which generally maps $KR^*(P\times TX)$ to $KR^*(P)$, where P and X are compact manifolds. $\pi^*: KR^*(S^1)\to KR^*(TS^1)$ is the map induced from the projection $\pi: TS^1\to S^1$. α is multiplication in $KR^*(T(S^1\times M))$ and β is multiplication in $KR^*(TS^1)$. The multiplicative property of t-ind ensures that that diagram commutes. Thus the only problem is to find η such that $t\text{-ind}\circ\beta\circ(\pi^*\otimes\eta)$ is the identity map from $KR^{-k}(S^1)$ to $KR^{-(k+1)}(\text{pt.})$. It is easily checked that this η is the symbol of the operator ∂_ϵ acting on $\Gamma(\mathbf{H})$, which proves the proposition. ■

One can easily generalize the Proposition to the case of a fibration over S^1. In a special case, the element of $\pi_1(\mathscr{F}_1 R)$ represented by a circle of real skew-adjoint operators can be computed by means of spectral flow. Suppose that for all $e^{2\pi i\epsilon}\in S^1$, $T(\epsilon)$ commutes with a fixed $J\in B(H_{\mathbf{R}})$ satisfying $J^2=-I, J^*=-J$. Then J provides a complex structure on $H_{\mathbf{R}}$ and we can write $T(\epsilon)$ as $\begin{pmatrix} A(\epsilon) & B(\epsilon) \\ -B(\epsilon) & A(\epsilon)\end{pmatrix}$. Over the complexes this is equivalent to $\begin{pmatrix} A+iB(\epsilon) & 0 \\ 0 & (A-iB)(\epsilon)\end{pmatrix}$ and for each eigenvector $x\in H_{\mathbf{R}}\otimes C$ of $(A+iB)(\epsilon)$ with eigenvalue $i\lambda$, there is a corresponding eigenvector $\bar{x}$ of $(A-iB)(\epsilon)$ with eigenvalue $-i\lambda$. It follows that each eigenvalue $i\lambda$ of $A+iB$ gives a pair $(i\lambda,-i\lambda)$ of eigenvalues of T, and the spectral flow of $i(A+iB)$ equals the number of eigenvalue rearrangements of $T(\text{mod }2)$. Thus the class in $\pi_1(\mathscr{F}_1 R)=\mathbf{Z}_2$ represented by $T(\epsilon)$ is labeled by the spectral flow of $i(A+iB)(\epsilon)$ (mod 2).

IV. QFT's ON COMPACT SPACES

The topology of real operator spaces arises in two distinct ways in QFT. The first way uses the π_0 invariant to ensure zero eigenvalues for some differential operator T. The physical interpretation of such a zero eigenvalue depends on whether the operator arises from a Lagrangian or a Hamiltonian. If T enters in a Euclidean fermionic Lagrangian in the form $\langle\Psi,T\Psi\rangle$ then a zero eigenvalue can prevent tunnelling between different "θ vacua."[14,15] On the other hand, if T gives the spatial Hamiltonian for a fermionic theory then there are degenerate ground states arising from ker T.[16]

The second way uses the π_1 invariant to label global anomalies. This means that there may be an obstruction to defining a renormalized determinant function for a family of operators. If one is dealing with a circle of operators then it is possible that when one attempts to define the the determinant smoothly along the loop, the spectral properties of the operator force the putative determinant to change sign when going around the circle. (For a more precise interpretation, see Sec. VI.)

Our examples all involve Dirac-type operators. Because the Clifford algebra structure depends strongly on the dimension of the manifold, we will list examples by dimension and restrict to the case of perturbatively renormalizable field theories, i.e., dim$\leqslant 4$. Of course, there are mathematically interested examples in all dimensions. In general, one has that on a k-dimensional manifold, the real Dirac operator (involving only the metric) lies in $\mathscr{F}_k R$.[17]

A. One dimension

Let M be an oriented Riemannian manifold with $\pi_1(M)=0$, let $\gamma: S^1\to M$ be a smooth path in M, and let $(\gamma,\Psi): S^1\to TM$ cover γ. The Lagrangian for $N=\frac{1}{2}$ supersymmetric geodesic motion is

$$L(\gamma,\Psi)=\frac{1}{2}\int_{S^1}[|\dot{\gamma}|^2-\langle\Psi,\nabla_{\dot{\gamma}}\Psi\rangle]\,dT,$$

where the Ψ fields are formally anticommuting. Upon doing a formal integration over Ψ in the functional integral $\int e^{-L}\mathscr{D}\gamma\mathscr{D}\Psi$, one is left with $\int e^{-(1/2)\int_{S^1}|\dot{\gamma}|^2\,dT}\times(\det{}^{1/2}\nabla_{\dot{\gamma}})\mathscr{D}\gamma$. If one tries to define $\det^{1/2}\nabla_{\dot{\gamma}}$ by a regularized product of the positive eigenvalues of $i\nabla_{\dot{\gamma}}$ then the obstruction to a smooth definition is the possibility of an odd number of eigenvalue rearrangements of $i\nabla_{\dot{\gamma}}$ when going around a loop of γ's, i.e., the possibility of a map $S^1\to[S^1,M]$ giving a nontrivial element of $\pi_1(\mathscr{F}_1 R)$. In this example one can compute $|\det^{1/2}\nabla_{\dot{\gamma}}|$ explicitly and see whether there is a smooth definition of $\det^{1/2}\nabla_{\dot{\gamma}}$,[18] but one can also see this via Proposition 1.

Proposition 2: There is a loop in Map (S^1,M) whose image is nontrivial in $\pi_1(\mathscr{F}_1 R)$ iff M is not spin.

Proof: Let $\gamma: T^2\to M$ be a loop in Map(S^1,M). Because M is oriented, γ^*TM is an SO(N) bundle over T^2. Let A be the pullback of the Riemannian connection on TM to γ^*TM. Let S be the flat $\mathbf{R}$ bundle on $T^2=S^1\times S^1$ with the holonomy -1 on the first S^1 factor. By Proposition 1, the element of $\pi_1(\mathscr{F}_1 R)$ given by γ is the same as the element of $\pi_0(\mathscr{F}_2 R)$ given by $\begin{pmatrix}\partial_0 & \partial_1+A_1 \\ -\partial_1+A_1 & -\partial_0\end{pmatrix}$ acting on $\Gamma(E)$, $E=(\gamma^*TM\oplus\gamma^*TM)\otimes S$. Because the index of D in $\pi_0(\mathscr{F}_2 R)$ is a homotopy invariant, it only depends on the topological class of the real vector bundle E. For $n>2$ the SO(n) bundles on T^2 are classified by $H^2(T^2,\mathbf{Z}_2)=\mathbf{Z}_2$, which can be considered to be the element of $\pi_1(\text{SO}(n))$ used in gluing the ends of $S^1\times I$ to construct a bundle over T^2.

Let V denote a nontrivial $\mathbf{R}^3$ bundle over T_2. Now γ^*TM is classified as a real bundle by $\gamma^*w_2(M)$, where $w_2(M)\in\mathscr{H}^2(M,\mathbf{Z}_2)$ is the second Stiefel–Whitney class, and so we can instead compute the index of $\tilde{D}=\begin{pmatrix}\partial_0 & D_1 \\ D_1 & -\partial_0\end{pmatrix}$ acting on either $\Gamma((T^2\times\mathbf{R}^{\dim M})\otimes S)$ if $\gamma^*w_2(M)$ is trivial, or $\Gamma(((T^2\times\mathbf{R}^{\dim M-3})\oplus V)\otimes S)$ if $\gamma^*w_2(M)$ is nontrivial. However, this is computed to be $\gamma^*w_2(M)[T^2]\in\mathbf{Z}_2$, the evaluation of $\gamma^*w_2(M)$ on T^2. As one can pick up a nontrivial

$w_2(M)$ by some mapping of T^2, it follows that the index of $\nabla_{\dot\gamma}$ in $\pi_1(\mathscr{F}_1 R)$ is zero for all γ: $T^2 \to M$ iff $w_2(M) = 0$, i.e., M is spin. One has the same story for $n = 2$. ■

Under canonical quantization one sees that the Hamiltonian corresponding to $L(\gamma,\Psi)$ is $\frac{1}{2}\not{D}^2_M$,[18] which makes sense iff M is spin. Thus in this case a global anomaly causes nonexistence of the quantum theory.

B. Two dimensions

As one has Majorana–Weyl spinors in two-dimensional Minkowski spaces, one can consider the fermionic Lagrangian $L(\Psi) = \int \overline{\Psi}\partial_+\Psi$, where ∂_+ maps S_+ to S_-. The total integral $\int e^{iL(\Psi)}\mathscr{D}\Psi$ is formally $\det^{1/2}\partial_+$, which Wick rotates to $\det^{1/2}\partial_{\bar z}$. Now on a compact two-dimensional Riemannian spin manifold one only has Majorana spinors, and the real Dirac operator can be written as $\not{D} = \sigma^1 D_0 + \sigma^3 D_1$. The Minkowski–Weyl property can be seen in the fact that $\not{D}$ lies in $\mathscr{F}_2 R$, as it anticommutes with $i\sigma_2$. Then $D_0 + iD_i$ is skew symmetric and one can form the complex Berezin integral $\int\mathscr{D}\Psi \exp -\int\Psi^T(D_0 + iD_1)\Psi$, with total formal integral $\det^{1/2}\partial_{\bar z}$.

The class of $\not{D}$ in $\pi_0(\mathscr{F}_2 R)$ is labeled by $\frac{1}{2}\dim_{\mathbf{R}}\ker\not{D}$ (mod 2). On a Riemann surface of genus g there are $2^{g-1}(2^g + 1)$ spin structures for which it is nontrivial and $2^{g-1}(2^g - 1)$ for which it is trivial.[19]

C. Three dimensions

Let A be a real gauge field and consider the Minkowski–Majorana action

$$L = \int \overline{\Psi}(i\sigma^2 D_0 + \sigma^1 D_1 + \sigma^3 D_2)\Psi\, d^3x.$$

After integrating out the fermions in $\int e^{iL(\Psi)}\mathscr{D}\Psi$ one is left with $\det^{1/2}\not{D}_A$.

As there are no Euclidean Majorana spinors in three dimensions, let S be the complex spinor bundle over S^3, let E be an $\mathbf{R}^N$ vector bundle over S^3 with connection A, and consider the Euclidean Lagrangian

$$L = \int_{S^3}\Psi^\dagger(\sigma^1 D_1 + \sigma^2 D_2 + \sigma^3 D_3)\Psi, \text{ for } \Psi\in\Gamma(S\otimes E).$$

The Minkowski–Majorana property is seen in the fact that $\not{D}_A = \sigma^1 D_1 + \sigma^2 D_2 + \sigma^3 D_3$ lies in $\mathscr{F}_3 R$, which implies that all eigenspaces of $i\not{D}_A$ are even dimensional. We may try to define the formal integral $\int e^{-L(\Psi)}\mathscr{D}\Psi = \det^{1/2}\not{D}_A$ by multiplying the eigenvalues of $i\not{D}_A$ with half-multiplicity. This will only be well-defined when going around a circle of operators if one-half of the spectral flow around the circle is even, i.e., if the circle is trivial in $\pi_1(\mathscr{F}_3 R)$ (mod 2).

Proposition 3: Let $A(\epsilon)$ be a one-parameter family of connections on E, $0\leqslant\epsilon\leqslant 1$, with $A(1) = g\cdot A(0)$ for a gauge transformation g: $S^3\to SO(N)$. Then the class of $\not{D}_{A(\epsilon)}$ in $\pi_1(\mathscr{F}_3 R) = \mathbf{Z}$ is $g^*\omega[S^3]$, where $\omega\in H^3(SO(N),\mathbf{Z})$ is given by the three-form $\omega = (-1/48\pi^2)\mathrm{Tr}(g^{-1}\,dg)^3$. This can be odd for some choice of $A(\epsilon)$ iff $N > 3$.

Proof: Let L be the $\mathbf{R}^N$ bundle over $S^1\times S^3$ formed from the trivial bundle over $I\times S^3$ by identifying the fiber over $\{1\}\times S^3$ with the g-twisted fiber over $\{0\}\times S^3$, and by then tensoring with the pullback of $\mathbf{H}$ to $S^1\times S^3$. Let T denote the real Dirac operator on $S^1\times S^3$ twisted by L and let $\tilde T$ denote the complex Dirac operator on $S^1\times S^3$ from $S^+\otimes L$ to $S^-\otimes L$. By Proposition 1, the class of $\not{D}_{A(\epsilon)}$ in $\pi_1(\mathscr{F}_3 R)$ equals the class of T in $\pi_0(\mathscr{F}_4 R)$, which is one-half of the index of $\tilde T$. Now the family of connections $A(\epsilon)$ give a connection B on the trivial bundle over $I\times S^3$ by $B_0 = 0$, $B_i(\epsilon,X) = A_i(\epsilon)(X)$, which extends to a connection on L. We can homotopy B to $B_0 = 0$, $B_i(\epsilon,X) = \epsilon(g^{-1}\,dg)_i$ without changing the index. Then the index of $\tilde T$ is given by

$$\int_{S^1\times S^3}\mathrm{Tr}\, e^{F(B)/2\pi i}$$
$$= -\frac{1}{8\pi^2}\int_{S^1\times S^3}\mathrm{Tr}(d\epsilon\wedge g^{-1}\,dg + (\epsilon-\epsilon^2)(g^{-1}\,dg)^2)^2 = -\frac{1}{24\pi^2}\int_{S^3}\mathrm{Tr}(g^{-1}\,dg)^3,$$

so the class in $\pi_0(\mathscr{F}_4 R)$ is

$$-\frac{1}{48\pi^2}\int_{S^3}\mathrm{Tr}(g^{-1}\,dg)^3.$$

One can check that for $N = 3$, the pullback of ω from SO(3) to SU(2) is twice the generator of $\mathscr{H}(SU(2),\mathbf{Z})$. Since every map from S^3 to SO(3) factors through SU(2), it follows the evaluation of $g^*\omega$ on S^3 on S^3 is always even. On the other hand, for $N = 4$ the pullback of ω from SO(4) to SU(2) via $SU(2)\to SU(2)\times SU(2)\to SO(4)$ gives the generator of $\mathscr{H}^3(SU(2),\mathbf{Z})$. As one can embed SO(4) in SO(N) for $N\geqslant 4$, it follows that $g^*\omega[S^3]$ can always be odd for some g if $N\geqslant 4$. ■

One could also compute this invariant by computing one-half of the spectral flow directly. This is perhaps more physical, as for fixed ϵ there will be a term in the Euclidean effective action equal to $\pm\frac{1}{4}i\pi\eta(i\not{D}_{A(\epsilon)})$.[20,21]

D. Four dimensions

In four-dimensional Minkowski space we have massive or massless Majorana spinors, or massless Weyl spinors, but not both simultaneously. To see how this is reflected in the Euclidean action, consider the real Euclidean Dirac operator $\not{D} = \Sigma_{\mu=0}^3\gamma^\mu D_\mu$ with $\gamma^0 = \gamma^0_M\otimes i\tau^2$ and $\gamma^j = \gamma^j_M\otimes I$. As D anticommutes with the operators $\rho(e_1) = \gamma^0_M\otimes\tau^1$, $\rho(e_2) = \gamma^0_M\otimes\tau^3$, and $\rho(e_3) = \gamma^5_M\otimes I$, it lies in $\mathscr{F}_4 R$ and gives a quaternionic operator. The natural way to form a massive Dirac operator is by $\not{D}_m = \not{D}1 + m\rho(e_3)$, which lies in $\mathscr{F}_3 R$, as it anticommutes with $\rho(e_1)$ and $\rho(e_2)$. Using the complex structure provided by $\rho(e_1)\rho(e_2)$, one can see that $\not{D}_m$ is the underlying real operator for the complex skew-adjoint operator $T_m = i\gamma^5_E(\Sigma_{\mu=0}^3\gamma^\mu_E D_\mu + m)$. One can then use the action $L(\Psi) = \int\Psi^\dagger T_m\Psi$ to form a complex Berezin integral $\int\mathscr{D}\Psi\, e^{-L(\Psi)}$ to describe massive Euclidean Dirac spinors. This Berezin integral satisfies reflection positivity and the reconstructed Hilbert space is the Fock space of the massive Minkowski Dirac spinor, with the standard second-quantized Dirac Hamiltonian. Although this way of handling Euclidean Dirac spinors may be unconventional, one can see, for example, that the total Berezin integral is formally $\det i\gamma^5_E(\Sigma_{\mu=0}^3\gamma^\mu_E D_\mu + m)$, which formally equals

$\det(\Sigma_{\mu=0}^{3}\gamma_E^{\mu}D_\mu + m)$, the result obtained from the usual field-doubling method.[22]

In order to deal with Euclidean Majorana spinors, one must use the symmetries of $\not{D}_m$. Because $\not{D}_m$ is $\mathscr{F}_3R$, there is an operator A satisfying $A^T = -A$ and $AT_m^* + T_mA = 0$. Then $(A^{-1}T_m)^T = -A^{-1}T_m$ and one can use $A^{-1}T_m$ to form the action $\tilde{L}(\Psi) = \int \Psi^T A^{-1}T_m\Psi$ for complex four-component Ψ. In terms of the charge conjugation operator, $\tilde{L}$ can be written in the following way. In a given representation $\{\gamma_\mu^E\}_{\mu=0}^3$ of the Dirac algebra, let C satisfy $C\gamma^\mu C^{-1} = -\gamma^{\mu T}$, $C^T = -C$, $C^\dagger = C^{-1} = -C^*$. Then the charge conjugation operator $\Psi \to \Psi^c = C^{-1}\Psi^*$ is intrinsically defined and one can write $\tilde{L}(\Psi)$ as $\int(\Psi^c)^\dagger(\Sigma_{\mu=0}^3\gamma^\mu D_\mu + m)\Psi$, with $C(\Sigma_{\mu=0}^3\gamma^\mu D_\mu + m)$ a skew-symmetric operator. One can form the complex Berezin integral $\int e^{-\tilde{L}(\Psi)}\mathscr{D}\Psi$ whose total integral is the complex Pfaffian $\det^{1/2} C(\Sigma_{\mu=0}^3\gamma^\mu\mathscr{D}_\mu + m)$, the $\frac{1}{2}$ reflecting that one is dealing with Majorana spinors. This gives the same way to handle Euclidean Majorana spinors as was probably by Nicolai.[23]

The Weyl property is seen in the fact that $\not{D}$ anticommutes with the self-adjoint operator $\rho(e_1)\rho(e_2)\rho(e_3)$. Writing $\not{D}$ as $\begin{pmatrix} 0 & B \\ -B^T & 0 \end{pmatrix}$, the quaternionic operator B is the Euclidean equivalent of the chiral Minkowski Dirac operator ∂ and $\det^{1/2}\not{D} \cong \det B$ is the Wick rotation of $\det\not{\partial}$. One can couple an $O(N)$ gauge field A to $\not{D}$ to obtain an operator $\not{D}_A$ in $\mathscr{F}_4R$, but one can go further and use the quaternionic nature of $\not{D}$ to naturally couple an $\mathrm{Sp}(N)$ gauge field V. Let us write V as $V^{(1)} + V^{(i)}i + V^{(j)}j + V^{(k)}k$ with $V^{(1)}$ skew symmetric and $V^{(i)}$, $V^{(j)}$, and $V^{(k)}$ symmetric. Then

$$\not{D}_V = \sum_{\mu=0}^{3}\gamma^\mu(\partial_\mu + V_\mu^{(1)} + V_\mu^{(i)}i + V_\mu^{(j)}j + V_\mu^{(k)}k)$$

lies in $\mathscr{F}_1R$ and anticommutes with $\rho(e_1)\rho(e_2)\rho(e_3)$.

The class of $\not{D}$ in $\pi_0(\mathscr{F}_4R) = \mathbf{Z}$ is labeled by $\frac{1}{4}$ Index B, which is $\frac{1}{2}\hat{A}(M)$ for a pure Dirac operator acting on the real spinors $\Gamma(S)$. As the other homotopy groups of $\mathscr{F}_4R$ vanish up to π_3, a more interesting example is given by $\not{D}$ coupled to an $\mathrm{Sp}(N)$ gauge field, the original global anomaly of Witten.[3] If $V(\epsilon)$ is a one-parameter family of $\mathrm{Sp}(N)$ connections on an $\mathscr{H}^N$ vector bundle E over M, with $V(1)$ differing from $V(0)$ by a gauge transformation g, then by Proposition 1 the class of $\not{D}_V$ in $\pi_1(\mathscr{F}_1R) = \mathbf{Z}_2$ is given by $\frac{1}{2}\dim_{\mathbf{R}}\ker T$ (mod 2), with $T = \begin{pmatrix} \partial_\epsilon & \not{D}_{V(\epsilon)} \\ \not{D}_{V(\epsilon)} & -\partial_\epsilon \end{pmatrix}$ acting on cross sections of the $\mathscr{H}^{2N}$ vector bundle over $S^1\times M$ created by twisting the ends of $I\times((E\otimes S)\oplus(E\otimes S))$ together by $-g$. Now

$$\begin{aligned}
\tfrac{1}{2}\dim_{\mathbf{R}}\ker T &= \tfrac{1}{2}\dim_{\mathbf{C}}\ker T \\
&= \tfrac{1}{2}\dim_{\mathbf{C}}\ker\rho(e_1)\rho(e_2)\rho(e_3)T \\
&= \tfrac{1}{2}\dim_{\mathbf{C}}\ker\rho(e_1)\rho(e_2)\rho(e_3) \\
&\quad\times\begin{pmatrix} 0 & -i\partial_\epsilon + \not{D}_{V(\epsilon)} \\ i\partial_\epsilon + \not{D}_{V(\epsilon)} & 0 \end{pmatrix} \\
&= \tfrac{1}{2}\dim_{\mathbf{C}}\ker\rho(e_1)\rho(e_2)\rho(e_3)(\partial_\epsilon + i\not{D}_{V(\epsilon)}) \\
&\quad + \tfrac{1}{2}\dim_{\mathbf{C}}\ker\rho(e_1)\rho(e_2)\rho(e_3) \\
&\quad\times(\partial_\epsilon - i\not{D}_{V(\epsilon)}).
\end{aligned}$$

However, over the complexes both $\rho(e_1)\rho(e_2)\rho(e_3)(\partial_\epsilon + i\not{D}_{V(\epsilon)})$ and $\rho(e_1)\rho(e_2)\rho(e_3)(\partial_\epsilon - i\not{D}_{V(\epsilon)})$ are equivalent to the real operator $\rho(e_1)\rho(e_2)\rho(e_3)\partial_\epsilon + \not{D}_{V(\epsilon)}$. Thus

$$\tfrac{1}{2}\dim_{\mathbf{R}}\ker T = \dim_{\mathbf{R}}\ker(\rho(e_1)\rho(e_2)\rho(e_3)\partial_{(\epsilon)} + \not{D}_{V(\epsilon)}).$$

In the case of $M = S^4$, the $\mathrm{Sp}(N)$ bundles over $S^1\times S^4$ are classified by $\mathscr{H}^5(S^1\times S^4, \mathbf{Z}_2) = \mathbf{Z}_2$, which can be thought of as the element of $\pi_4(\mathrm{Sp}(N)) = \mathbf{Z}_2$, used to join the ends of $I\times S^4$. Upon twisting by the Hopf bundle over S^1, one can check that $\dim_{\mathbf{R}}\ker(\rho(e_1)\rho(e_2)\rho(e_3)\partial_\epsilon + \not{D}_{V(\epsilon)})$ (mod 2), equals the class of g in $\pi_4(\mathrm{Sp}(N))$, which is Witten's orignal calculation. [In this case, because one is dealing with chiral spinors, one can also see that there is a global anomaly using the results of Witten–Bismut–Freed.[24] They showed that the holonomy of the Quillen connection on the determinant line bundle is, when going around the loop $V(\epsilon)$, $\exp -\pi i(\eta(\not{D}) + \dim_{\mathbf{C}}\ker\not{D})$ where, in our case, $\not{D} = i(\rho(e_1)\rho(e_2)\rho(e_3)\partial_{(\epsilon)} + \not{D}_{V(\epsilon)})$. Because $-i\not{D}$ lies in $\mathscr{F}_1R$, the spectral symmetry ensures that $\eta(\not{D})$ is zero. Thus the holonomy is 1 if $\dim_{\mathbf{C}}\ker\not{D}$ is even and -1 if $\dim_{\mathbf{C}}\ker\not{D}$ is odd, showing that in the latter case there is a global anomaly in the sense that the Quillen connection has nontrivial holonomy.]

V. QFT's ON NONCOMPACT SPACES

In general the index of a family of Fredholm operators on a noncompact space is harder to compute than in the compact case. We will only consider complex skew-adjoint operators T whose underlying real operator $T_{\mathbf{R}}$ lies in $\mathscr{F}_1R$. In general $\det^{1/2} T_{\mathbf{R}} = |\det T|$, and so an odd spectral flow in a family $T(\epsilon)$ prevents the smooth definition of $\det^{1/2} T_{\mathbf{R}}(\epsilon)$. This is seen in the fact that a class $[T_{\mathbf{R}}(\epsilon)]$ in $\pi_1(\mathscr{F}_1R) = \mathbf{Z}_2$ can be computed using the spectral flow (mod 2) of a circle of operators $T(\epsilon)$, which in turn can be computed using the generalized eta invariant. If $H(\epsilon)$ is a skew-adjoint operator which arises in the Lagrangian of a d-dimensional Euclidean QFT, the most practical way that we know to compute $\eta(H(\epsilon))$ is to regard $H(\epsilon)$ as the Hamiltonian for a $(d+1)$-dimensional Minkowski QFT and compute the vacuum expectation of the change operator $Q = \int <j^0(X)> d^dX$, which then gives $\eta(H(\epsilon))$ via the equation $Q = -\frac{1}{2}\eta$.[12] One can calculate Q (or more precisely, $dQ/d\epsilon$) for the $(d+1)$-dimensional theory via a gradient expansion[25] and then the spectral flow is simply the change in Q when going around the circle.

One way of producing the d-dimensional Lagrangian is as follows: let $\{M_i\}_{i=1}^{2d+1}$ be mutually anticommuting self-adjoint matrices and for a map ϕ: $\mathbf{R}^d \to \mathbf{R}^{d+1}$, consider the operator $T = \Sigma_{j=1}^{d} M^j\partial_j + \Sigma_{j=d+1}^{2d+1} iM^j\phi^j$ with ϕ approaching constants radially. This will be Fredholm iff $|\phi(x)|^2$ is bounded away from zero for large x and then the large x behavior of $\phi/|\phi|$ gives a map $\tilde{\phi}$: $S^{d-1}\to S^d$. We will show that under the one-parameter family of $\tilde{\phi}$'s that starts and ends with a point map, and covers S^d, there is an odd spectral flow and so a global anomaly.

A d-dimensional Euclidean Lagrangian incorporating T is

$$L(\phi,\psi)=\int_{\mathbf{R}^d}\frac{1}{2}\sum_{j=1}^{d+1}|\nabla\phi^j|^2+V\left(\sum_{j=1}^{d+1}\phi_j^2\right)+\frac{1}{2}\psi^\dagger T\psi.$$

The ϕ-ψ couplings are Yukawa type, and these Lagrangians can be thought to be analogs of the Gell-Mann–Lévy σ model[4] (as opposed to the more recent definitions of a σ model[26]!). L has an $SO(d+1)$ global symmetry, which we will argue to be broken by the global anomaly.

A. One dimension

Consider the Euclidean action

$$L^{(1)}(\phi,\psi)=\int_{\mathbf{R}}\frac{1}{2}(\partial_1\phi_1)^2+\frac{1}{2}(\partial_1\phi_2)^2+V(\phi_1^2+\phi_2^2)$$
$$+\tfrac{1}{2}\Psi^\dagger(\sigma^3\partial_1+i\sigma^1\phi_1+i\sigma^2\phi_2)\Psi,$$

for $\phi_1,\phi_2\in C^\infty(\mathbf{R}^1)$ and $\Psi\in C^\infty(\mathbf{R}^1)\otimes\mathbf{C}^2$, where V goes to ∞ as its argument goes to ∞. Here $L^{(1)}$ has an $SO(2)$ symmetry given by

$$\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}\to\begin{pmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix},$$

$\Psi\to e^{(1/2)i\alpha\sigma_3}\Psi$. First consider the case that V has a minimum away from the origin. Then there will be finite action bosonic paths which (as x_1 ranges from $-\infty$ to ∞) go from one point in the minimum well to another. Formally, the integration over these paths gives the $SO(2)$ symmetry in the quantum theory.

Let T denote the skew operator $\sigma^3\partial_1+i\sigma^1\phi_1+i\sigma^2\phi_2$. Consider a family $\phi(\epsilon)$ of background bosonic configurations with $\phi_{1\epsilon}(-\infty)=\phi_{2\epsilon}(-\infty)=0$ and $\phi_{1\epsilon}(\infty)=\cos\epsilon$, $\phi_{2\epsilon}(\infty)=\sin\epsilon$, as sketched in Fig. 1. As the fermionic integration in $\int e^{-L^{(1)}}\mathscr{D}\phi\,\mathscr{D}\Psi$ leaves a factor of $|\det T|$, if there is an odd spectral flow in $T(\epsilon)$ then one might expect that the instanton sum is ill-defined and the $SO(2)$ symmetry is broken.

For the operator $iT(\epsilon)$ one can show that the derivative of the generalized η invariant is

$$\frac{d}{d\epsilon}\eta(iT(\epsilon))=\frac{1}{\pi}\frac{1}{\phi_1^2+\phi_2^2}\left(\frac{d\phi_1}{d\epsilon}\phi_2-\frac{d\phi_2}{d\epsilon}\phi_1\right)\Bigg|\Bigg|_{-\infty}^{\infty}$$
$$=\frac{1}{\pi}\frac{d}{d\epsilon}TAN^{-1}\frac{\phi_1}{\phi_2}\Bigg|\Bigg|_{-\infty}^{\infty}$$

(Refs. 12 and 25). This can be seen by computing the vacuum charge for the two-dimensional Minkowski Lagrangian

$$L^{(2)}(\Psi)=i\int_{\mathbf{R}^2}\frac{1}{2}\bar{\Psi}(\partial_0+\sigma^3\partial_1+i\sigma^1\phi_1+i\sigma^2\phi_2)\Psi,$$

as was done in Ref. 25; the relevant Feynman diagram is that of Fig. 2. Thus there is odd spectral flow as ϵ goes from 0 to 2π.

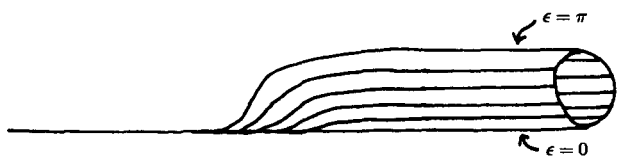


FIG. 1. A one-parameter family of background ϕ's in one dimension.

As our model is quantum mechanical, one can also analyze it in a Hamiltonian approach. The Hamiltonian $\mathscr{H}$, acting on $C^\infty(\mathbf{R}^2)\otimes\Lambda^*(\mathbf{C}^2)$, is

$$\mathscr{H}=\tfrac{1}{2}(-\partial_1^2-\partial_2^2)$$
$$+V(X_1^2+X_2^2)+\tfrac{1}{2}\Psi^\dagger(\sigma^2X_1-\sigma^1X_2)\Psi$$

where the operators Ψ_i satisfy $\{\Psi_i^*,\Psi_j\}=2\delta_{ij}$. The $U(1)$ charge is

$$Q=-i(X^1\partial_2-X^2\partial_1)+\tfrac{1}{4}i\Psi^\dagger\sigma_3\Psi,$$

and commutes with $\mathscr{H}$. Representing the complex Clifford algebra generated by the Ψ's on $\Lambda^*(\mathbf{C}^2)$ via $\Psi_i\to\sqrt{2}I(e_i)$, $\Psi_i^*\to\sqrt{2}E(e_i)$, we can split H as $H_1\oplus H_2$ where H_1 and H_2 act on $C^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{even}}(\mathbf{C}^2)$ and $C^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{odd}}(\mathbf{C}^2)$, respectively, and are given by

$$H_1=\tfrac{1}{2}(-\partial_1^2-\partial_2^2)+V(X_1^2+X_2^2)$$

and

$$H_2=\tfrac{1}{2}(-\partial_1^2-\partial_2^2)$$
$$+V(X_1^2+X_2^2)+\begin{pmatrix}0 & -iX_1-X_2\\ iX_1-X_2 & 0\end{pmatrix}.$$

Because Q has integer spectrum on $C^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{even}}(\mathbf{C}^2)$ and half-integer spectrum on $C^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{odd}}(\mathbf{C}^2)$, the $SO(2)$ symmetry of the ground state will be broken iff the ground state is in $C^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{odd}}(\mathbf{C}^2)$. However, for all $\Psi\in C_0^\infty(\mathbf{R}^2)$, $\langle\Psi|H_1|\Psi\rangle=\langle\Psi\otimes\binom{1}{0}|H_2|\Psi\otimes\binom{1}{0}\rangle$ and so

$$(\inf_{\Psi\in C_0^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{odd}}(\mathbf{C}^2)}\langle\Psi|H_2|\Psi\rangle)$$
$$\leqslant(\inf_{\Psi\in C_0^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{even}}(\mathbf{C}^2)}\langle\Psi|H_1|\Psi\rangle),$$

implying that the ground state is indeed in $C^\infty(\mathbf{R}^2)\otimes\Lambda^{\text{odd}}(\mathbf{C}^2)$. In this example it is clear that the existence of a global anomaly does not make a theory inconsistent but merely breaks a global symmetry; this appears to be related to the fact that the anomaly occurs in a global rather than local symmetry.

The functional integral argument for global symmetry breaking required that V have a nontrivial minimum in order that the fermionic operator in the background field be Fredholm. However, from the Hamiltonian argument one sees that symmetry is broken no matter what V is. This can be seen in the functional integral approach by compactifying the space-time from $\mathbf{R}$ to $[-\beta,\beta]$. If there is a symmetry breaking for each β then one would expect the same as β goes to ∞. A convenient choice of fermionic boundary conditions which preserves the $SO(2)$ symmetry is the Atiyah–Patodi–Singer (APS) boundary condition.[11] This requires that $\Psi(\beta)$ lie in the positive eigenspace of $-\sigma^2\phi_1(\beta)+\sigma^1\phi_2(\beta)$ and that $\Psi(-\beta)$ lie in the negative eigenspace of $-\sigma^2\phi_1(-\beta)+\sigma^1\phi_2(-\beta)$.

Proposition 4: Let $T(\epsilon)$ be a family of operators on $C^\infty([-\beta,\beta])\otimes\mathbf{C}^2$ given by $\sigma^3\partial_1+i\sigma^1\phi_1+i\sigma^2\phi_2$ with the

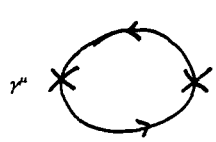


FIG. 2. One-dimensional spectral flow.

APS boundary condition, where $\begin{pmatrix}\phi_{1\epsilon}(-\beta)\\ \phi_{2\epsilon}(-\beta)\end{pmatrix}$ is a nonzero vector independent of ϵ and $\begin{pmatrix}\phi_{1\epsilon}(\beta)\\ \phi_{2\epsilon}(\beta)\end{pmatrix} = \begin{pmatrix}\cos\epsilon\\ \sin\epsilon\end{pmatrix}$. Then as ϵ goes from 0 to 2π, there is an odd spectral flow of $iT(\epsilon)$.

Proof: Let $V(\epsilon)$ denote the vector $\begin{pmatrix}\sqrt{\phi_1^2+\phi_2^2}\,(-\beta)\\ -(\phi_2 - i\phi_1)(-\beta)\end{pmatrix}$ and let $W(\epsilon)$ denote the vector $\begin{pmatrix}\sqrt{\phi_1^2+\phi_2^2}\,(\beta)\\ (\phi_2 - i\phi_1)(\beta)\end{pmatrix}$; the APS condition is that $\Psi_\epsilon(-\beta)$ is proportionate to $V(\epsilon)$ and $\Psi_\epsilon(\beta)$ is proportionate to $W(\epsilon)$. Because the spectral flow is a homotopy invariant, we can compute it for any loop in $\mathcal{F}_1$ homotopic to $T(\epsilon)$. In particular, for $0\leqslant\alpha\leqslant 1$, consider the loop of operators on $C^\infty([-\beta,\beta])\otimes\mathbf{C}^2$ given by $\sigma^3\partial_1 + \alpha i\sigma^1\phi_1 + \alpha i\sigma^2\phi_2$ with the boundary condition that $\Psi_\epsilon(-\beta)$ is proportionate to $V(\epsilon)$ and $\Psi_\epsilon(\beta)$ is proportionate to $W(\epsilon)$. One can check that this gives a smooth homotopy within the class of elliptic self-adjoint boundary value problems[27] and so it suffices to compute the spectral flow at $\alpha = 0$. Then the spectrum is

$$\left\langle \frac{1}{4\beta}\left[(2n+1)i\pi + \ln\frac{\phi_1+i\phi_2}{\sqrt{\phi_1^2+\phi_2^2}}(-\beta) - \ln\frac{\phi_1+i\phi_2}{\sqrt{\phi_1^2+\phi_2^2}}(\beta)\right]\right\rangle_{n\in\mathbf{Z}},$$

which has an odd spectral flow as ϵ goes from 0 to 2π. ■

In higher dimensions, we will only consider the case when V has a minimum away from the origin, so that the instantonlike background fields give Fredholm fermionic operators. Presumably one could put the theory in a finite volume, as we have done in one dimension, and conclude that there is a global anomaly with no restriction on V.

B. Two dimensions

Consider the two-dimensional Euclidean Lagrangian

$$L^{(2)} = \int_{\mathbf{R}^2}\sum_{j=1}^{3}\frac{1}{2}(\partial_\mu\phi_j)^2 + V\left(\sum_{j=1}^{3}\phi_j^2\right) + \frac{1}{2}\Psi^\dagger\left(\sum_{j=1}^{2}\gamma_E^j\partial_j + i\gamma_E^0\phi_1 + i\gamma_E^3\phi_2 + i\gamma_E^5\phi_3\right)\Psi$$

for $\phi_1,\phi_2,\phi_3\in C^\infty(\mathbf{R}^2)$ and $\Psi\in C^\infty(\mathbf{R}^2)\otimes\mathbf{C}^4$. Here $L^{(2)}$ has an SO(3) symmetry which rotates the ϕ's. If $T(\epsilon)$ is a one-parameter family of skew Fredholm operators of the form

$$T = \sum_{j=1}^{2}\gamma_E^j\partial_j + i\gamma_E^0\phi_1 + i\gamma_E^3\phi_2 + i\gamma_E^5\phi_3$$

then we will compute the generalized η invariant of $T(\epsilon)$ by considering iT to be equivalent to the Hamiltonian of the three-dimensional Minkowski Lagrangian

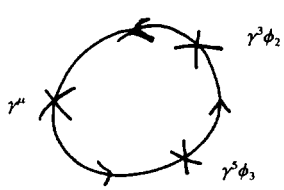


FIG. 3. Three-dimensional spectral flow.

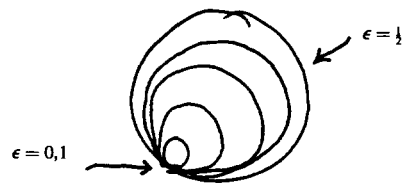


FIG. 4. A one-parameter family of background ϕ's (at ∞) in two dimensions.

$$L^{(3)} = \int_{\mathbf{R}^3}\bar{\Psi}\times\left(-i\gamma_E^0\partial_0 + \sum_{j=1}^{2}\gamma_E^j\partial_j + \phi_1 - i\gamma_E^3\phi_2 - i\gamma_E^5\phi_3\right)\Psi.$$

The Feynman diagram to compute the vacuum charge is that of Fig. 3 and letting n^a denote $\phi^a/|\phi|$, one finds

$$\frac{dQ}{d\epsilon} = \frac{d}{d\epsilon}\int_{\mathbf{R}^2}\frac{1}{8\pi}\epsilon_{abc}n^a dn^b\wedge dn^c.$$

This is simply the infinitesimal change in the volume on S^2 swept out by the curve $\phi/|\phi|: S^1\to S^2$, where the S^1 is a large circle in $\mathbf{R}^2$, and we have normalized the volume form on S^2 to have mass 1. Consider a one-parameter family of loops on S^2 as in Fig. 4. If each loop represents the behavior of $\phi/|\phi|$ for large radius in $\mathbf{R}^2$, for some ϵ, then as ϵ goes from 0 to 1 it follows that there is an odd change in the vacuum charge, and so an odd spectral flow in $iT(\epsilon)$. Presumably this spectral flow breaks the global SO(3) symmetry.

C. Three dimensions

Consider the three-dimensional Euclidian Lagrangian

$$L^{(3)} = \int_{\mathbf{R}^3}\frac{1}{2}\sum_{j=0}^{3}(\partial_\mu\phi_j)^2 + V\left(\sum_{j=0}^{3}\phi_j^2\right) + \frac{1}{2}\Psi^\dagger\left(\sum_{j=1}^{3}\gamma_E^j\partial_j + i\gamma_E^0\phi_0 + i\gamma_E^5\otimes\vec{\phi}\cdot\vec{\tau}\right)\Psi,$$

with $\phi_0,\ldots,\phi_3\in C^\infty(\mathbf{R}^3)$ and $\Psi\in C^\infty(\mathbf{R}^3)\otimes\mathbf{C}^8$. There is a naive SO(4) symmetry which rotates the ϕ's, and as before we will compute the spectral flow for the fermionic differential operator by computing the vacuum charge of the four-dimensional Minkowski Langrangian

$$L^{(4)} = \int_{\mathbf{R}}^{4}\frac{1}{2}\bar{\Psi}\left(-i\gamma_E^0\partial_0 + \sum_{j=1}^{3}\gamma_E^j\partial_j + \phi_0 + i\gamma^5\bar{\phi}\cdot\bar{\tau}\right)\Psi.$$

This calculation was done in Ref. 25 and the relevant Feynman diagram is that of Fig. 5. Letting n^a denote $\phi^a/|\phi|$, the result was

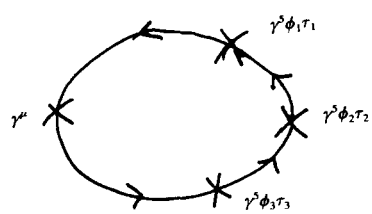


FIG. 5. Three-dimensional spectral flow.

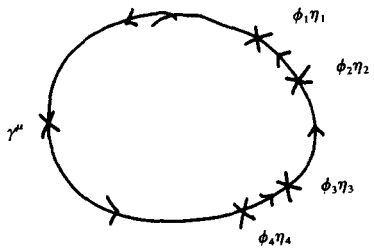


FIG. 6. Four-dimensional spectral flow.

$$\frac{dQ}{d\epsilon} = \frac{d}{d\epsilon}\int_{\mathbf{R}^3} \frac{1}{12\pi^2}\epsilon_{abcd} n^a\, dn^b \wedge dn^c \wedge dn^d.$$

As before, this is the infinitesimal change in the normalized volume on S^3 swept out by a family of immersed S^2's, and by choosing the family to cover S^3, we ensure that there is an odd spectral flow in the fermionic differential operator.

D. Four dimensions

Let $\{\eta^i\}_{i=1}^4$ be another copy of the Dirac matrices $\{\gamma_E^i\}_{i=0}^3$ and put $\eta^5 = \eta^1\eta^2\eta^3\eta^4$. Consider the four-dimensional Euclidean action

$$L^{(4)} = \int_{\mathbf{R}^4} \sum_{j=0}^4 (\partial_\mu \phi_j)^2 + V\left(\sum_{j=0}^4 \phi_j^2\right) + \frac{1}{2}\Psi^\dagger\left(\sum_{\mu=1}^4 \gamma_E^\mu \partial_\mu + i\gamma_E^5(\phi_0\eta^5 + \vec{\phi}\cdot\vec{\eta})\right)\Psi,$$

where $\phi_0,\ldots,\phi_4 \in C^\infty(\mathbf{R}^4)$ and $\Psi \in C^\infty(\mathbf{R}^4)\otimes \mathbf{C}^{16}$. This is an analog of the linear σ model[4] with the target space being $\mathbf{R}^5$ instead of $\mathbf{R}^4$. The naive SO(5) global symmetry rotates the ϕ's. The corresponding five-dimensional Minkowski Lagrangian is

$$L^{(5)} = \int_{\mathbf{R}^5} \frac{1}{2}\bar{\Psi} \times\left(-i\gamma_E^5\partial_0 + \sum_{\mu=1}^4 \gamma_E^\mu\partial_\mu + \phi_0\eta^5 + \vec{\phi}\cdot\vec{\eta}\right)\Psi.$$

The Feynman diagram to compute the vacuum charge is that of Fig. 6, and the result is (letting $n^a = \phi^a/|\phi|$)

$$\frac{dQ}{d\epsilon} = \frac{d}{d\epsilon}\frac{1}{64\pi^2}\int_{\mathbf{R}^4} \epsilon_{abcde} n^a\, dn^b \wedge dn^c \wedge dn^d \wedge dn^e.$$

Once again, this is the infinitesimal volume on S^4 swept out by the ϕ field at ∞, and by a suitable choice of $\phi(\epsilon)$ there will be an odd spectral flow in the fermionic differential operator.

VI. DETERMINANT BUNDLES

Over the space of Fredholm operators $\mathcal{F}$ one has the virtual index bundle Index and its highest exterior power, the line bundle Det. For Dirac-type operators Quillen showed how to define a natural metric on Det.[5] We wish to show how to extend these constructions to the other classes of Fredholm operators.

First, consider the space $\mathcal{F}_1$ of skew-adjoint complex Fredholm operators. The heuristic obstruction to defining a determinant function on $\mathcal{F}_1$ is the possible change of sign in going around a loop, that is, the mod 2 reduction of $\pi_1(\mathcal{F}_1) = \mathbf{Z}$. Abstractly one can form a flat $\mathbf{R}$ bundle over $\mathcal{F}_1$ via the homomorphism ρ: $\pi_1(\mathcal{F}_1) \to \mathrm{End}(\mathbf{R})$ which takes $1 \in \pi_1(\mathcal{F}_1)$ to the operator of multiplication by -1. To be more concrete, let us consider a space $\mathcal{S}$ of skew-adjoint Dirac type operators $\not{D}_s$ on a compact spin manifold, possibly coupled to an external vector bundle. As in Ref. 5, $\mathcal{S}$ can be covered by open sets $\{U_\alpha\}_{\alpha\in\mathbf{R}^+}$, so that for $s\in U_\alpha$, $i\not{D}_s$ has no eigenvalue of $\pm\alpha$. Then the transition functions (for $\alpha<\beta$) $g_{\alpha\beta}(S) = \Pi_{\alpha<|\lambda_i|<\beta}\lambda_i(i\not{D}_s)$ define an $\mathbf{R}$ bundle DET over $\mathcal{S}$. That is, $v_\alpha \in \mathbf{R}$ in a trivialization over U_β corresponds to $v_\beta = g_{\alpha\beta}v_\alpha$ in trivialization over U_β, and so there is a well-defined Quillen metric on DET given by

$$\|v_\alpha\|^2 = v_\alpha^2\left(\prod_{|\lambda_i|>\alpha}\lambda_i^2(i\not{D}_s)\right),$$

where the product is understood to be defined using zeta-function regularization. The unique connection on $\Gamma(\mathrm{DET})$ which preserves $\|\cdot\|$ is given in trivialization α by

$$A_\alpha = \lim_{s\to 0}\sum_{|\lambda_i|>\alpha}\lambda_i^{-s-1}\,d\lambda_i$$

and is flat. Thus under parallel transport in patch α, the quantity $(\Pi_{|\lambda_i|>\alpha}|\lambda_i|)v_\alpha$ is constant. One can convince oneself that the holonomy around a loop is the spectral flow (mod 2).

Now consider the space $\mathcal{F}_1 R$ of real skew-adjoint Fredholm operators. We can abstractly define a flat $\mathbf{R}$ bundle via the homomorphism ρ: $\pi_1(\mathcal{F}_1 R) = \mathbf{Z}_2 \to \mathrm{End}(\mathbf{R})$ which takes 1 to -1. For a space $\mathcal{S}$ of real skew-adjoint Dirac-type operators, define the covering $\{U_\alpha\}_{\alpha\in\mathbf{R}^+}$, as above. Over U_α we have the $\mathbf{R}$ bundle $\Lambda^{\max}(V_\alpha)$, where $V_\alpha = \oplus_i\{\text{eigenspaces of eigenvalue } \lambda_i, |\lambda_i|<\alpha\}$. If $\alpha<\beta$ then over $U_\alpha \cap U_\beta$, T defines a two-form on $V_\beta - V_\alpha$ (by $\Sigma_{\alpha<\lambda_i<\beta}\lambda_i e_i\wedge\bar{e}_i$, e_i orthonormal) and an isomorphism from $\Lambda^{\max}(V_\alpha)$ to $\Lambda^{\max}(V_\beta)$ via exterior multiplication by $T^{(1/2)(\dim_{\mathbf{R}} V_\beta - \dim_{\mathbf{R}} V_\alpha)}$; then the bundles $\Lambda^{\max}(V_\alpha)$ patch together to give an $\mathbf{R}$ bundle Pfaff over $\mathcal{S}$. There is a metric on Pfaff given by

$$\|\Lambda(e_i)_a\|^2 = |\Lambda(e_i)|_H^2 \prod_{\lambda_i>\alpha}\lambda_i^2(i\not{D}_s),$$

where $|\cdot|_{\mathcal{H}}$ denotes the metric induced from the Hilbert space $\mathcal{H}$ and there is a compatible flat connection. One can see that the holonomy of the connection around a loop is the number of eigenspace rearrangements (mod 2).

Because the elements of $\mathcal{F}_2 R$, can be written as $A + iB$ with A and B skew symmetric, the natural function to consider is the complex Pfaffian. Freed has shown that for Dirac-type operators in $\mathcal{F}_2 R$, the determinant line bundle of Quillen has a natural square root, the complex Pfaffian line bundle, with induced metric and connection.[24]

Finally, the elements of $\mathcal{F}_3 R$ can be considered to be skew-adjoint complex operators which anticommute with a complex antilinear map. Then the even dimensionality of the eigenspaces allows us to canonically take the square root of the transition functions used to define DET for the $\mathcal{F}_1$ case. In this way one obtains a flat line bundle $\mathrm{DET}^{1/2}$ which has holonomy around a loop given by $\frac{1}{2}$ (spectral flow) (mod 2).

ACKNOWLEDGMENTS

I wish to thank L. Alvarez-Gaumé for discussions which caused the consideration of Proposition 4, and D. Freed for valuable help with Proposition 1 and Sec. VI. I also wish to thank M. Jacob, J. Ellis, and the CERN Theory Division for their hospitality while part of this research was performed.

[1]M. F. Atiyah and I.M. Singer, "Dirac operators coupled to vector potentials," Proc. Natl. Acad. Sci. **81**, 2597 (1984).
[2]M. F. Atiyah and I.M. Singer, "Index theory for skew adjoint Fredholm operators," Publ. Math. IHES **37**, 5 (1969).
[3]E. Witten, "An SU(2) anomaly," Phys. Lett. B **117**, 324 (1982).
[4]M. Gell-Mann and M. Lévy, "The axial vector current in beta decay," Nuovo Cimento **16**, 705 (1960).
[5]D. Quillen, "Determinants of Cauchy-Riemann operators over a Riemann suface," Funkts. Anal. i. Prilozn **19**, 37 (1985).
[6]R. Coquereaux, "Modulo 8 periodicity of real Clifford algebras and particle physics," Phys. Lett. B **115**, 389 (1982).
[7]M. F. Atiyah, *K-theory* (Benjamin, New York, 1967).
[8]I. M. Singer, "Operator theory and periodicity," Indiana Univ. Math. J. **20**, 949 (1971).
[9]G. Segal, "Faddeev's anomaly and Gauss' law," preprint, Oxford.
[10]M. F. Atiyah, "On K-theory and reality," Quart. J. Math. Oxford **17**, 367 (1966).
[11]M. F. Atiyah, V. K. Patodi, and I. M. Singer, "Spectral symmetry and Riemannian geometry III," Math. Proc. Cambridge Phil. Soc. **79**, 71 (1976).
[12]J. Lott, "Vacuum change and the eta function," Commun. Math. Phys. **93**, 533 (1984).
[13]M. F. Atiyah and I.M. Singer, "The index of elliptic operators IV," Ann. Math. **93**, 119 (1971).
[14]C. G. Callan, R. F. Dashen, and D. Gross, "The structure of the gauge theory vacuum," Phys. Lett. B **63**, 334 (1976).
[15]R. Jackiw and C. Rebbi, "Vacuum periodicity in a Yang–Mills quantum theory," Phys. Rev. Lett. **37**, 172 (1976).
[16]R. Jackiw and C. Rebbi, "Solitons with fermion number 1/2," Phys. Rev. D **13**, 3398 (1976).
[17]N. Hitchin, "Harmonic spinors," Adv. Math. **14**, 1 (1974).
[18]E. Witten, *Global Anomalies in String Theory*, Symposium on Geometry and Topology (World Scientific, Singapore, 1985).
[19]M. F. Atiyah, "Riemann surfaces and spin structures," Ann. Sci. Ecole Norm. Sup. **4**, 47 (1971).
[20]J. Lott, "The eta function and some new anomalies," Phys. Lett. B **145**, 179 (1984).
[21]L. Alvarez-Gaumé, S. Della Pietra, and G. Moore, "Anomalies and odd dimensions," Ann. Phys. (NY) **163**, 288 (1985).
[22]K. Osterwalder and R. Schrader, "Euclidean Fermi fields and a Feynman–Kac formula for Boson–Fermion models," Helv. Phys. Acta. **46**, 277 (1973).
[23]H. Nicolai, "A possible constructive approach to (Super $\phi^3)_4$ I," Nucl. Phys. B **140**, 294 (1978).
[24]D. Freed, "On determinant line bundles," in *Mathematical Aspects of String Theory*, edited by S. T. Yau (World Scientific, Singapore, to be published).
[25]J. Goldstone and F. Wilczek, "Fractional quantum numbers on solitons," Phys. Rev. Lett. **47**, 986 (1981).
[26]G. Moore and P. Nelson, "The aetiology of sigma-model anomalies," Commun. Math. Phys. **100**, 83 (1985).
[27]P. Gilkey, *Invariance Theory, The Heat Equation and the Atiyah-Singer Index Theorem* (Publish or Perish, Wilmington, DE, 1984).

Collapse and exponentiation of infinite symmetry algebras of Euclidean projective and Grassmannian σ models

Guy Arsenault
CRM and Département de Physique, Université de Montréal, C.P. 6128, Succ. A, Montréal H3C 3J7, Canada

Michel Jacques[a)]
CRM and Laboratoire de Physique Nucléaire, Université de Montréal, C.P. 6128, Succ. A, Montréal H3C 3J7, Canada

Yvan Saint-Aubin
CRM and Département de Mathématiques et de Statistique, Université de Montréal, C.P. 6128, Succ. A, Montréal H3C 3J7, Canada

(Received 18 November 1987; accepted for publication 3 February 1988)

Various two-dimensional σ models enjoy an infinite set of infinitesimal transformations acting on their solution space. The action of these symmetries is investigated for the Euclidean projective and Grassmannian σ models. On the (anti-) self-dual sector of the latter, the algebra of symmetries is shown to collapse to a finite-dimensional algebra isomorphic to $\mathrm{sl}(n+1,\mathbb{C})$ for the models with fields in the Grassmannians $G_{n+1,p}$. The finite action obtained by exponentiation is given in a closed form. For $\mathbb{CP}^n$ models, this result is extended to the whole space of finite action solutions and the structure of the algebra remains $\mathrm{sl}(n+1,\mathbb{C})$. Hence the action is not transitive on the solution space.

I. INTRODUCTION

A great deal of work is currently under way to investigate the structure and applications of infinite-dimensional Lie algebras. From the mathematical point of view, they constitute the most recent development of group theory. For physicists, they are expected to play a role analogous to the one played by usual group theory for the understanding of the properties of elementary particles. Indeed, one hopes that infinite-dimensional Lie algebras and their representation theory will shed some light on string theories.

Two-dimensional models play a special role in that respect. Let us recall that the conformal algebra in two dimensions has the structure of two commuting Virasoro algebras. Sigma models are interesting in their own way because of their privileged link with string theories and also because infinite-dimensional Lie algebras often show up in their properties. Here we want to concentrate on the infinite algebra spanned by symmetry transformations, first introduced by Dolan[1] for the unitary principal sigma models. This field of research has been developed significantly by Wu[2] and Uhlenbeck,[3] among others. The general case of a sigma model with values in an arbitrary Riemannian symmetric space has been treated in a previous paper.[4]

The results obtained in Ref. 4 are particularly important because they show similarities between the structure of sigma models and the structure of the Kadomtsev–Petviashvili (KP) equation introduced by the Kyoto school.[5] Assuming that the infinitesimal transformations can be integrated to a group action, we are led to the following interesting question: Can we expect this symmetry group to act transitively on the space of solutions or, at least, can we characterize physically the orbits under that group? This question constitutes the first motivation for the present paper.

A second motivation is given by the existence of finite-dimensional orbits under the action of the symmetry group. (Such orbits are known to exist for the KP equation.[5]) On these orbits, the infinite-dimensional group would then collapse into a finite-dimensional one. Some distinguished Euclidean sigma models (those with values in a Kähler manifold[6]) admit instanton solutions. The subspaces of instanton solutions of given charge are finite dimensional. If the symmetry transformations map k-instantons into k-instantons, we get an example of the collapse process described above.

We propose to examine these problems in the case of the Euclidean sigma models with values in a complex Grassmann manifold. The instantons of these models are well known.[6] Actually, much more is known when the Grassmann manifold is a projective manifold. Indeed, Din and Zakrzewski[7] gave a construction of all finite action solutions starting from the instantons. The present paper deals with the fate of the symmetry transformations on all these solutions.

The paper is organized as follows. Section II is a review of the projector formulation of the Grassmannian sigma models, their instanton solutions, and their Lax formulation. In Sec. III, the infinitesimal transformations derived in Ref. 4 are introduced for the subspace of instanton solutions: It is shown that they map instantons into instantons and that the infinite-dimensional symmetry algebra collapses into a finite-dimensional one when restricted to the instanton sector of the model. The resulting symmetry generators can be integrated explicitly into a finite action. However, the resulting group does not act transitively on the instanton subspace. Section IV is devoted to the same questions, but this time for the whole space of finite action solutions of the projective sigma models. After a description of the construction of Din

[a)] On leave of absence from the Institut de Physique Théorique, Université Catholique de Louvain, Belgium.

 0022-2488/88/061465-07$02.50

and Zakrzewski,[7] we give the proof that the infinite-dimensional symmetry algebra also reduces to a finite-dimensional algebra on the whole solution space of the $\mathbb{CP}^n$ models. This algebra is also integrated explicitly into a group action consistent with the construction of Din and Zakrzewski[7] and does not act transitively on the solution space either. Section V contains some comments and open questions, together with an explicit example for the $\mathbb{CP}^2$ case.

II. GRASSMANNIAN MODELS

Euclidean sigma models are defined on a two-dimensional space with the coordinates (x_+,x_-). The particular models we consider here are those where the field takes its values in the Grassmann manifold $G(n+1,p)$ of p planes in $\mathbb{C}^{n+1}$. A useful parametrization of points in $G(n+1,p)$ is given by $(n+1)\times(n+1)$ Hermitian rank p projectors Σ:

$$\Sigma^\dagger=\Sigma,\quad \Sigma^2=\Sigma. \tag{2.1}$$

The equations of motion are

$$[\Box\Sigma,\Sigma]=0. \tag{2.2}$$

A particular subset of the solution space of the field equation (2.2) is given by the solutions of the self-duality condition[6]:

$$\Sigma\,\partial_+\Sigma=0. \tag{2.3}$$

From now on, we only consider finite action solutions, i.e., configurations Σ such that $\int \mathrm{tr}(\partial_+\Sigma\,\partial_-\Sigma)d^2x<\infty$. With this additional condition, solutions of (2.3) are called instantons.

Let us now turn to the Lax formulation of the model. It is well known[8,9] that Eqs. (2.2) are the integrability conditions for the following linear system:

$$\partial_\pm R(x_+,x_-;\lambda) = [1/(1\pm\lambda)]\,2I[\partial_\pm\Sigma,\Sigma]IR(x_+,x_-;\lambda), \tag{2.4}$$

where R is an $\mathrm{SL}(n+1,\mathbb{C})$-valued function depending on an additional complex parameter λ and $I=\mathrm{diag}(\mathbb{1}_p, -\mathbb{1}_{n+1-p})$. On the other hand, the Hermiticity condition $\Sigma^\dagger=\Sigma$ imposes the following condition on R:

$$R^\dagger(x_+,x_-;\lambda)=R^{-1}(x_+,x_-;-\bar\lambda). \tag{2.5}$$

In the particular case of instantons, a solution of (2.4) satisfying condition (2.5) is found to be[10]

$$R(x_+,x_-;\lambda)=I(\mathbb{1}-[2/(1-\lambda)]\Sigma). \tag{2.6}$$

III. SYMMETRY TRANSFORMATIONS IN THE INSTANTON SECTOR OF THE GRASSMANNIAN MODELS

A. Infinitesimal transformations

As explained in the Introduction, one of the goals of this paper is to examine the fate of the infinite-dimensional Lie algebra acting on the solution space of the Grassmannian models when restricted to the instanton sector. We refer the reader to a previous paper[4] for the detailed derivation of the action of the symmetry algebra. In the case at hand, the generating function for the infinitesimal transformations takes the form

$$\Sigma\to\Sigma'=\Sigma+\Delta^T(\lambda)\Sigma, \tag{3.1a}$$

with

$$\Delta^T(\lambda)\Sigma\equiv[IR(\lambda)TR^{-1}(\lambda)I,\Sigma]. \tag{3.1b}$$

The generating function is to be understood as defining an infinite set of symmetries $\Delta^{T(i)}\Sigma$:

$$\Delta^T(\lambda)\Sigma\equiv\sum_{i=0}^{\infty}\lambda^{-i}\Delta^{T(i)}\Sigma. \tag{3.2}$$

In order to preserve the Hermiticity condition (2.5), the parameter $T\in\mathrm{sl}(n+1,\mathbb{C})$ in $\Delta^{T(i)}\Sigma$ must be such that $T=-T^\dagger$ if i is even and $T=T^\dagger$ if i is odd.

Using (2.6), the symmetry generators for the instantons read as

$$\Delta^T(\lambda)\Sigma=[(1+\lambda^2)/(1-\lambda^2)][T,\Sigma] + [2\lambda/(1-\lambda^2)](\{\Sigma,T\}-2\Sigma T\Sigma), \tag{3.3}$$

where again, in the first term, T has to be taken anti-Hermitian, while in the second term, it has to be taken Hermitian.

Although the general results[4] ensure that Σ' is a new solution of the full equations of motion (2.2), they do not guarantee that Σ' is still a solution of the self-duality condition (2.3). In fact, one can check explicitly that $\Sigma'\,\partial_+\Sigma'=0$ by going into the basis where Σ is diagonal. On the other hand, it will be possible to integrate the infinitesimal law (3.3) and obtain the explicit form of the finite transformation to which it corresponds. We thus postpone the proof that Σ' is indeed an instanton to Sec. III B.

The next step is to identify the algebraic structure spanned by the symmetry generators $\{\Delta^{T(i)},\ i\in\mathbb{N}, T=-T^\dagger$ for i even, $T=T^\dagger$ for i odd$\}$. A general result may be deduced immediately from the explicit form of the transformation (3.3): All the generators $\Delta^{T(i)}$ with an even index will act in the same way on Σ, as will all the generators $\Delta^{T(i)}$ with an odd index. This observation solves one of the problems discussed in this paper: The infinite-dimensional symmetry algebra of the Euclidean sigma models with values in a Grassmann manifold collapses into a finite-dimensional algebra when restricted to the instanton sector of the model.

The structure of this finite-dimensional algebra is easily read by computing the commutators between the generators that are now referred to by the notation

$$\Delta_e^T\Sigma\equiv[T,\Sigma],\quad \text{with } T^\dagger=-T, \tag{3.4a}$$

$$\Delta_o^T\Sigma\equiv\{\Sigma,T\}-2\Sigma T\Sigma,\quad \text{with } T^\dagger=T. \tag{3.4b}$$

The result is

$$[\Delta_e^U,\Delta_e^V]\Sigma=\Delta_e^{[U,V]}\Sigma,\quad \text{with } U=-U^\dagger,\quad V=-V^\dagger, \tag{3.5a}$$

$$[\Delta_o^U,\Delta_o^V]\Sigma=\Delta_e^{[U,V]}\Sigma,\quad \text{with } U=U^\dagger,\quad V=V^\dagger, \tag{3.5b}$$

$$[\Delta_e^U,\Delta_o^V]\Sigma=\Delta_o^{[U,V]}\Sigma,\quad \text{with } U=-U^\dagger,\quad V=V^\dagger. \tag{3.5c}$$

The finite-dimensional symmetry algebra thus has the structure of $\mathrm{sl}(n+1,\mathbb{C})$ by using the decomposition $\mathrm{sl}(n+1,\mathbb{C})=\mathfrak{h}\oplus\mathfrak{m}$, where $\mathfrak{h}$ is the subalgebra $\mathrm{su}(n+1)$ (corresponding to Δ_e^T) and $\mathfrak{m}$ is the subset of Hermitian matrices (corresponding to Δ_o^T).

This finite-dimensional structure embeds into the infinite-dimensional structure in the following way. We expect

the above symmetric space decomposition to fit into the structure of a twisted loop algebra by identifying all the generators with even indices in the gradation and all those with odd indices. However, we have to check that this idea corresponds indeed to the structure spanned by the infinity of generators in (3.3). The commutators are

$$[\Delta^U(\lambda),\Delta^V(\lambda')]\Sigma = [1/(\lambda'-\lambda)]\{\lambda'\Delta^{[U,V]}(\lambda)-\lambda\Delta^{[U,V]}(\lambda')\}\Sigma + [1/(\lambda\lambda'-1)]\{\Delta^{[U,V]}(\lambda)+\Delta^{[U,V]}(\lambda')\}\Sigma. \tag{3.6}$$

We have already encountered such a form for a commutator of generating functions in the general case.[4] In Ref. 4 it was proved that, under a change of basis, these commutation rules have the structure of the twisted loop algebra

$$\hat{\mathfrak{g}} = \left(\bigoplus_{i\in\mathbb{N}}(\mathrm{su}(n+1)\otimes t^{2i})\right)\oplus\left(\bigoplus_{i\in\mathbb{N}}(\mathfrak{m}\otimes t^{2i+1})\right),$$

where $\mathfrak{m}$ is the space of $((n+1)\times(n+1))$ Hermitian matrices.

The conclusion of this analysis is that the algebraic structure spanned by the symmetry generators of the instanton sector of the Grassmannian models is the above twisted loop algebra. However, all the even (odd) generators in the gradation act in the same way as the subspace of index 0 (1) and the infinite-dimensional structure thus collapses into the structure of the finite-dimensional algebra $\mathrm{su}(n+1)\oplus\mathfrak{m} = \mathrm{sl}(n+1,\mathbb{C})$.

B. Finite transformations

Returning to the transformation laws (3.4), we see that (3.4a) can be integrated immediately. This is not the case for (3.4b). However, both cases can be treated simultaneously if we observe that they are both special cases of a linear transformation on a Grassmann manifold, expressed in affine coordinates.[11] With this remark, the transformations (3.4) can be seen as the infinitesimal forms of

$$\Sigma_e' = e^T\Sigma e^{-T}, \quad \text{with } T^\dagger = -T, \tag{3.7a}$$

$$\Sigma_o' = e^T\Sigma(e^{-T}+2\sinh T\,\Sigma)^{-1} = e^T\Sigma(e^{-2T}(\mathbf{1}-\Sigma)+\Sigma)^{-1}e^{-T}, \quad \text{with } T^\dagger = T. \tag{3.7b}$$

We are now left with the problem of proving that the laws (3.7) are indeed symmetries of the instanton sector. For (3.7a), this fact is trivial. The new information is contained in (3.7b), which is a far from obvious symmetry of the self-duality equation.

First, let us show that the inverse in (3.7b) always exists. This will be done in the basis where Σ is diagonal: $\Sigma\equiv\mathrm{diag}[\mathbf{1}_p,0]$. Denoting e^{-2T} in this basis as

$$e^{-2T}\equiv\begin{bmatrix}\tau_1 & \tau_2\\ \tau_3 & \tau_4\end{bmatrix}, \quad \text{with } \tau_1^\dagger=\tau_1,\ \tau_4^\dagger=\tau_4,\ \tau_2^\dagger=\tau_3, \tag{3.8a}$$

we obtain

$$e^{-2T}(\mathbf{1}-\Sigma)+\Sigma = \begin{bmatrix}\mathbf{1} & \tau_2\\ 0 & \tau_4\end{bmatrix}.$$

This proves that the matrix appearing in (3.7b) is invertible as soon as τ_4 is invertible. However, τ_4, being a submatrix of a positive definite matrix, is itself positive definite and hence always invertible.

For later convenience, we also denote e^{2T} in this basis as

$$e^{2T} = \begin{bmatrix}\tilde{\tau}_1 & \tilde{\tau}_2\\ \tilde{\tau}_3 & \tilde{\tau}_4\end{bmatrix}, \quad \text{with } \tilde{\tau}_1^\dagger=\tilde{\tau}_1,\ \tilde{\tau}_4^\dagger=\tilde{\tau}_4,\ \tilde{\tau}_3^\dagger=\tilde{\tau}_2. \tag{3.8b}$$

The properties of the Σ_o' are (i) $\Sigma_o'^\dagger=\Sigma_o'$, (ii) $\Sigma_o'^2=\Sigma_o'$, and (iii) $\Sigma_o'\,\partial_+\Sigma_o'=0$ and should be proved in that order. Since the proofs are similar, we give only the calculation for the self-duality condition (iii). Observe first that

$$\partial_+\Sigma_o' = e^T\partial_+\Sigma(e^{-2T}(\mathbf{1}-\Sigma)+\Sigma)^{-1}e^{-T} - \Sigma_o'e^T(\mathbf{1}-e^{-2T})\partial_+\Sigma \times(e^{-2T}(\mathbf{1}-\Sigma)+\Sigma)^{-1}e^{-T}$$

so that

$$\Sigma_o'\,\partial_+\Sigma_o' = \Sigma_o'e^{-T}\partial_+\Sigma(e^{-2T}(\mathbf{1}-\Sigma)+\Sigma)^{-1}e^{-T}.$$

Now let Λ be the unitary matrix that diagonalizes Σ. Then

$$\Sigma_o'e^{-T} = e^T\Sigma(e^{-2T}(\mathbf{1}-\Sigma)+\Sigma)^{-1}e^{-2T} = e^T\Sigma((\mathbf{1}-\Sigma)+e^{2T}\Sigma)^{-1} = e^T\Lambda^{-1}\begin{bmatrix}\mathbf{1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\tilde{\tau}_1 & 0\\ \tilde{\tau}_3 & \mathbf{1}\end{bmatrix}^{-1}\Lambda = e^T\Lambda^{-1}\begin{bmatrix}\tilde{\tau}_1^{-1} & 0\\ 0 & 0\end{bmatrix}\Lambda = e^T\Lambda^{-1}\begin{bmatrix}\tilde{\tau}_1^{-1} & 0\\ 0 & 0\end{bmatrix}\Lambda\Sigma$$

and thus

$$\Sigma_o'\,\partial_+\Sigma_o' = e^T\Lambda^{-1}\begin{bmatrix}\tilde{\tau}_1^{-1} & 0\\ 0 & 0\end{bmatrix}\Lambda\Sigma\,\partial_+\Sigma \times(e^{-2T}(\mathbf{1}-\Sigma)+\Sigma)^{-1}e^{-T} = 0. \tag{3.9}$$

Before concluding this section, we want to make two remarks. First, let v be an eigenvector of Σ with eigenvalue 1. Then $\Sigma_0'e^Tv = e^Tv$, which means that the eigenspace of Σ is "rotated" by e^T with T Hermitian for the odd case, whereas it is obviously rotated by e^T with T anti-Hermitian in the even case.

Second, there is an obvious discrepancy between the dimension of the symmetry group $\mathrm{SL}(n+1,\mathbb{C})$ and the number of parameters of the instanton solution.[6] Thus the action of the above symmetry group is not transitive on the space of solutions since it is not even transitive on the subspace of self-dual solutions.

IV. SYMMETRIES ON THE SOLUTION SPACE OF THE $\mathbb{C}P^n$ MODELS

Among all (complex) Grassmannian models, it is the projective models whose solution spaces have been investigated most thoroughly. Indeed, because of the studies of Borchers and Garber[12] and Din and Zakrzewski,[7] the space of solutions with finite action of the Euclidean $\mathbb{C}P^n$ models is well understood: Its characterization is explicit enough so that we can extend the results obtained for the self-dual and anti-self-dual solutions of the Grassmannian models to the whole finite action solution space of the $\mathbb{C}P^n$ models.

As a first step toward the description of the symmetries

on the whole space of finite action solutions, we shall summarize the results of Din and Zakrzewski.[7] Since the fields of the $\mathbb{CP}^n$ models are rank-1 projectors $P(x_+,x_-)$, it is convenient to introduce the unit vector field $z(x_+,x_-) \in \mathbb{C}^{n+1}$, $|z|=1$, on which $P(x_+,x_-)$ projects. The field equations are

$$D_-D_+z + (z^\dagger D_-D_+z)z = 0 \tag{4.1}$$

and can be shown to be equivalent to (2.2) for P. Here, the covariant derivatives $D_\pm$ are defined by

$$D_\pm \equiv \partial_\pm - (z^\dagger \partial_\pm z). \tag{4.2}$$

In terms of the unit vector field z, the (anti-) self-duality condition takes the simple form

$$D_\mp z = 0, \tag{4.3}$$

where the lower sign is for anti-self-duality. The general instanton solution is given by $(n+1)$ polynomials $p_l(x_+)$ of x_+ only, $0 \leqslant l \leqslant n$, with no common root, and has the form

$$z_l = p_l(x_+)\left(\sum_{m=0}^{n} |p_m(x_+)|^2\right)^{-1/2}. \tag{4.4}$$

The anti-instantons bear the same definition, with x_+ replaced by x_-. The (anti-) instanton number is $k = \max_{0<l<n}(\deg(p_l))$.

Define the operators $P_\pm$ by the following action on a field $f(x_+,x_-) \in \mathbb{C}^{n+1}$:

$$P_\pm f = \partial_\pm f - \frac{(f^\dagger \partial_\pm f)}{|f|^2} f. \tag{4.5}$$

Let z be any finite action solution of the $\mathbb{CP}^n$ model. Din and Zakrzewski[7] were able to show the following remarkable properties: (i) there exist k and m such that $P_+^{k+1}z = P_-^{m+1}z = 0$ with $k+m=n$; (ii) every element of the set

$$\left\{\frac{P_-^m z}{|P_-^m z|}, \frac{P_-^{m-1}z}{|P_-^{m-1}z|}, \ldots, \frac{P_- z}{|P_- z|}, z, \frac{P_+ z}{|P_+ z|}, \ldots, \frac{P_+^{k-1}z}{|P_+^{k-1}z|}, \frac{P_+^k z}{|P_+^k z|}\right\} \tag{4.6}$$

is a solution of the $\mathbb{CP}^n$ model with finite action; and (iii) all the elements of the set (4.6) are mutually orthogonal,

$$\begin{aligned} &(P^i_- z)^\dagger (P^j_+ z) = 0, \quad \text{all } i,j, \\ &(P^i_- z)^\dagger (P^j_- z) = (P^i_+ z)^\dagger (P^j_+ z) = 0, \quad i \neq j. \end{aligned} \tag{4.7}$$

Hence the set (4.6) constitute a moving frame of $\mathbb{C}^{n+1}$ whose elements are solutions of the $\mathbb{CP}^n$ model. We shall refer to such a set as a *family* of solutions. Since $P_+^{k+1}z = P_-^{m+1}z = 0$, $P_+^k z$ and $P_-^m z$ are, respectively, an anti-instanton and an instanton. [According to Din and Zakrzewski,[7] it might occur, for particular z's, that $P_+^{k+1}z = P_-^{m+1}z = 0$ for k and m such that $k+m<n$. In these degenerate cases, the family (4.6) is not a $\mathbb{C}^{n+1}$ frame. However, these z's can be approximated continuously by nondegenerate families. We shall not consider any further these peculiar solutions.] Since any solution z with finite action gives rise to an instanton by repeated action of the operator P_-, it is easier to describe the families as generated from an instanton f, i.e., $P_-f = 0$:

$$\{f, P_+f/|P_+f|, \ldots, P_+^n f/|P_+^n f|\}, \tag{4.8}$$

where $P_+^n f/|P_+^n f|$ is the anti-instanton of the family.

This family structure can be translated in terms of a family of projectors $\{P_0,P_1,\ldots,P_n\}$. If $z_k \equiv P_+^k f$, for $0 \leqslant k \leqslant n$, define P_k to be the Hermitian projector on z_k. Since the z_k's are solutions of (4.1), the projectors P_k are solutions of (2.2). Moreover, the orthogonality of the z_k's is equivalent to $P_iP_j = \delta_{ij}P_j$. There is no compact form for the action of the operator P_+ *on the projectors* P_k and the recursive construction of the family (4.8) is more easily written in terms of the z_k's:

$$P_{k+1}(P_+ z_k) = (P_+ z_k). \tag{4.9}$$

A set of projectors $\{P_0,P_1,\ldots,P_n\}$ verifying the above conditions gives rise to a unique family $\{z_0,z_1,\ldots,z_n\}$ verifying conditions (i)–(iii).

To take full advantage of the formalism used to describe the symmetries introduced in Sec. III, another description of the families $\{z_0,z_1,\ldots,z_n\}$ or $\{P_0,P_1,\ldots,P_n\}$ is also useful. Let the Σ_k, $k=0,1,\ldots,n+1$, be the matrices

$$\Sigma_0 \equiv 0 \text{ and } \Sigma_k \equiv \sum_{l=0}^{k-1} P_l. \tag{4.10}$$

Obviously, the set

$$\{\Sigma_0,\Sigma_1,\ldots,\Sigma_n,\Sigma_{n+1} \equiv \mathbb{1}\} \tag{4.11}$$

contains the same information as the families $\{z_0,z_1,\ldots,z_n\}$ and $\{P_0,P_1,\ldots,P_n\}$ since the P_k's can be retrieved from (4.11) simply by $P_k = \Sigma_{k+1} - \Sigma_k$. The conditions on the family $\{P_0,P_1,\ldots,P_n\}$ are equivalent to the conditions (4.12)–(4.15) on the corresponding family $\{\Sigma_0,\Sigma_1,\ldots,\Sigma_{n+1}\}$. The Σ_k are projectors of rank k:

$$\Sigma_k^2 = \Sigma_k^\dagger = \Sigma_k, \quad \text{with rank } \Sigma_k = k, \tag{4.12}$$

satisfying the self-duality equation (2.3) of the Grassmannian models

$$\Sigma_k \partial_+ \Sigma_k = 0. \tag{4.13}$$

The image of Σ_i is a subspace of the image of Σ_j if $i \leqslant j$:

$$\Sigma_i\Sigma_j = \Sigma_j\Sigma_i = \Sigma_i, \quad \text{if } i \leqslant j. \tag{4.14}$$

Finally, the projectors Σ_k satisfy the supplementary differential conditions

$$\Sigma_{k+1}\,\partial_+\Sigma_k = \partial_+\Sigma_k, \quad 0 \leqslant k \leqslant n, \tag{4.15}$$

where $\Sigma_{n+1} = \mathbb{1}$. The equivalence of the purely algebraic conditions on the family $\{P_0,P_1,\ldots,P_n\}$ on one hand, and on the family $\{\Sigma_0,\Sigma_1,\ldots,\Sigma_{n+1}\}$ on the other hand, is straightforward. The equivalence between the differential relations on the two families can be derived by direct manipulations. We give here only an outline of the proof that $[\Box P_k, P_k] = 0$ holds given (4.13) and (4.15). The self-duality equations (4.13) on Σ_k imply the field equations $[\Box\Sigma_k, \Sigma_k] = 0$ for all k. Then

$$
\begin{aligned}
[\Box P_k,P_k] &= [\Box(\Sigma_{k+1}-\Sigma_k),(\Sigma_{k+1}-\Sigma_k)] \\
&= -[\Box\Sigma_{k+1},\Sigma_k]-[\Box\Sigma_k,\Sigma_{k+1}] \\
&= -\partial_+(\partial_-\Sigma_{k+1}\Sigma_k)+\partial_-\Sigma_{k+1}\,\partial_+\Sigma_k \\
&\quad +\partial_+(\Sigma_k\,\partial_-\Sigma_{k+1})-\partial_+\Sigma_k\,\partial_-\Sigma_{k+1} \\
&\quad -\partial_-(\partial_+\Sigma_k\Sigma_{k+1})+\partial_+\Sigma_k\,\partial_-\Sigma_{k+1} \\
&\quad +\partial_-(\Sigma_{k+1}\,\partial_+\Sigma_k)-\partial_-\Sigma_{k+1}\,\partial_+\Sigma_k.
\end{aligned}
\tag{4.16}
$$

Using various derivatives of $\Sigma_{k+1}\Sigma_k=\Sigma_k\Sigma_{k+1}=\Sigma_k$, the self-duality equations, and (4.15), Eq. (4.16) becomes

$$
[\Box P_k,P_k] = -\partial_-(\partial_+\Sigma_k)+\partial_-(\partial_+\Sigma_k)=0.
$$

Since each element of the family $\{\Sigma_k,\ 0\leqslant k\leqslant n+1\}$ is a self-dual solution of a Grassmannian σ model, the symmetries studied in Sec. III transform each of them into a new solution of the same model. If the symmetries can be shown to leave the conditions (4.12)–(4.15) satisfied, they will induce symmetries of the families $\{P_k,\ 0\leqslant k\leqslant n\}$ and then, of the whole space of finite action solutions of the $\mathbb{CP}^n$ model. In the case of the even generators, the transformations (3.7a) are

$$
\Sigma_k' = e^T\Sigma_k e^{-T},\quad \text{for } 0\leqslant k\leqslant n+1 \text{ and } T\in\mathrm{su}(n+1),
\tag{4.17}
$$

and this verification is trivial since all the projectors Σ_k and P_k become transformed in the same rigid fashion $\Sigma_k\to\Sigma_k'=e^T\Sigma_k e^{-T}$, $P_k\to P_k'=e^TP_ke^{-T}$. The conditions (4.12)–(4.15) are obviously satisfied by the Σ_k''s if they are by the unprimed Σ_k's. For the odd generators, the finite action (3.7b) is

$$
\begin{aligned}
&\Sigma_k' = e^T\Sigma_k[e^{-2T}(\mathbf{1}-\Sigma_k)+\Sigma_k]^{-1}e^{-T}, \\
&\qquad \text{for } 0\leqslant k\leqslant n+1 \text{ and } T \text{ Hermitian},
\end{aligned}
\tag{4.18}
$$

but the verification here is, however, a nontrivial one. Condition (4.12) was discussed in Sec. III. Condition (4.14) can be shown to hold following steps similar to the ones leading to (3.9). Since the self-duality equation has been verified in Sec. III, only condition (4.15) still remains. The derivative $\partial_+\Sigma_k'$ has the form

$$
\begin{aligned}
\partial_+\Sigma_k' = e^T\{&\partial_+\Sigma_k-\Sigma_k[e^{-2T}(\mathbf{1}-\Sigma_k)+\Sigma_k]^{-1} \\
&\times(\mathbf{1}-e^{-2T})\partial_+\Sigma_k\} \\
&\times[e^{-2T}(\mathbf{1}-\Sigma_k)+\Sigma_k]^{-1}e^{-T}.
\end{aligned}
$$

With the notation $[\,_k]\equiv[e^{-2T}(\mathbf{1}-\Sigma_k)+\Sigma_k]$, condition (4.15) becomes, using (4.14) and (4.15) for the unprimed Σ_k's,

$$
\begin{aligned}
&\Sigma_{k+1}'\,\partial_+\Sigma_k' \\
&\quad = e^T\Sigma_{k+1}[\,_{k+1}]^{-1}\{\Sigma_{k+1}\,\partial_+\Sigma_k-\Sigma_{k+1}\Sigma_k[\,_k]^{-1} \\
&\qquad \times(\mathbf{1}-e^{-2T})\partial_+\Sigma_k\}[\,_k]^{-1}e^{-T} \\
&\quad = e^T\Sigma_{k+1}\{\partial_+\Sigma_k-\Sigma_k[\,_k]^{-1}(\mathbf{1}-e^{-2T})\partial_+\Sigma_k\} \\
&\qquad \times[\,_k]^{-1}e^{-T}
\end{aligned}
$$

because of $(\Sigma_{k+1}')^2=\Sigma_{k+1}'$. Hence

$$
\begin{aligned}
\Sigma_{k+1}'\,\partial_+\Sigma_k' &= e^T\{\Sigma_{k+1}\,\partial_+\Sigma_k-\Sigma_{k+1}\Sigma_k[\,_k]^{-1} \\
&\quad \times(\mathbf{1}-e^{-2T})\partial_+\Sigma_k\}[\,_k]^{-1}e^{-T} \\
&= \partial_+\Sigma_k'.
\end{aligned}
$$

Thus all the symmetries described in Sec. III are symmetries not only of the (anti-) instanton solutions, but also of the whole space of finite action solutions. However, the explicit action of the odd generators on the z_k, which are neither instanton nor anti-instanton solutions, cannot be expressed solely in terms of z_k and e^T, but also requires the first few derivatives $\partial^i{}_+\,z_k$ and $\partial^j{}_-\,z_k$. Since the family structure is conserved, the simplest expression for the z_k' for $1\leqslant k\leqslant n$ is in terms of $f\equiv z_0$, the instanton of the family (see the remark at the end of Sec. III):

$$
z_k' = P^k{}_+f'/|P^k{}_+f'| = P^k{}_+\,(e^Tf)/|P^k{}_+\,(e^Tf)|.
\tag{4.19}
$$

Because T is Hermitian, $|f'|^2=|f^\dagger e^{2T}f|$ is not equal to $|f|^2$ in general and then $P_+(e^Tf)\neq e^TP_+f$. Section V provides an example of an "odd" transformation.

The action just described is not transitive on the solution space. It is not even transitive on the instanton solutions, as was underlined for the general Grassmannian case. In the $\mathbb{CP}^n$ model, the general k-instanton solution takes the form (4.4), where the maximum degree among the $p_l(x_+)$, $0\leqslant l\leqslant n+1$, is k and the polynomials $p_l(x_+)$ have no common root. Since both the even and odd symmetries act on z_0 as $z_0\to e^Tz_0/|e^Tz_0|$, where T is in $\mathrm{sl}(n+1,\mathbb{C})$ and e^T is always invertible, e^Tz_0 have no common root since, otherwise, e^Tz_0 would represent an l-instanton solution with $l<k$ and $e^{-T}(e^Tz_0)$ would not be a k-instanton. Hence, the action of the group described above is onto the k-instanton solution subspace. Since the algebra $\mathrm{sl}(n+1,\mathbb{C})$ has $2n(n+2)$ real dimensions and the general k-instanton solution depends on $2(n+1)(k+1)-2$ real parameters, the action is obviously not transitive for $k>n$.

To conclude this section, we reformulate the result just obtained through the technique used in Sec. III to see whether that same technique could provide a group of symmetries for the solutions z_k, $1\leqslant k\leqslant n-1$, bigger than the one just described here. The answer will be that both techniques coincide. (The details will be skipped; only the major steps will be outlined.) Din *et al.*,[10] gave the solution of the linear system associated to a given P_k in terms of the P_l, $0\leqslant l\leqslant k$, as

$$
R_k(\lambda) = I\{\mathbf{1}+[4\lambda/(\lambda-1)^2]\Sigma_k-[2/(1-\lambda)]P_k\}.
\tag{4.20}
$$

The generating function for the symmetry generators is

$$
\begin{aligned}
\Delta^T(\lambda)P_k &= [(1+\lambda^2)/(1-\lambda^2)][T,P_k] \\
&\quad +[2\lambda/(1-\lambda^2)](\{P_k,T_k\}-2P_kT_kP_k).
\end{aligned}
\tag{4.21}
$$

The (point-dependent) matrix generator T_k appearing in the odd part of the generating function is

$$
T_k\equiv(1-2\Sigma_k)T(1-2\Sigma_k).
\tag{4.22}
$$

Again, all the even generators coincide and are equal to

$$
\Delta_e^TP_k = [T,P_k],\quad T\in\mathrm{su}(n+1);
\tag{4.23}
$$

their exponentiation is obvious and has been given before. The odd generators are also identical:

$$\Delta_o^T P_k = \{P_k, T_k\} - 2P_k T_k P_k, \quad T \text{ Hermitian}, \tag{4.24}$$

but their exponentiation is difficult since it depends on the evolution of the P_l, $0 \leqslant l \leqslant k$, under that symmetry. One does not know *a priori* what the elements z_l' of the family of z_k' are, except by the action of P_- on z_k'. However, it is not difficult to prove, up to first order in T, that the set $\{P_0 + \Delta_o^T P_0, P_1 + \Delta_o^T P_1, \ldots, P_n + \Delta_o^T P_n\}$, where the $\Delta_o^T P_k$ are given by (4.24), is still a family. This very fact allows us to sum the $\Delta_o^T P_l$ and then define the family $\{\Sigma_0 + \Delta_o^T \Sigma_0, \Sigma_1 + \Delta_o^T \Sigma_1, \ldots, \Sigma_n + \Delta_o^T \Sigma_n\}$, where the $\Delta_o^T \Sigma_n$ are obtained from (4.24):

$$\Delta_o^T \Sigma_k = \sum_{l=0}^{k-1} \Delta_o^T P_l = \{\Sigma_k, T\} - 2\Sigma_k T \Sigma_k, \tag{4.25}$$

where T now appears *nonconjugated*. However, this is exactly the infinitesimal form (3.4b) obtained in Sec. III for the projector Σ. The conclusion is then that the finite symmetries introduced earlier in this section coincide, in the infinitesimal limit, with the transformations obtained through the solution of the linear system and the technique used in Sec. III.

V. CONCLUDING REMARKS

As a first remark, we present a nontrivial example of the action of the odd generators on a family of the projective model $\mathbb{CP}^n$. In order to be nontrivial, the family has to be from a model with $n \geqslant 2$ since the $\mathbb{CP}^1$ model has only instantons and anti-instantons as finite action solutions. We choose the simplest of these models, the $\mathbb{CP}^2$ model.

The family will be generated from the following two-instanton solution:

$$f = z_0 = \frac{1}{\sqrt{1 + x^2 + x^4}} \begin{bmatrix} 1 \\ x_+ \\ x_+^2 \end{bmatrix}, \tag{5.1}$$

where $x^2 \equiv x_+ x_-$. The solutions z_1 and z_2 are then

$$z_1 = \frac{1}{\sqrt{1 + 5x^2 + 6x^4 + 5x^6 + x^8}} \begin{bmatrix} -x_-(1 + 2x^2) \\ 1 - x^4 \\ x_+(2 + x^2) \end{bmatrix} \tag{5.2}$$

and

$$z_2 = \frac{1}{\sqrt{1 + 4x^2 + x^4}} \begin{bmatrix} x_-^2 \\ -2x_- \\ 1 \end{bmatrix}. \tag{5.3}$$

Observe that z_2 is an anti-instanton, as it should be. The Hermitian generator T is chosen to be

$$T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{5.4}$$

so that

$$z_0^{(\theta)} \equiv \frac{e^{\theta T} z_0}{|e^{\theta T} z_0|} = \alpha(x_+, x_-) \begin{bmatrix} c + s x_+^2 \\ x_+ \\ s + c x_+^2 \end{bmatrix}, \tag{5.5}$$

where $c \equiv \cosh\theta$, $s \equiv \sinh\theta$, and $\alpha(x_+, x_-)$ is the real function of x_+ and x_- such that $z_0^{(\theta)}$ is of unit length. The second member of the new family is

$$z_1^{(\theta)} \equiv \frac{P_+ z_0^{(\theta)}}{|P_+ z_0^{(\theta)}|} = \beta(x_+, x_-) \begin{bmatrix} s x_+(x^2 - 2) - c x_-(1 + 2x^2) \\ (2c^2 - 1)(1 - x^4) + 2cs(x_-^2 - x_+^2) \\ c x_+(2 + x^2) + s x_-(2x^2 - 1) \end{bmatrix} \tag{5.6}$$

and the new anti-instanton is

$$z_2^{(\theta)} \equiv \frac{P_+^2 z_0^{(\theta)}}{|P_+^2 z_0^{(\theta)}|} = \gamma(x_+, x_-) \begin{bmatrix} c x_-^2 - s \\ -2x_- \\ -s x_-^2 + c \end{bmatrix}, \tag{5.7}$$

where, again, the functions β and γ normalize the length of $z_1^{(\theta)}$ and $z_2^{(\theta)}$ to unity. It can be checked explicitly that these new $z_0^{(\theta)}$, $z_1^{(\theta)}$, and $z_2^{(\theta)}$ are indeed solutions of the equation of motion. The explicit form of $z_1^{(\theta)}$ can hardly be traced back to z_1. However, the form $z_2^{(\theta)} \equiv P_+^2 z_0^{(\theta)} / |P_+^2 z_0^{(\theta)}|$ happens to be precisely $z_2^{(\theta)} = e^{-\theta T} z_2 / |e^{-\theta T} z_2|$. At first sight, this might seem surprising, since the action defining $z_0^{(\theta)}$ is through $e^{\theta T}$, not $e^{-\theta T}$. However, this can be shown to hold in general. Indeed, using (4.24) for $P_n = \mathbb{1} - \Sigma_n$, one finds

$$\Delta_o^T P_n = -\{P_n, T\} + 2P_n T P_n \tag{5.8}$$

for the anti-instanton, instead of

$$\Delta_o^T P_0 = \{P_0, T\} - 2P_0 T P_0 \tag{5.9}$$

for the instanton. Hence the two vectors $z_0^{(\theta)}$ and $z_n^{(\theta)}$ are obviously orthogonal:

$$z_n^{(\theta)\dagger} z_0^{(\theta)} \propto (e^{-\theta T} z_n)^\dagger (e^{\theta T} z_0) = z_n^\dagger z_0 = 0,$$

as it should be. Finally, a direct calculation shows that neither in the instanton $z_0^{(\theta)}$ nor in the anti-instanton $z_n^{(\theta)}$ is there any common root (for any θ) in the three polynomial components.

As last remarks, we want to list a few open questions. The first to come to mind is whether a similar construction of the symmetries for the whole space of finite action solutions of the Grassmannian models could be calculated along the lines of the discussion of Sec. IV. The problem here is that there is no explicit description of the whole solution space for Grassmannian models other than for the $\mathbb{CP}^n$ models. Second, it would be interesting to have a geometrical explanation of why the larger group $\mathrm{Sl}(n+1, \mathbb{C})$ appears as a symmetry group of the field equations since only its real form $\mathrm{SU}(n+1)$ is an explicit symmetry. It probably originates in the Kählerian nature of the Grassmannian manifolds and from the fact that the self-dual equations are nothing but the Cauchy–Riemann equations for this complex structure. This conjecture allows one to hope that a similar result holds for any σ model whose field takes its values in a Kähler symmetric space. A third (more ambitious) question is to try to calculate the explicit action of the infinite symmetry algebra around solutions of other models. This task might be too difficult or even worthless for solutions on Minkowski space. Even though solitons known to exist for principal σ

models on Minkowski space bear some conceptual similarities with the Euclidean instantons, they are very different at the mathematical level. As pointed out by Uhlenbeck,[3] the unitarity condition (2.5) on Minkowski space $[R^{\dagger -1}(\bar{\lambda}) = R(\lambda)]$ does not have any finite power series solutions. Hence the collapse of the infinite-dimensional algebra to a finite algebra might be a peculiarity of the Euclidean sector. Nevertheless, the question of finding the explicit action of the symmetry algebra could be interesting for other models on Euclidean space. Finally, let us recall that, in the Kyoto school approach[5] of the Korteweg–de Vries and Kadomtsev–Petviashvili equations, an infinite-dimensional (affine) algebra also acts on the solution space. Moreover, this action defines, as in the present case, finite-dimensional orbits. It would be of prime interest to know if the finite orbits in both cases bear any similarity more than coincidental.

ACKNOWLEDGMENT

Supported in part by the Natural Sciences and Engineering Research Council of Canada and the "Fonds FCAR pour l'aide et le soutien à la recherche."

[1]L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981).
[2]Y.-S. Wu, Nucl. Phys. B **211**, 160 (1983).
[3]K. Uhlenbeck, "Harmonic maps into Lie groups (Classical solutions of the chiral model)," preprint of the Mathematics Department, University of Chicago, 1985.
[4]M. Jacques and Y. Saint-Aubin, J. Math. Phys. **28**, 2463 (1987).
[5]M. Sato, RIMS Kokyuroku **439**, 30 (1981); M. Jimbo and T. Miwa, Publ. RIMS, Kyoto Univ. **19**, 943 (1983).
[6]A. M. Perelomov, Physica D **4**, 1 (1981).
[7]A. M. Din and W. J. Zakrzewski, Nucl. Phys. B **174**, 397 (1980).
[8]H. Eichenherr and M. Forger, Nucl. Phys. B **155**, 381 (1979); **164**, 528 (1980).
[9]V. E. Zakharov and A. V. Mikhailov, Sov. Phys. JETP **47**, 1017 (1978).
[10]A. M. Din, Z. Horvath, and W. J. Zakrzewski, Nucl. Phys. B **233**, 269 (1984).
[11]P. Winternitz, "Lie groups and solutions of nonlinear partial differential equations," in *Lecture Notes in Physics*, Vol. 189 (Springer, New York, 1983); J. Harnad, Y. Saint-Aubin, and S. Shnider, Commun. Math. Phys. **93**, 33 (1984).
[12]M. J. Borchers and W. D. Garber, Commun. Math. Phys. **71**, 299 (1980); **72**, 77 (1980).

On the quantization of Poincaré and de Sitter gauge models

R. Aldrovandi and J. G. Pereira
Instituto de Física Teórica, Universidade Estadual Paulista, Rua Pamplona 145, 01405-São Paulo, SP, Brazil

(Received 29 September 1987; accepted for publication 17 February 1988)

The gauge model based on the Yang–Mills equations for the Poincaré group cannot be consistently quantized, at least in a perturbative approach. The regulated theory, obtained by adding the counterterms required by consistency and renormalizability, is just the gauge theory for a de Sitter group.

I. INTRODUCTION

Gauge theories for the Poincaré and de Sitter groups have been extensively studied as alternative theories for gravitation.[1] In this paper, "gauge theories" are to be considered as synonymous for models in which the field equations are the Yang–Mills equations for the group. That the gauge model for the Poincaré group could describe gravitation has already been shown elsewhere.[2] On the other hand, the quantization of such a model is expected from the start to face difficulties because of two peculiarities of the group: it is nonsemisimple and it acts on space-time itself. As a consequence of the first peculiarity, the Yang–Mills equations are not derivable from a Lagrangian.[2] As a result of the second peculiarity, all source fields belong, besides some tensor or spinor representation, to a "kinematic" representation whose generators are derivative fields on space-time. The number of derivatives appearing in currents and invariants is thereby augmented, representing a great threat to renormalizability. It will be shown here that such a model presents an inconsistency in the gauge field vertices, a problem that seems to stem from the absence of a Lagrangian. In order to illustrate what happens let us consider an unrealistic but instructive model. Suppose we did not know the Yukawa coupling Lagrangian $\mathscr{L}_I = g\varphi\Psi\bar{\Psi}$, but we had somehow arrived at the field equations in the form

$$\vec{D}\Psi = g\varphi\Psi, \tag{1.1}$$

$$\bar{\Psi}\overleftarrow{D} = -g\varphi\bar{\Psi}, \tag{1.2}$$

$$(\Box^2 + m^2)\varphi = g'\bar{\Psi}\Psi, \tag{1.3}$$

where $\vec{D} = i\gamma^\mu\partial_\mu - m$. Suppose further that we had some evidence (say, "experimental") that $g' \neq g$. This is a baffling situation from an intuitive point of view, but the problem can be made more definite if, ignoring the Lagrangian, we try to quantize the system by the Källén–Yang–Feldman (KYF) formalism.[3] The trouble is clear: as seen from the channels of Ψ and $\bar{\Psi}$, the coupling constant is g; as seen from the φ channel, it would be g'. The $\varphi\bar{\Psi}\Psi$ vertex obtained from Eqs. (1.1) and (1.2) would be different from that obtained from (1.3). This trivial remark points to a fundamental inconsistency of those equations, which are coherent only when $g = g'$. On the other hand, if we examine them in the light of Vainberg's theorem,[4] which gives necessary and sufficient conditions for the existence of a Lagrangian for a given set of equations, we find that $g = g'$ is necessary for (1.1)–(1.3) to be derivable from a Lagrangian.

We show in Sec. II, by using the KYF formalism, that this kind of inconsistency is present in the Yang–Mills equations for the Poincaré group.

The fact that the Poincaré group comes out as an Inönü–Wigner contraction limit of the de Sitter groups is exploited in Sec. III to provide more insight on the problem. The de Sitter groups being semisimple, a Lagrangian model can be built up, the path integral formalism may be used to supply the Feynman rules, and the Poincaré model is then seen as a limit case. The comparison of the de Sitter and Poincaré cases sheds some light on the way the inconsistencies, absent in the former, emerge in the latter. Geometrical considerations suggest that the de Sitter models can be viewed as smoothed versions of the Poincaré model.

Inconsistencies in field theories appear mainly when renormalization is involved, and sometimes find remedy in the addition of counterterms to the Lagragian, with consequent modifications in the field equations. A notorious example is the electrodynamics of scalar mesons, which only becomes renormalizable if a self-interaction term $\lambda\varphi^4$ is added to the purely electromagnetic Lagrangian. As here no Lagrangian is at hand, we may think of changing the equations directly. A study of the possibilities arising in this line of thought is given in Sec. IV, where, by combining requirements of vertex consistency and renormalizability, successive counterterms are introduced in the Yang–Mills equations. Curiously enough, the final well-behaved resulting theory is just a de Sitter gauge model, which in this way appears as a "functionally corrected" Poincaré model.

II. VERTEX INCONSISTENCY

The Poincaré Lie algebra is the semidirect product of the Lorentz algebra and the algebra of the translations in space-time. It is convenient to use the double index notation $J_{\alpha\beta}$ ($\alpha,\beta = 1,2,3,4$, with $\alpha < \beta$), for the Lorentz generators and to take J_α for the translation generators. Individual indices can be raised and lowered by the Minkowski metric $\eta_{\alpha\beta}$.

Taking $A^\alpha{}_{\beta\mu}$ and $B^\gamma{}_\mu$ as the gauge potentials related, respectively, to the Lorentz sector (which constitutes a gauge subtheory) and the translation sector, the corresponding field tensors turn out to be[2]

$$F^{\alpha\beta}{}_{\mu\nu} = \partial_\mu A^{\alpha\beta}{}_\nu - \partial_\nu A^{\alpha\beta}{}_\mu - gA^\alpha{}_{\gamma\mu}A^{\gamma\beta}{}_\nu + gA^\alpha{}_{\gamma\nu}A^{\gamma\beta}{}_\mu, \tag{2.1}$$

 0022-2488/88/061472-05$02.50

$$\tau^{\alpha}{}_{\mu\nu} = \partial_\mu B^{\alpha}_{\nu} - \partial_\nu B^{\alpha}{}_{\mu} - gA^{\alpha}{}_{\gamma\mu} B^{\gamma}_{\nu} + gA^{\alpha}{}_{\gamma\nu} B^{\gamma}{}_{\mu}. \quad (2.2)$$

The Yang–Mills equations for the Poincaré group are

$$\partial_\mu F^{\alpha\beta\mu\nu} - gA^{\alpha}{}_{\gamma\mu} F^{\gamma\beta\mu\nu} + gF^{\alpha}{}_{\gamma}{}^{\mu\nu} A^{\gamma\beta}\mu = gS^{\alpha\beta\nu}, \quad (2.3)$$

$$\partial_\mu \tau^{\alpha\mu\nu} - gA^{\alpha}{}_{\gamma\mu} \tau^{\gamma\mu\nu} + gF^{\alpha}{}_{\gamma}{}^{\mu\nu} B^{\gamma}{}_{\mu} = gl^2 \theta^{\alpha\nu}, \quad (2.4)$$

where $S^{\alpha\beta\nu}$ is the source spin density, $\theta^{\alpha\nu}$ is a source energy-momentum including coupling to the gauge fields, and l is the Planck length.

There is no Lagrangian density from which the above field equations can be derived.[2] It will be seen in Sec. IV that some pieces of this system of coupled equations can have Lagrangians, but the fact is that the whole system cannot. Attempts to redefine the fields so as to make the theory more tractable either disfigure its character by changing the meaning of the fundamental fields or make it trivial. For example, if the treatment used for the Korteweg–de Vries equation is applied here, the fields $B^{\alpha}{}_{\mu}$ must be some derivative $\partial_\mu \varphi^{\alpha}$, corresponding to the vacuum of the model.

In the absence of a Lagrangian, the natural way possibly open to quantization is the KYF formalism. It is convenient to use (2.1) and (2.2) in (2.3) and (2.4) so that equations acquire the form

$$\Box A^{\alpha\beta}{}_{\nu} - \partial_\nu (\partial^\mu A^{\alpha\beta}{}_{\mu}) = gV^{\alpha\beta}{}_{\nu}[A] - g^2 W^{\alpha\beta}{}_{\nu}[A] + S^{\alpha\beta}{}_{\nu}, \quad (2.5)$$

$$\Box B^{\alpha\nu} - \partial^\nu (\partial^\mu B^{\alpha}{}_{\mu}) = gU^{\alpha\nu}_{\beta\lambda}[A] B^{\beta\lambda} - g^2 Z^{\alpha\nu}_{\beta\lambda}[A] B^{\beta\lambda} + l^2 \theta^{\alpha\nu}, \quad (2.6)$$

where

$$V^{\alpha\beta}{}_{\nu}[A] = (A_{\gamma}{}^{\beta}{}_{\rho}\delta^{\alpha}_{\epsilon} - A^{\alpha}{}_{\epsilon\rho}\delta_{\gamma}{}^{\beta}) \times (\delta_{\nu}{}^{\rho}\partial^{\sigma} - \eta^{\rho\sigma}\partial_\nu) A^{\epsilon\gamma}{}_{\sigma}, \quad (2.7)$$

$$W^{\alpha\beta}{}_{\nu}[A] = (\delta_{\epsilon}{}^{\alpha}\eta^{\beta\gamma} - \delta_{\epsilon}{}^{\beta}\eta^{\alpha\gamma}) \times (\delta_{\nu}{}^{\sigma}\eta^{\lambda\rho} - \delta_{\nu}{}^{\lambda}\eta^{\sigma\rho}) A^{\epsilon}{}_{\varphi\rho} A^{\varphi}{}_{\delta\lambda} A^{\delta}{}_{\gamma\sigma}, \quad (2.8)$$

$$U^{\sigma\nu}_{\beta\lambda}[A] = \delta_\lambda{}^\nu (\partial_\mu A^{\alpha}{}_{\beta}{}^{\mu} + 2A^{\alpha}{}_{\beta\mu}\partial^\mu) - A^{\alpha}{}_{\beta}{}^{\nu}\partial_\lambda - A^{\alpha}{}_{\beta\lambda}\partial^\nu + \partial^\nu A^{\alpha}{}_{\beta\lambda} - 2\partial_\lambda A^{\alpha}{}_{\beta}{}^{\nu}, \quad (2.9)$$

$$Z^{\alpha\nu}_{\beta\lambda}[A] = A^{\alpha}{}_{\gamma\mu}(A^{\gamma}{}_{\beta}{}^{\mu}\delta_\lambda{}^\nu - 2A^{\gamma}{}_{\beta}{}^{\nu}\delta_\lambda{}^\mu) + A^{\alpha}{}_{\gamma}{}^{\nu} A^{\gamma}{}_{\beta\lambda}. \quad (2.10)$$

Gauge-fixing terms should be added to the left-hand side but they will not be important for the argument that follows.

Let us consider the sourceless case. To simplify the discussion, we shall rewrite (2.5) and (2.6) symbolically as

$$KA = gV[A] - g^2 W[A], \quad (2.11)$$

$$KB = gU[A]B - g^2 Z[A]B. \quad (2.12)$$

In the KYF formalism, we look for a perturbative solution in the form

$$A = \mathring{A} + gK^{-1}V[A] - g^2 K^{-1} W[A], \quad (2.13)$$

$$B = \mathring{B} + gK^{-1}U[A]B - g^2 K^{-1} Z[A]B. \quad (2.14)$$

Iteration to the desired order is then performed by replacing the A's and B's successively in terms of the free solutions $\mathring{A}$ and $\mathring{B}$. The operator K^{-1} represents a convolution with the Green's function of the differential operator in the lhs of (2.5) and (2.6) with Feynman boundary conditions.[5] We shall refer to K^{-1} simply as the Feynman propagator in some supposedly fixed gauge. The Feynman rules are obtained by projecting each one of these perturbative solutions on outgoing fields of the same kind. Each time they "hit" the free propagator, these outgoing fields produce free-fields of the same kind, so that the first contributions give precisely the basic vertices for the Feynman rules. In the case (2.11), such vertices are of the form $gAV[A]$ and $g^2 AW[A]$ and from them just the expected three-leg and four-leg vertices for a gauge model for the Lorentz group are obtained. Equation (2.11) is, of course, a set of coupled equations, one for each potential $A^{\alpha}{}_{\beta\mu}$. Take, for instance, the component $A^{1}{}_{2\mu}$. The projection is to be made on an outgoing field, $\mathring{A}^{1}{}_{2\mu}$, of exactly the same kind. Other potentials $\mathring{A}^{2}{}_{3\mu}$, $\mathring{A}^{3}{}_{0\mu}$, etc. appear in the vertices. In the equations for $A^{2}{}_{3\mu}$ and $A^{3}{}_{0\mu}$, the projections are made on outgoing fields $\mathring{A}^{2}{}_{3\mu}$ and $\mathring{A}^{3}{}_{0\mu}$, respectively. The important point is that the three-leg vertex involving $A^{1}{}_{2\mu}$, $A^{2}{}_{3\mu}$, and $A^{3}{}_{0\mu}$ will appear the same when obtained from each one of their respective equations. In other words, the expression for a vertex can be obtained from the equation related to any of its legs, and the result is independent of the choice of the leg. This general fact of perturbative field theory is easily found for (2.11), which are in reality the Yang–Mills equations for a Lorentz gauge model. Ghost fields could be introduced in principle through the old laborious Feynman method,[6,7] but (2.11) alone has a Lagrangian and in fact it would be simpler to pursue the whole treatment for the Lorentz sector by the path integration methods.

Now we come to the main point. The same considerations above, when applied to the whole set (2.11) and (2.12), lead to an insurmountable difficulty: vertices like $gB(\partial A)B$, $gBA(\partial A)$, and $g^2 BAAB$ do come out from (2.12) but not from (2.11). There are AB couplings in (2.12) but no field B appears in (2.11). Thus the expression for a vertex is no longer obtained from the equation for any of its legs, it is now dependent on the choice of the leg. With some freedom of language, we might say that the B's are able to "feel" the A's, but not the other way round. Or still, that vertices involving B's and A's are present for outgoing B's but not for outgoing A's. The same kind of inconsistency would appear in our defective Yukawa model [(1.1)–(1.3)] with $g = 0$ and $g' \neq 0$.

From this fundamental vertex inconsistency we conclude that, at least from the point of view of the KYF formalism, a model with (2.3) and (2.4) as field equations is not amenable to quantization.

III. RELATION TO DE SITTER MODELS

For usual gauge models, it is simpler to obtain the whole set of Feynman rules by the path integral approach and it will be instructive to examine our special case in the light of this standard procedure. It requires a Lagrangian, which is missing, but we can resort to the well-known fact that the Poincaré group P is an Inonü-Wigner contraction[8] of the two de Sitter (dS) groups.[9] As dS is semisimple, we can easily write down both the Yang–Mills equations and the corresponding Lagrangian for a dS gauge model. The com-

parison of the two cases will allow us to see why and where the procedure breaks down in the Poincaré model.

The relations between classical gauge models for P and dS have been studied in detail[2,10] and here we shall only recall the main points. In order to see what happens to gauge fields in the contraction process it is convenient to look at the contraction as acting on the group parameters ω^{ab} $(a,b=1,...,5,\ a<b)$. The parameters $\omega^{\alpha\beta}$ $(\alpha,\beta=1,...,4)$, related to the Lorentz subgroup, remain untouched. The parameters $\omega^{\alpha 5}$ represent "rotation" angles, compact or not, depending on the relative sign of $\eta_{\alpha\alpha}$ and η_{55}, where η_{ab} is the diagonalized dS invariant metric. Contraction requires redefining such angles as $L\omega^{\alpha 5}=a^{\alpha}$, where the a^{α} are the translation parameters and L is a length parameter taken to infinity in the contraction limit. A translation is thereby viewed as the limit of some infinitesimal rotation with an infinite radius. The dS generators J_{ab} obey

$$[J_{cd},J_{ef}]=-if^{ab}{}_{cd,ef}J_{ab}, \tag{3.1}$$

where

$$f^{ab}{}_{cd,ef}=\eta_{de}\delta_c^{[a}\delta_f^{b]}-\eta_{df}\delta_c^{[a}\delta_e^{b]}-\eta_{ce}\delta_d^{[a}\delta_f^{b]}+\eta_{cf}\delta_d^{[a}\delta_e^{b]}, \tag{3.2}$$

with $[ab]$ meaning antisymmetrization in the indices. If $A^{ab}{}_{\mu}$ are the gauge potentials for the dS gauge model, then $A^{\alpha\beta}{}_{\mu}$ remain the same through the contraction process, but $A^{\alpha 5}{}_{\mu}$ must be redefined so that $A^{\alpha 5}{}_{\mu}=L^{-1}B^{\alpha}{}_{\mu}$, where $B^{\alpha}{}_{\mu}$ is the translation gauge potential of the previous section. This can be checked, for example, by comparing the vacuum potentials $A^{\alpha 5}{}_{\mu}=\partial_{\mu}\omega^{\alpha 5}$ and $B^{\alpha}{}_{\mu}=\partial_{\mu}a^{\alpha}$. By the same process, if $F^{ab}{}_{\mu\nu}$ are the dS field strengths, the $F^{\alpha\beta}{}_{\mu\nu}$ become the field strengths (2.1) related to the Lorentz subgroup, while $\tau^{\alpha}{}_{\mu\nu}=LF^{\alpha 5}{}_{\mu\nu}$ become the translation field strengths (2.2). The Yang–Mills equations for the dS model,

$$\partial_{\mu}F^{ab\mu\nu}-gA^{a}{}_{c\mu}F^{cb\mu\nu}+gF^{a}{}_{c}{}^{\mu\nu}A^{cb}{}_{\mu}=0, \tag{3.3}$$

reduce exactly to the sourceless versions of (2.3) and (2.4) in the contraction limit $L\to\infty$, and the same happens to the corresponding Bianchi identities.

The contraction procedure has been frequently used to approach questions involving P,[11] mainly because it allows a point to point comparison to the better behaved dS group. It has been so in the demonstration of the nonexistence of a Lagrangian for the set of equations (2.3) and (2.4).[2] Equation (3.3) comes from the typical Lagrangian

$$\mathcal{L}=-\tfrac{1}{4}F^{ab}{}_{\mu\nu}F_{ab}{}^{\mu\nu}, \tag{3.4}$$

in which the algebra double indices are lowered and raised by the Cartan–Killing metric of dS. In the contraction limit, such a metric becomes degenerate and the field equations lose some terms. In particular, the cubic term in B, present in (3.3) and related to the four-leg vertex typical of gauge theories, is suppressed (as discussed below).

Path integral quantization can be performed without too much ado and Feynman rules of the usual kind are obtained for the dS model. For the P model, we start by making the substitution $A^{\alpha 5}{}_{\mu}=L^{-1}B^{\alpha}{}_{\mu}$ and follow the same procedure while keeping in mind what happens at the limit. Also, the ghost fields with $(\alpha 5)$ indices must be substituted in an analogous way, but as they will not be important for our central problem we shall not discuss them. In reality we shall concentrate on the inconsistency of the P model, leaving aside all the details having no bearing upon it. Once the substitution is made, (3.4) becomes

$$\mathcal{L}=-\tfrac{1}{4}[(F^{\alpha\beta}{}_{\mu\nu}+gL^{-2}B^{\alpha}{}_{\mu}B^{\beta}{}_{\nu})^2+L^{-2}(\tau^{\alpha}{}_{\mu\nu})^2]. \tag{3.5}$$

It is clear that the limit cannot be taken immediately: only the part

$$\mathcal{L}_{\mathcal{L}}=-\tfrac{1}{4}(F^{\alpha\beta}{}_{\mu\nu})^2, \tag{3.6}$$

corresponding to a Lorentz group gauge model, with only (2.3) for the field equations, would remain. As is frequently the case in the contraction formalism, we should first perform all the calculations and take the limit at the last step, although here we shall keep an eye on the relations to the field equations. Because it will be enough to make our point, we shall only examine in detail the three-field vertex: (3.5) is written as

$$\begin{aligned}\mathcal{L}=\mathcal{L}_{\mathcal{L}}-(L^{-2}/4)[&(\partial_{[\mu}B^{\alpha}{}_{\nu]})^2\\&+2gf_{\alpha\beta,\gamma 5,\epsilon 5}(\partial_{[\mu}A^{\alpha\beta}{}_{\nu]}B^{\gamma}{}_{\mu}B^{\gamma}{}_{\nu}B^{\epsilon}{}_{\nu}\\&+\partial_{[\mu}B^{\epsilon}{}_{\nu]}A^{\alpha\beta}{}_{[\mu}B^{\gamma}{}_{\nu]})+o(g^2)],\end{aligned} \tag{3.7}$$

where we have kept the dS structure constants (3.2).

We can obtain (2.3) and (2.4) from (3.5) simply by taking variations with respect to $A_{\alpha\beta\nu}$ and $B_{\alpha\nu}$, respectively, and then taking $L\to\infty$. An important point is that (2.4) is obtained with an overall factor L^{-2}, which cancels out. A consequence is that the contributions coming from the three-field terms in (3.7), proportional to L^{-2}, will remain in (2.4) but will be suppressed in (2.3). We see in this way how it happens that the BA coupling, present in (2.4), vanishes in (2.3), and find the same inconsistency of the previous section. The same happens to the terms A^2B^2 omitted in (3.7). The terms in B^4 have a L^{-4} factor and are totally suppressed.

Another consequence of (3.5) is that, once the $B^{\alpha}{}_{\mu}$ become (beside the $A^{\alpha\beta}{}_{\nu}$) the fundamental fields in substitution to the $A^{\alpha 5}{}_{\mu}$, the conjugate momenta become ill-defined. The vanishing of their time components is usual in a gauge theory, but here also the space components vanish: the momenta conjugate to $B^{\alpha}{}_{j}$ is $\pi^{\alpha}{}_{j}=L^{-2}\tau^{\alpha}{}_{j4}$, so that in the limit the canonical quantization is jeopardized.

In the Feynman rules for gauge models, the group dependence rests basically in the structure constants,[12] whose cyclic symmetry is used precisely to make the vertices symmetric in the external legs.[13] The cyclic symmetry is absent for nonsemisimple groups, which suggests that the inconsistency found here might be a common illness of all models involving such groups.

We have seen that, as long as we take the Yang–Mills equations as the very foundations of the theory, the Poincaré model is inevitably inconsistent. Let us forget the equations for a moment and use the contraction procedure to obtain a quantized theory. This amounts to taking (3.5) seriously and obtaining the resulting Feynman rules. The task is rather lengthy albeit standard. The results are simple and, once found, easily understood. Here we shall only describe the

main points of the resulting theory, trying to justify them by general arguments.

(i) The Lorentz sector constitutes a gauge subtheory, with the usual rules.

(ii) As seen in (3.5), the propagator of the B fields will be just the usual one, in some fixed gauge, times a factor L^2; the same applies to the corresponding ghosts.

(iii) Vertices are as usual, with the difference that each B leg (or corresponding ghost) gains a factor L^{-1} (an obvious consequence of the $A^{a5}{}_{\mu} \to L^{-1}B^{a}{}_{\mu}$ substitution).

Note that no final factor of L comes out from internal B lines in a diagram, since the L^2 factor in the propagator is just compensated by the L^{-1} factors in the two vertices connected. Graphs with external B legs will retain L^{-1} factors. However, if we calculate an S matrix element with N external B legs, the same L^{-N} factor will appear in each term in the perturbative series and, consequently, cancel out. Only when graphs with different numbers of external B legs are compared will the L^{-1} factors play a role.

The geometric setting for a P gauge model is best seen as an associated bundle, with Minkowski space as the base manifold and the fibers being tangent (also Minkowski) spaces on which the group acts. In the analogous setting for a dS model,[2] each tangent space is replaced by a dS space characterized by a length parameter L. When $L \to \infty$, each dS space approaches a tangent Minkowski space. If we use conformal coordinates[8] for each dS space, its points will be projected on a Minkowski space. In such coordinates, the natural dS group parameters are precisely ω^{ab} and a^{a}, and the gauge fields become naturally $A^{ab}{}_{\mu}$ and $B^{a}{}_{\mu}$. The quantized theory sketched above is in reality a dS model, viewed in conformal coordinates. To use an analogy, a dS model stands to a P model like a parabola to its asymptote, which is approached more and more when L becomes larger and larger, but it is never really attained. The dS model appears as a "smoothing" of the incongruous P model and seems to be its nearest quantizable theory. In Sec. IV we shall arrive again at a dS model from a rather different approach.

IV. CONSISTENCY AND LAGRANGIAN CHARACTER

Lagrangian theories do not exhibit the inconsistency described above. We could ask whether or not vertex consistency implies the presence of a Lagrangian or, in other words, whether only Lagrangian theories are quantizable in a coherent way. We shall not consider this very general question here. We shall restrict ourselves to Eqs. (2.3) and (2.4) in the sourceless case and proceed to a kind of naive patchwork, trying to see which terms should be dropped or added to make them into consistent equations. We find that every time they become consistent, they also become derivable from a Lagrangian.

We can start by simply dropping all terms coupling B to A in (2.4). The field equations become

$$\partial_{\mu}F^{ab\mu\nu} - gA^{a}{}_{c\mu}F^{cb\mu\nu} + gF^{a}{}_{c}{}^{\mu\nu}A^{cb}{}_{\mu} = 0, \tag{4.1}$$

$$\partial_{\mu}(\partial^{\mu}B^{a\nu} - \partial^{\nu}B^{a\mu}) = 0, \tag{4.2}$$

which are derivable from the Lagrangian $\mathcal{L} = -\frac{1}{4}(F^{ab}{}_{\mu\nu})^2 - \frac{1}{4}(\partial^{\mu}B^{a\nu} - \partial^{\nu}B^{a\mu})^2$. They are the field equations of gauge models for the Lorentz group $\mathcal{L}$ and for the Abelian translation group $T_{3,1}$. Their set would describe a model for the direct product $\mathcal{L} \otimes T_{3,1}$. Notice, however, that, as the fields $B^{a}{}_{\mu}$ are Lorentz vector fields, they should in reality couple to a Lorentz gauge potential. We take this into account by treating $B^{a}{}_{\mu}$ as a source field: usual derivatives are replaced by covariant ones and a source current appears in (4.1). As $B^{a}{}_{\mu}$ is a vector, it is its rotational that goes into the covariant derivative $\tau^{a}{}_{\mu\nu}$ given in (2.2). Also the divergence in (4.2) becomes covariant. Vertex consistency then fixes the source current, and the new equations are

$$\partial_{\mu}F^{ab\mu\nu} - gA^{a}{}_{c\mu}F^{cb\mu\nu} + gF^{a}{}_{c}{}^{\mu\nu}A^{cb}{}_{\mu} = g\tau^{ab\mu\nu}B^{b}{}_{\mu}, \tag{4.3}$$

$$\partial_{\mu}\tau^{ab\mu\nu} - gA^{a}{}_{c\mu}\tau^{cb\mu\nu} = 0. \tag{4.4}$$

These equations are derivable from $\mathcal{L} = -\frac{1}{4}F^2 - \frac{1}{4}\tau^2$, from which it can be checked that the source current in (4.3) is, as it should be, the spin density. We have been treating $B^{a}{}_{\mu}$ as "normal" vector fields with the canonical dimension (mass)^1. In reality, they have a defective dimension, as is clear from the redefinition $A^{a5}{}_{\mu} = L^{-1}B^{a}{}_{\mu}$ used in the contraction procedure. In order to correct this in the above equations, it is enough to add a factor L^{-1} to each $B^{a}{}_{\mu}$ field (and consequently to every $\tau^{a}{}_{\mu\nu}$). The only novelty will be a factor L^{-2} in the spin density.

We can now compare the resulting equations with the sourceless cases of (2.3) and (2.4); the only difference is the term $gF^{a}{}_{c}{}^{\mu\nu}B^{c}{}_{\mu}$ in (2.4). If we simply add this term to (4.4), vertex inconsistency comes out, but now we can relate it to a simple cause: such a term is obtained from a Lagrangian $\mathcal{L}' = -(g/2)F_{ab}{}^{\mu\nu}B^{a}{}_{\mu}B^{b}{}_{\nu}$ when variations are taken with respect to $B_{a\nu}$; however, $\mathcal{L}'$ should also contribute to (2.3) or (4.3) through its variations with respect to $A_{ab\nu}$. This contribution to (2.3) reestablishes vertex consistency. The new Lagrangian,

$$\mathcal{L} = -\tfrac{1}{4}F_{ab}{}^{\mu\nu}(F^{ab}{}_{\mu\nu} + 2gL^{-2}B^{a}{}_{\mu}B^{b}{}_{\nu}) - 4L^{-2}\tau_{a}{}^{\mu\nu}\tau^{a}{}_{\mu\nu}, \tag{4.5}$$

leads to a rather complicated theory. Then comes a beautiful point: this theory, as it is, is nonrenormalizable because of the graphs with four external B legs and exchange of two or more A's. When we look for the necessary counterterms, we find that $[-(g^2/4)B^{a}{}_{\mu}B^{b}{}_{\nu}B_{a}{}^{\mu}B_{b}{}^{\nu}]$ must be added to (4.5). This is quite natural for the four-legged graphs because they have a zero divergence degree. This situation is analogous to the case of scalar electrodynamics, where the renormalization of the higher order graphs with four external scalar legs, also of vanishing divergence degree, enforces the presence of a $\lambda\varphi^4$ term in the Lagrangian.[12] The addition of the B^4 term puts (4.5) into the form

$$\mathcal{L} = -\tfrac{1}{4}(F_{ab}{}^{\mu\nu} + gL^{-2}B_{a}{}^{\mu}B_{b}{}^{\nu})^2 - (L^{-2}/4)(\tau_{a}{}^{\mu\nu})^2. \tag{4.6}$$

This is the same Lagrangian as (3.5). The added B^4 term leads to a cubic term in (2.4), just that one we have seen suppressed by contraction in Sec. III. Therefore, summing up, by adding to (2.3) and (2.4) the terms necessary to wash out the vertex inconsistency, and then adding a last term to

make the model renormalizable, we arrive at a de Sitter theory.

V. FINAL COMMENTS

The absence of a Lagrangian is a most grievous flaw in a field theory. In the case considered here, the group contraction procedure can be used to show that the conjugate momenta of the translation gauge potentials are vanishing, so precluding a coherent canonical quantization. The existence of a Lagrangian for the Yang–Mills equation is closely related to the structure constants cyclic symmetry,[2] which fails for nonsemisimple groups. Such a symmetry is used to obtain the Feynman rules for gauge models,[13] which have consequently to be reexamined. We have seen that, for the Poincaré group, the very definition of a vertex becomes impossible and quantization, at least in a perturbative approach, unfeasible. The addition of counterterms required by consistency leads to an intricate theory. Interestingly enough, the addition of a B^4 term required by renormalizability turns the model into a gauge theory for the de Sitter group, which appears as the nearest coherently quantizable theory.

ACKNOWLEDGMENTS

The authors are very grateful to G. Francisco for discussions and a critical reading of the manuscript.

This work was partially supported by the Financiadora de Estudos e Projetos, Rio de Janeiro. R. A. was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasília. J. G. P. thanks the Fundação de Amparo à Pesquisa do Estado de São Paulo, for financial support.

[1]See, for example, F. W. Hehl, "Four Lectures on Poincaré Gauge Field Theory," in *Proceedings of the 6th Course of the International School of Cosmology and Gravitation on Spin, Torsion and Supergravity*, edited by P. G. Bergmann and V. de Sabbata (Plenum, New York, 1980); E. A. Ivanov and J. Niederle, Phys. Rev. D **25**, 976 (1982); D. Ivanenko and G. Sardanashvily, Phys. Rep. **94**, 1 (1983); E. A. Lord, J. Math. Phys. **27**, 3051 (1986); W. Drechsler, *ibid.* **26**, 41 (1985); R. J. MacKellar, *ibid.* **25**, 161 (1984).

[2]R. Aldrovandi and J. G. Pereira, Phys. Rev. D **33**, 2788 (1986).

[3]G. Källen, Ark. Fys. **2**, 187, 371 (1950); C. N. Yang and D. Feldman, Phys. Rev. **79**, 972 (1950).

[4]M. M. Vainberg, *Variational Methods for the Study of Nonlinear Operators* (Holden–Day, San Francisco, 1964); R. W. Atherton and G. M. Homsy, Stud. Appl. Math. **54**, 31 (1975); B. A. Finlayson, Phys. Fluids **15**, 963 (1972).

[5]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[6]R. P. Feynman, Acta. Phys. Pol. **24**, 697 (1963).

[7]B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).

[8]F. Gürsey, *Introduction to the de Sitter Groups, Lecture in the Istambul Summer School of Theoretical Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1962).

[9]M. Levy-Nahas, J. Math. Phys. **8**, 1211 (1967).

[10]R. Aldrovandi and E. Stédile, Int. J. Theor. Phys. **23**, 301 (1984).

[11]S. W. Mac Dowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).

[12]See, for instance, C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).

[13]P. Ramond, *Field Theory* (Benjamin/Cummings, Reading, MA, 1981).

The solution of the *s*-wave Bethe–Goldstone equation with a standard hard core potential

Michael K. Ong
Department of Mathematics, Bowdoin College, Brunswick, Maine 04011

(Received 19 May 1987; accepted for publication 27 January 1988)

This paper extends the results of an earlier article [J. Math. Phys. **27**, 1154 (1986)] to include the solution of the *s*-wave Bethe–Goldstone equation for the interaction of two nucleons characterized by a potential with an infinite repulsive core and a nearby attractive well. The solution again exhibits band-limiting behavior and is obtained in closed form via the prolate spheroidal wave functions. It is shown that the attractive part of the interaction potential perturbs the far-field scattered wave so that healing of the nucleon wave function is achieved only when the attractive part is weak. Finally, asymptotic results for the case of small core radius are also calculated.

I. INTRODUCTION

This paper is an extension of an earlier article, where we treated the *s*-wave Bethe–Goldstone equation for a pure hard core potential and obtained a closed form solution. The results are in excellent agreement with the approximate iterative solution found in the original paper by Bethe and Goldstone.[2] Here, we utilize the methods developed earlier and apply them to the solution of the Bethe–Goldstone equation for the so-called standard hard core potential of Moszkowski and Scott.[3] It is simply an interaction potential with a repulsive hard core plus an attractive well. Again, we seek only the *s*-wave solutions because they are the only ones that penetrate to small relative distances, where the effects of the singular potential are strongest. Any perturbative treatment of the problem is still inapplicable in this case since the matrix elements of the singular part of the interaction potential with respect to the independent particle wave functions are all divergent. The present problem exhibits band-limiting behavior and we are, once again, prompted to seek a solution in terms of prolate spheroidal wave functions.

As a brief refresher, consider two nucleons interacting in the Fermi sea according to the Brueckner independent pair model. In the center-of-mass coordinate system, the two-nucleon wave function $\psi(\mathbf{r})$ satisfies the Schrödinger equation[4–6]

$$[(\hbar^2/m^*)\Delta + \epsilon]\psi(\mathbf{r}) = Q_F v(r)\psi(\mathbf{r}) , \tag{1.1}$$

with relative coordinates $\mathbf{r}$; effective mass m^*; and the Pauli projection operator Q_F, which effectively removes from $v(r)\psi(\mathbf{r})$ those Fourier components with relative momentum $k < k_F$, the Fermi momentum.

The projection operator, defined by

$$Q_F = \sum_{\lambda\mu} |\lambda\mu\rangle\langle\lambda\mu| , \tag{1.2}$$

is idempotent, i.e., $Q_F^2 = Q_F$. Consequently,[7] its spectrum is $\sigma(Q_F) = \{0,1\}$, which, when translated in terms of the Fermi distribution, is

$$Q_F|\lambda\mu\rangle = \begin{cases} |\lambda\mu\rangle, & k > k_F, \\ 0, & \text{otherwise} . \end{cases} \tag{1.3}$$

If we consider only *s*-wave solutions in the form $\psi(r) = r^{-1}u(r)$, Eq. (1.1), together with Eq. (1.3), can be transformed into the scalar integrodifferential equation (2.1).[1,2] This will be the focal point of the present paper.

We have demonstrated in the present paper that the far-field scattered wave is perturbed, the severity of which depends on the strength of the attractive part of the interaction potential, contrary to the normal plane wave assumption usually adopted in most nuclear matter calculations. Modulation of the scattered wave is slight only, however, when the attractive part is weak.

We have also shown that the contributions of the hard core and the attractive part can be distinctively separated. In short, the present paper concentrates only on the effects of the attractive part of the interaction potential. To achieve consistency with our earlier notations,[1] all new terms representing the contributions from the attractive well are indicated by the superscript asterisk(s). Finally, in some of the present sections the calculations are quite involved. In order to preserve the continuity of the discussions, we have delegated some of these calculations to the Appendixes.

This paper is organized as follows: In Sec. II, we present the *s*-wave Bethe–Goldstone equation for the standard hard core potential. Section III briefly summarizes some useful results concerning band-limited and prolate spheroidal wave functions. A closed form solution of the Bethe–Goldstone equation is given in Sec. IV. In Sec. V, we demonstrate that an approximation for small core radius c (the case of most physical significance) can be obtained quite readily. Section VI is concerned with the far-field scattering problem, wherein we show that the scattered wave is perturbed. Finally, an approximation of the normalization constant for small c is calculated in Sec. VII.

II. THE *s*-WAVE EQUATION WITH THE STANDARD HARD CORE POTENTIAL

Consider the dimensionless integrodifferential equation[1]

$$\left(\frac{d^2}{dr^2} + K^2\right)u(r) = v(r)u(r) - \int_0^\infty \chi(r,r')v(r')u(r')dr' , \tag{2.1a}$$

 0022-2488/88/061477-09$02.50

where the kernel is

$$\chi(r,r') = (1/\pi) \times [\sin(r-r')/(r-r') - \sin(r+r')/(r+r')] . \tag{2.1b}$$

This will be referred to as the s-wave Bethe–Goldstone equation.

The interaction potential $v(r)$ to be considered is the phenomenological standard hard core potential of Moszkowski and Scott,[3] with a single bound state at zero energy and an effective range of 2.5 fm (fermi) defined by (dimensionless)

$$v(r) = \begin{cases} +\infty, & r<c, \\ -v_0 e^{-\mu(r-c)}, & r>c, \end{cases} \tag{2.2}$$

where v_0, μ, and c are dimensionless quantities consistent with Eq. (2.1). More precisely, $c = k_F \tilde{c}$, $\mu = \tilde{\mu}/k_F$, and $v_0 = 2m^*_{red}\tilde{v}_0/\hbar^2 k_F^2$, with m^*_{red} the reduced nucleon effective mass. The original parameters of Moszkowski and Scott[3] are $\tilde{c} = 0.4$ fm, $\tilde{\mu} = 2.084$ fm^{-1}, and $\tilde{v}_0 = 260$ MeV. In the following analysis, it should be clear that all calculations are done in terms of dimensionless quantities.

This more realistic choice of potential, although it oversimplifies the actual nucleon–nucleon force, is a one-step improvement of the grossly simplistic pure hard core potential. The attractive exponential well is chosen to be rather weak so that it does not contain large momentum components. This guarantees that the two-nucleon wave function cannot be "bent" to produce more than one bound state. The expression "the attractive part is *rather* weak" is then taken to mean that the attractive part of the interaction potential does not cause appreciable modulation of the two-nucleon wave function after scattering. This, as will be seen later, is essential to the convergence of any nuclear matter calculations using this potential. A similar observation, albeit based on a different approach from this paper, is discussed in Refs. 4 and 6. The criterion for weakness is determined in the discussion following Eq. (6.5), wherein it is shown that the depth of the attractive potential $\tilde{v}_0$ must be much smaller than the effective kinetic energy of the nucleon pair in the Fermi sea.

The two-nucleon wave function $u(r)$ and its slope vanish inside the hard core and are finite elsewhere. Furthermore, in order to obtain a nontrivial solution of Eq. (2.1), it is necessary for the slope to be discontinuous at the core boundary $r = c$. Consequently, the product $v(r)u(r)$ has a δ-function discontinuity at $r = c$ and must remain finite for $r > c$. Finally, to include the attractive part of the potential, we write

$$v(r)u(r) = A\,[\delta(r-c) - v_0 e^{-\mu(r-c)}\theta(r-c)] + \omega(r)\theta(c-r) , \tag{2.3}$$

where $\theta(\cdot)$ is the Heaviside step function, $\delta(\cdot)$ is the Dirac delta function, and A is the normalization constant to be determined approximately from the condition that the wave function asymptotically approaches the unperturbed free-particle wave function so that there is no s-wave phase shift. This condition is

$$\lim_{r\to+\infty} [u(r)/r] \to j_0(Kr) , \tag{2.4}$$

where j_0 is the spherical Bessel function of zeroth order.

This important stipulation [(2.4)] is a consequence of the Pauli projection operator Q_F which effectively blocks any interaction between nucleons at large separation. In effect, the nucleons return to their independent particle states before the next collision occurs and the two-nucleon wave function is said to have "healed." However, it will be shown in Sec. VI that, unlike the pure hard-core problem, this is only true when the attractive part of the interaction potential is very weak. Otherwise, healing will not be complete and there will be an appreciable phase shift. The strength of the attractive well is embodied in the number v_0. In Sec. V [cf. Eqs. (5.7) and (5.8f)], the wave function is calculated with an explicit corrective term due to v_0. The overall normalization constant A is also calculated in Sec. VII with corrective terms arising from v_0 [cf. Eq. (7.8)]. It is interesting to note that this "healing" phenomenon is related mathematically to the fact that the kernel $\chi(r,r')$ has no singularity.

A few words of clarification are necessary at this point. In the absence of the attractive part of the interaction potential, we have $v_0 \equiv 0$ so that Eq. (2.3) reduces to the same expression used in Ref. 1, in agreement with the original work of Bethe and Goldstone.[2] It has been shown that for this pure hard core problem, the asymptotic limit (2.4) is reached very rapidly, resulting in virtually no scattering and hence no modulation of the wave function for large separation distance r.[1,2,4,6] The healing of the wave function for large r is complete in this case. A healing distance has also been calculated[5] and was shown to be less than the internucleon distance so that, in essence, the nucleons return to their independent-particle wave function before the next collision takes place. In the present problem, however, due to the presence of the attractive well, we do not expect complete healing. This means that, depending on the strength of the attractive well, the far-field scattered wave may not quite reach the plane wave limit, viz. $j_0(Kr)$. In fact, as demonstrated by the result of Eq. (6.5), healing will be complete only if the term involving v_0 is negligibly small. The wave function in this case is said to be "wounded." Gomez *et al.*,[4] using very crude plane wave approximations, have calculated the phase shift due to the attractive part of the potential (modeled as a weak square well); it turns out to be significantly small for large separation distance. A similar discussion concerning the weakness of the square well potential can also be found in Ref. 8. The asymptotic limit (2.4) is therefore used only as an approximation in calculating the normalization constant A under the condition that the attractive part is weak. This is entirely consistent within the framework of Brueckner's independent pair theory.

The extra contribution $\omega(r)$ in Eq. (2.3) is nonvanishing only inside the core and is to be determined from the condition that $u(r)$ vanishes inside the hard core. From Eqs. (2.1) and (2.3), this translates to

$$\omega(r) - \int_0^c \chi(r,r')\omega(r')dr'$$
$$= A\chi(r,c) - v_0 A \int_c^\infty \chi(r,r') e^{-\mu(r'-c)}\, dr', \quad r<c, \tag{2.5}$$

subject to the condition that

$$\omega(r) = \begin{cases} 0, & r>c, \\ \text{finite}, & r<c. \end{cases} \tag{2.6}$$

Equation (2.5) is a Fredholm integral equation of the second kind and can be solved for the unknown $\omega(r)$ once the integral

$$I(r,c) \equiv \int_c^\infty \chi(r,r') e^{-\mu(r'-c)}\, dr' \tag{2.7}$$

is evaluated. Clearly, the integral $I(r,c)$ accounts for the contributions due to the attractive part of the potential. This innocent looking term considerably complicates the calculations to follow.

The equation for a pure hard core, i.e., for $v_0 \equiv 0$, was originally solved by Bethe and Goldstone using an approximate iterative procedure.[2] The approximate solution is valid only for small c. Recognizing that condition (2.6) implies that the function $\omega(r)$ is band limited, we have solved the corresponding hard core equation in closed form using prolate spheroidal wave functions. Here, the contribution $\omega(r)$ remains band limited; this suggests that Eq. (2.5) can also be solved using the method developed in Ref. 1.

To facilitate the solution of the integral equation (2.5) via band-limited functions, it is first necessary to extend the equation into the region $-c<r<0$. This requires the extension of $\omega(r)$ and the nonhomogeneous part $H(r,c)$, defined by

$$H(r,c) \equiv \chi(r,c) - v_0 I(r,c), \tag{2.8}$$

as an odd function of r in the interval $-c<r<0$, i.e.,

$$\omega(r) = \begin{cases} \omega(r), & 0<r<c, \\ -\omega(-r), & -c<r<0; \end{cases} \tag{2.9a}$$

$$H(r,c) = \begin{cases} H(r,c), & 0<r<c, \\ -H(-r,c), & -c<r<0. \end{cases} \tag{2.9b}$$

In doing so, Eq. (2.5) becomes

$$\omega(r) - \frac{1}{\pi}\int_{-c}^{c} \frac{\sin(r-r')}{r-r'}\omega(r')dr' = AH(r,c), \quad |r|<c, \tag{2.10}$$

subject to the condition

$$\omega(r) = \begin{cases} 0, & |r|>c, \\ \text{finite}, & |r|<c. \end{cases} \tag{2.11}$$

The integral equation (2.10) is of the same form obtained in Ref. 1, but with a very different nonhomogeneous term $H(r,c)$. The problem at hand is still of the same band-limiting variety as before and its solution can be obtained similarly using prolate spheroidal wave functions. In Ref. 1 we have presented a summary of the theory concerning the band-limited solutions of the homogeneous integral equation

$$\lambda f(t) = \int_{-1}^{1} f(s) \frac{\sin c(t-s)}{\pi(t-s)} ds, \quad |t|<1, \tag{2.12}$$

via prolate spheroidal wave functions. This topic will be touched upon briefly in Sec. III.

III. SUMMARY OF KNOWN RESULTS ON BAND-LIMITED FUNCTIONS AND PROLATE SPHEROIDAL FUNCTIONS

The following theorem summarizes the known results regarding the eigenvalues and eigenfunctions of the integral equation (2.12).

Theorem: For $f \in L_2(-1,1)$ the integral operator

$$\mathbf{K}f = \int_{-1}^{1} \frac{\sin c(t-s)}{\pi(t-s)} f(s) ds,$$

with a continuous and symmetric kernel for all $-1<t, s<1$, is a positive self-adjoint compact operator in $L_2(-1,1)$ so that the integral equation

$$\mathbf{K}f = \lambda f, \quad |t|<1, \tag{3.1}$$

has a denumerable set of eigenvalues

$$1>\lambda_0>\lambda_1>\cdots>0,$$

to which each λ_i is an associated real-valued eigenfunction $f_i(t)$, which forms a complete set in $L_2(-1,1)$ and which satisfies the orthogonality condition

$$\text{(i)} \quad \int_{-1}^{1} f_i(t) f_j(t) dt = \lambda_i \delta_{ij} \tag{3.2a}$$

and the eigenvalue problem

$$\text{(ii)} \quad \lambda_i f_i(t) = \int_{-1}^{1} f_i(s) \frac{\sin c(t-s)}{\pi(t-s)} ds, \quad |t|<1. \tag{3.2b}$$

The eigenfunctions $f_i(t)$ are also bounded continuous solutions of the differential operator

$$L(\cdot) \equiv \left\{ \frac{d}{dt}(t^2-1)\frac{d}{dt} + c^2 t \right\}(\cdot) \tag{3.3}$$

satisfying the eigenvalue problem

$$Lf_i = \chi_i f_i, \quad -\infty<t<\infty,$$

with eigenvalues $0<\chi_0<\chi_1<\cdots$.

The eigenfunctions $f_i(t)$ are band limited and expressible in terms of prolate spheroidal wave functions of zeroth order, $S_{0n}(c,t)$, $n \in \mathbb{Z}^+$. The prolate spheroidal wave functions form a complete set in $L_2(-1,1)$ and are odd or even functions of t according to whether n is odd or even.

For fixed $n \in \mathbb{Z}^+$ and small c, we have

$$\lambda_n(c) = \frac{2}{\pi}\left[\frac{2^{2n}(n!)^3}{(2n)!(2n+1)!}\right]^2 c^{2n+1} \times \exp\left\{ -\frac{(2n+1)c^2}{(2n-1)^2(2n+3)^3}[1+O(c^4)] \right\}, \tag{3.4a}$$

$$S_{0n}(c,t) = P_n(t) + c^2\left[\frac{n(n-1)}{2(2n-1)^2(2n+1)} P_{n-2}(t) - \frac{(n+1)(n+2)}{2(2n+3)^3(2n+1)} P_{n+2}(t)\right] + O(c^4), \tag{3.4b}$$

where the P_n's are the Legendre polynomials so that, in the limit as $c \to 0$,

$$\chi_n(0) = n(n+1), \quad n \in \mathbb{Z}^+, \tag{3.5a}$$

$$S_{0n}(c,t) \to P_n(t). \tag{3.5b}$$

□

The theory of prolate spheroidal wave functions appears in several sources. The Theorem summarizes only selected results pertinent to this paper. A list of references can be found in Ref. 1.

IV. SOLUTION OF THE BETHE–GOLDSTONE EQUATION

After the change of variables

$$r = ct, \quad r' = cs, \quad A^{-1}\omega(r) = f(t), \quad H(r,c) = g(t), \tag{4.1}$$

Eq. (2.10) becomes

$$f(t) - \int_{-1}^{1} f(s) \frac{\sin c(t-s)}{\pi(t-s)} ds = g(t), \quad |t| < 1. \tag{4.2}$$

Since **K** is symmetric and compact, by the Hilbert–Schmidt theorem the solution of Eq. (4.2) can be expanded in terms of the prolate spheroidal wave functions in an absolutely and uniformly convergent series

$$f(t) = g(t) + \sum_{n=1}^{\infty} \left(\frac{\lambda_n}{1-\lambda_n} \right) \langle g(t'), f_n(t') \rangle f_n(t), \tag{4.3}$$

where the eigenfunctions are normalized as follows:

$$f_n(t) = [\lambda_n(c)/u_n(c)]^{1/2} S_{0n}(c,t),$$
$$u_n^2(c) = \int_{-1}^{1} [S_{0n}(c,t)]^2 dt. \tag{4.4}$$

The expansion coefficients can be calculated as

$$\langle g(t), f_n(t) \rangle = \int_{-1}^{1} g(t) f_n(t) dt = \mathrm{I} - v_0 \mathrm{II}, \tag{4.5a}$$

where

$$\mathrm{I} = \begin{cases} (2/c)\lambda_n(c) f_n(1), & n \text{ odd}, \\ 0, & n \text{ even}, \end{cases} \tag{4.5b}$$

and

$$\mathrm{II} = \int_{-1}^{1} I(ct,c) f_n(t) dt = \begin{cases} \dfrac{e^{\mu c}}{\pi} \displaystyle\int_{-1}^{1} e^{-\mu c t} \operatorname{Im}\{E_1[(\mu - i)c(1-t)]\} f_n(t) dt, & n \text{ odd}, \\ 0, & n \text{ even}. \end{cases} \tag{4.5c}$$

In arriving at I and II, we have used the parity $f_n(-t) = (-1)^n f_n(t)$ and the fact that $g(t)$ is an odd function of t. The quantity I, representing the contribution of the hard core, was calculated in Ref. 1. Again, it is clear that the quantity II is due to the contribution from the attractive part of the potential. A derivation of II can be found in Appendix A.

Therefore, the complete solution of the integral equation (4.2) is

$$f(t) = g(t) + \sum_{n \text{ odd}} \left[\frac{2\lambda_n^2}{c(1-\lambda_n)} f_n(1) - v_0 \frac{e^{\mu c}}{\pi} \times \operatorname{Im}\left\{ \int_{-1}^{1} e^{-\mu c t'} E_1[(\mu - i)c(1-t')] \times f_n(t') dt' \right\} \right] f_n(t), \quad |t| < 1. \tag{4.6}$$

In terms of the functions $S_{0n}(c,t)$, we have

$$f(t) = g(t) + \sum_{n \text{ odd}} \left[\frac{2\lambda_n^3}{c(1-\lambda_n)u_n} S_{0n}(c,1) - \frac{\lambda_n^2}{\pi(1-\lambda_n)u_n} v_0 e^{\mu c} \operatorname{Im}\{I^*(c)\} \right] S_{0n}(c,t), \quad |t| < 1, \tag{4.7}$$

where

$$I^*(c) \equiv \left\{ \int_{-1}^{1} e^{-\mu c t'} E_1[(\mu - i)c(1-t')] f_n(t') dt' \right\}. \tag{4.8}$$

This is the complete closed form solution of the integral equation for the extra contribution $\omega(r)$. In principle, an approximation can be obtained from Eq. (4.7) for any desired order of c. Unfortunately, the integral $I^*(c)$ cannot be integrated analytically in closed form. We shall present an approximation for small c in Appendix B.

The two-nucleon wave function vanishes inside the hard core so that the integrodifferential equation (2.1) becomes

$$\left(\frac{d^2}{dr^2} + K^2 \right) u(r) = A[\delta(r-c) - \chi(r,c) - v_0 e^{-\mu(r-c)}] - \int_0^c \chi(r,r')\omega(r') dr' - A v_0 \int_c^{\infty} \chi(r,r') e^{-\mu(r'-c)} dr', \quad r > c,$$
$$\equiv F(r) - F^*(r), \quad r > c, \tag{4.9}$$

where

$$F(r) \equiv A[\delta(r-c) - \chi(r,c)] - \int_0^c \chi(r,r')\omega(r') dr' \tag{4.10a}$$

and

$$F^*(r) \equiv A v_0 e^{-\mu(r-c)} - A v_0 \int_c^\infty \chi(r,r') e^{-\mu(r'-c)}\, dr' . \tag{4.10b}$$

With the requirement that $u(0) = 0$, Eq. (4.9) has the solution

$$u(r) = \frac{1}{K} \int_0^r [F(s) - F^*(s)] \sin K(r-s) ds, \quad r > c . \tag{4.11}$$

The closed form solution of the (dimensionless) s-wave Bethe–Goldstone equation for the standard hard core potential is now complete apart from the normalization constant A.

Thus far, no approximation schemes have been invoked and the solution of the problem is exact. It may be mathematically aesthetic at this point, but all the subtleties of the physics invoked remain to be seen. In Sec. V we shall seek for an approximate solution for the small core radius; some interesting observations will also be presented in Secs. VI and VII.

V. APPROXIMATION FOR SMALL CORE RADIUS

Since the typical model of the nucleus exhibits only a short-range hard-core potential, it is necessary to analyze the problem only for the case of small core radius c. In terms of the original interval $0 < r < c$, Eq. (4.7) becomes

$$A^{-1}\omega(r) = \chi(r,c) + \sum_{n \text{ odd}} [B'_n - B''_n \operatorname{Im}\{I^*(c)\}] \times S_{0n}\left(c, \frac{r}{c}\right), \quad 0 < r < c , \tag{5.1a}$$

where

$$B'_n \equiv [2\lambda_n^3 / c(1-\lambda_n) u_n(c)] S_{0n}(c,1) \tag{5.1b}$$

and

$$B''_n \equiv [\lambda_n^2 / (1-\lambda_n) u_n(c)] v_0 e^{\mu c} . \tag{5.1c}$$

The integral in Eq. (4.10b) is simply $I(r,c)$ given in Eq. (2.7); thus

$$A^{-1} F^*(r) = v_0 e^{-\mu(r-c)} + v_0 I(r,c) . \tag{5.2}$$

For small c, we can crudely approximate $I(r,c)$ as follows:

$$\begin{aligned} I(r,c) &\equiv \int_c^\infty \chi(r,r') e^{-\mu(r'-c)}\, dr' \\ &= e^{\mu c} \int_c^\infty e^{-\mu r'} \frac{\sin(r-r')}{r-r'}\, dr' \\ &\quad - e^{\mu c} \int_c^\infty e^{-\mu r'} \frac{\sin(r+r')}{r+r'}\, dr' \\ &= e^{\mu(c-r)} \int_{c-r}^\infty e^{-\mu s} \frac{\sin s}{s}\, ds \\ &\quad - e^{\mu(c+r)} \int_{c+r}^\infty e^{-\mu s} \frac{\sin s}{s}\, ds \\ &\sim e^{\mu(c-r)} \tan^{-1}(1/\mu) - e^{\mu(c+r)} \tan^{-1}(1/\mu), \\ &\quad \text{for } \textit{small } c, \quad r < c, \\ &= -2 e^{\mu c} \sinh(\mu r) \tan^{-1}(1/\mu) , \\ &\quad \text{for } \textit{small } c, \quad r < c . \end{aligned} \tag{5.3}$$

Hence, Eq. (5.2) approximately becomes

$$A^{-1} F^*(r) \sim v_0 e^{\mu c} [e^{-\mu r} - 2 \sinh(\mu r) \tan^{-1}(1/\mu)] , \quad \text{for } \textit{small } c, \quad r < c . \tag{5.4}$$

Upon substituting $\omega(r)$ from Eq. (5.1a) and using the identity

$$\int_0^c \chi(s,s') S_{0n}\left(c, \frac{s'}{c}\right) ds' = \lambda_n S_{0n}\left(c, \frac{s}{c}\right), \tag{5.5}$$

the quantity $F(r)$ in Eq. (4.10a) becomes

$$F(r) = A[\delta(r-c) - \chi(r,c)] - \int_0^c \chi(s,s')\chi(s',c) ds' - \sum_{n \text{ odd}} [B_n - B_n^* \operatorname{Im}\{I^*(c)\}] S_{0n}\left(c, \frac{s}{c}\right), \tag{5.6}$$

where

$$B_n \equiv \lambda_n B'_n, \quad B_n^* \equiv \lambda_n B''_n .$$

The above expression is identical to the one obtained in Ref. 1, but with an additional term B_n^*.

With the above expressions, the two-nucleon wave function from Eq. (4.11) becomes

$$\begin{aligned} u(r) &= \frac{1}{K} \int_0^r [F(s) - F^*(s)] \sin K(r-s) ds \\ &= (A/K)\,(\mathrm{I} - \mathrm{II} + \mathrm{III} + \mathrm{III}^* - \mathrm{III}^{**}) , \end{aligned} \tag{5.7}$$

with

$$\mathrm{I} = \int_0^r [\delta(s-c) - \chi(s,c)] \sin K(r-s) ds , \tag{5.8a}$$

$$\mathrm{II} = \int_0^r \left[\int_0^c \chi(s,s')\chi(s',c) ds' \right] \sin K(r-s) ds \sim \frac{4c^4}{27\pi^3} \left(\frac{r}{k} - \frac{\sin Kr}{K^2} \right), \tag{5.8b}$$

$$\mathrm{III} = \sum_{n \text{ odd}} \int_0^c S_{0n}\left(c, \frac{s}{c}\right) \sin K(r-s) ds \sim O(c^{10}) . \tag{5.8c}$$

The analysis for the above quantities has been done in Ref. 1.

The contributions of the attractive part of the potential to the wave function are

$$\mathrm{III}^* = \sum_{n \text{ odd}} B_n^* \operatorname{Im}\{I^*(c)\} \times \int_0^c S_{0n}\left(c, \frac{s}{c}\right) \sin K(r-s) ds , \tag{5.8d}$$

$$\mathrm{III}^{**} = A^{-1} \int_0^r F^*(s) \sin K(r-s) ds . \tag{5.8e}$$

The integral in III* can be integrated approximately for small c by using the limiting value of S_{0n} in Eq. (3.4b). After some work and using the results of the Theorem, it can be shown that

$$\mathrm{III}^* \sim O(c^9) .$$

For small c, the integral III**, upon using Eq. (5.3), becomes

$$\mathrm{III}^* \sim [v_0 e^{\mu c}/(\mu^2 + K^2)]\,\{[\mu \sin(Kr) - K \sinh(\mu r)] - 2\tan^{-1}(1/\mu) \times [\mu \sin(Kr) - K\cos(Kr) - Ke^{-\mu r}]\}\,. \quad (5.8f)$$

Equations (5.8c) and (5.8f) show that the contributions of III and III* are negligible and that the major contributions to the wave function come from I and II, which represents the hard core, and from III**, which represents the attractive well. The contribution III** represents the corrective term to the two-nucleon wave function in the presence of the attractive well. Without the attractive part of the potential, i.e., $v_0 \equiv 0$, the major contributions come from I and II and the result is identical to the pure hard-core problem solved in closed form in Ref. 1.

VI. FAR-FIELD SCATTERING

In this section, we analyze the asymptotic behavior of the two-nucleon wave function for large separation distance r. The two-nucleon wave function (4.11) can be rewritten as

$$u(r) = \frac{1}{K}\int_0^r [F(s) - F^*(s)]\sin K(r-s)ds = \frac{\sin Kr}{K}\int_0^r [F(s) - F^*(s)]\cos Ks\,ds - \frac{\cos Kr}{K}\int_0^r [F(s) - F^*(s)]\sin Ks\,ds, \quad r > c\,. \quad (6.1)$$

The far-field scattering condition (2.4) requires that

$$\lim_{r\to+\infty}\int_0^r [F(s) - F^*(s)]\sin Ks\,ds \to 0 \quad (6.2a)$$

and

$$\lim_{r\to+\infty}\int_0^r [F(s) - F^*(s)]\cos Ks\,ds \to 1\,. \quad (6.2b)$$

The condition (2.4) follows from the requirement that the two-nucleon wave function $u(r)$ asymptotically approaches the unperturbed free-particle wave function for large separation distance r. However, it will be shown in the following that this is not entirely true unless the attractive exponential well is rather weak so that it does not appreciably perturb the two-nucleon wave function for large separation distance r. We have also pointed out earlier that in the presence of the attractive well, healing of the wave function need not be complete. The following analysis demonstrates it.

In Ref. 1, we have shown that for small c,

$$\lim_{r\to+\infty}\int_0^r F(s)\sin Ks\,ds \to 0\,. \quad (6.3)$$

Let us now consider the limiting value for $F^*(s)$. Upon using Eq. (4.10b), we have

$$\begin{aligned}\lim_{r\to+\infty}\int_0^r F^*(s)\sin Ks\,ds &= \lim_{r\to+\infty} Av_0 e^{\mu c}\int_0^r \left[e^{-\mu s} + \int_c^\infty e^{-\mu r'}\chi(s,r')dr'\right] \times \sin Ks\,ds \\ &= Av_0 e^{\mu c}\left(\frac{K}{\mu^2+K^2}\right) + Av_0 e^{\mu c}\lim_{r\to+\infty}\int_c^\infty e^{-\mu r'} \times \left[\int_0^r \chi(s,r')\sin Ks\,ds\right]dr', \\ &= \frac{Av_0 e^{\mu c}}{\mu^2+K^2}[K(1+\cos\mu c) + \mu\sin Kc]\,, \quad 0 \leqslant K < 1\,, \end{aligned} \quad (6.4)$$

where the interchange in the order of integration on the second line is justified since the integrals involved are uniformly convergent.[9] The integral between the square brackets is calculated in Appendix C.

Combining Eqs. (6.3) and (6.4), we arrive at the interesting result

$$\lim_{r\to+\infty}\int_0^r [F(s) - F^*(s)]\sin Ks\,ds \to \frac{Av_0 e^{\mu c}}{\mu^2+K^2}[K(1+\cos\mu c) + \mu\sin Kc]\,, \quad 0 \leqslant K < 1\,, \quad (6.5)$$

which shows that the plane wave limit is not reached and the far-field scattered wave is therefore perturbed. This is an important observation because it is assumed customarily in most nuclear matter calculations that the nucleons return to their independent particle states before the next collision occurs. This should therefore be taken only as a rough approximation for a potential with a very weak attractive part. For an attractive well with appreciable strength, the calculations must therefore be handled with care. With the parameters chosen for the Moszkowski and Scott[3] potential in Eq. (2.2) and the fact that $K < 1$, the rhs of Eq. (6.5) has an order of magnitude less than unity times Av_0. Since $A \neq 0$, clearly the rhs of Eq (6.5) is negligibly small only if $v_0 \ll 1$. When translated in terms of conventional units, since $v_0 = 2m^*_{\mathrm{red}}\tilde{v}_0/\hbar^2 k_{\mathrm{F}}^2$ [cf. Eq. (2.2)], we have

$$\tilde{v}_0 \ll \hbar^2 k_{\mathrm{F}}^2/2m^*_{\mathrm{red}}\,,$$

with m^*_{red} the effective nucleon mass in the Fermi sea. This implies that the far-field scattered wave is not appreciably perturbed only when the depth of the attractive potential is much less than the effective kinetic energy of the nucleon pair in the Fermi sea. Observe further that in the absence of the attractive well, the rhs of Eq. (6.5) vanishes and we have the same result obtained in Ref. 1. This is in agreement with the original work of Bethe and Goldstone[2] for the pure hard-core problem.

Although the above analysis was performed for a small core radius, observe that Eq. (6.5) remains finite for $c \to 0$ so that, in effect, the conclusion reached is still true regardless of the size of the radius (but it must be within reasonable bounds of the phenomenological potential used in the calculation). This implies that the conclusion regarding the mod-

ulation of the far-field scattered wave also holds for other very repulsive phenomenological potentials, e.g., the Reid soft-core potential. A "soft-core" potential has, by definition, a very repulsive core that goes to infinity only for $r \to 0$, instead of extending over the region $0 < r < c$. The important aspect in this section is the strength of the attractive part of the interaction potential. Theoretically, the attractive part should not cause any appreciable modulation of the wave function since any modulation would contain mostly Fourier components with wave numbers inside the Fermi sphere; these components are not admitted in the Bethe–Goldstone equation because they are occupied by other nucleons.[4]

VII. NORMALIZATION CONSTANT A

Under the assumption that the attractive part of the interaction potential is rather weak, we can find an approximation of the normalization constant A for a small core radius.

Now, considering Eq. (6.4) with $\cos Ks$ instead of $\sin Ks$, we have

$$\lim_{r\to+\infty}\int_0^r F^*(s)\cos Ks\,ds = Av_0e^{\mu c}\left(\frac{\mu}{\mu^2+K^2}\right) + Av_0e^{\mu c}\lim_{r\to+\infty}\int_c^\infty e^{-\mu r'} \times\left[\int_0^r \chi(s,r')\cos Ks\,ds\right]dr' . \tag{7.1}$$

After some manipulations and a change of variables, and in the limit as $r \to +\infty$, the integral inside the square brackets becomes

$$(1/\pi)(\mathrm{IV}\cdot\cos Kr' + \mathrm{V}\cdot\sin Kr') , \tag{7.2}$$

where

$$\mathrm{IV} \equiv 2\int_0^{r'} \frac{\sin t\cos Kt}{t}\,dt = 4\,\mathrm{si}(r') - \mathrm{si}[r'(1-K)] - \mathrm{si}[r'(1+K)] + \pi \tag{7.3a}$$

and

$$\mathrm{V} \equiv \ln\left|\frac{1-K}{1+K}\right| + 2\int_0^{r'} \frac{\sin t\cos Kt}{t}\,dt = \ln|1-K| + 2\,\mathrm{ci}(r') - \mathrm{ci}[r'(1-K)] - \mathrm{ci}[r'(1+K)] . \tag{7.3b}$$

In doing so, we have used the formula in Ref. 10 [Eq. (2.641)] and the sine and cosine integrals defined by

$$\mathrm{si}(x) = -\int_x^\infty \frac{\sin t}{t}\,dt ,$$

and

$$\mathrm{ci}(x) = -\int_x^\infty \frac{\cos t}{t}\,dt .$$

Consequently, the integral in Eq. (7.1), viz.

$$\mathrm{VI} \equiv \int_c^\infty e^{-\mu r'}(\mathrm{IV}\cdot\cos Kr' + \mathrm{V}\cdot\sin Kr')\,dr' , \tag{7.4}$$

can now be integrated approximately. After repeated applications of the formulas in Ref. 10 [Eqs. (6.261) and (6.262)] and some rather lengthy calculations, we obtain, in the limit as $c \to 0$,

$$\begin{aligned}\mathrm{VI} \to{}& \frac{\mu + K\ln|1-K|}{\mu^2+K^2} - \frac{K}{4\pi(\mu^2+K^2)}\left\{K\ln\left[\frac{\mu^2+(2K+1)^2}{\mu^2+1}\right] + 4\ln\left[\frac{\mu^2+(1+K)^2}{\mu^2+(1-K)^2}\right]\right.\\ &\left. - \ln[(1-K^2+\mu^2)^2+4\mu^2K^2] - \ln\left[\frac{(1+2K+\mu^2)^2+4K^2\mu^2}{(1+K)^4}\right] - \ln\left[\frac{(1-2K+\mu^2)^2+4K^2\mu^2}{(1-K)^4}\right]\right\}\\ &- \frac{\mu}{2\pi(\mu^2+K^2)}\left\{4\tan^{-1}\left[\frac{2\mu}{1-K^2-\mu^2}\right] + \tan^{-1}\left[\frac{2(1+K)\mu}{1+2K-\mu^2}\right] - \tan^{-1}\left[\frac{2(1-K)\mu}{1-2K-\mu^2}\right]\right.\\ &\left. + \tan^{-1}\left[\frac{2\mu K}{1+2K+\mu^2}\right] + \tan^{-1}\left[\frac{2\mu K}{1-2K+\mu^2}\right] - 2\tan^{-1}\left[\frac{2\mu K}{1-K^2+\mu^2}\right]\right\} .\end{aligned} \tag{7.5}$$

Hence, for small c, Eq. (7.1) is

$$\lim_{r\to+\infty}\int_0^r F^*(s)\cos Ks\,ds \sim Av_0e^{\mu c}\left[\frac{\mu}{\mu^2+K^2} + \mathrm{VI}\right] . \tag{7.6}$$

Also, the first integral in Eq. (6.2b), representing the pure hard core, has already been calculated in Ref. 1:

$$\lim_{r\to+\infty}\int_0^r F(s)\cos Ks\,ds = A\lim_{r\to+\infty}\int_0^r \left\{[\delta(s-c) - \chi(s,c)] - \int_0^c \chi(s,s')\chi(s',c)\,ds' - \sum_{n\ \mathrm{odd}} B_nS_{0n}\left(c,\frac{s}{c}\right)\right\} \times \cos Ks\,ds. \tag{7.7}$$

Finally, using Eqs. (7.6) and (7.7), the normalization constant A can now be determined easily from Eq. (6.2b), viz.

$$A^{-1} \sim [\text{pure hard core terms Eq. (7.7)} - v_0e^{\mu c}(\mu/(\mu^2+K^2) + \mathrm{VI})] . \tag{7.8}$$

With this, knowledge of the two-nucleon wave function is now complete. Equations (5.7), (5.8), and (7.8) constitute the complete closed form solution of the problem for the small core radius c. Although we have performed a detailed analysis only for small c, in principle a similar analysis can also be carried out for other values of c; however, this is a somewhat academic exercise of seemingly no known importance since the currently accepted nuclear model has a small core radius of $\tilde{c} \approx 0.4$ fm. Similar important quantities such

as the reaction matrix and the binding energy can also be determined, but the calculations involved are quite difficult and cumbersome. Finally, observe that in the absence of the attractive part of the interaction potential, i.e., $v_0 \equiv 0$, Eq. (7.8) reduces to the same result obtained in Ref. 1.

In conclusion, it might be added that with the success of the present method, perhaps the Bethe–Goldstone equation, using more realistic (and currently popularly used) nucleon–nucleon potentials, e.g., the Hamada–Johnston and Reid soft-core potentials, may now also be amenable to a closed form solution.

APPENDIX A: DERIVATION OF EQ. (4.5c)

Consider Eq. (4.5c),

$$\mathrm{II} \equiv \int_{-1}^{1} I(r,c) f_n(r) dr , \tag{A1}$$

where

$$I(r,c) \equiv \int_c^{\infty} \chi(r,r') e^{-\mu(r'-c)} dr' . \tag{A2}$$

Upon the substitution for $\chi(r,r')$ from Eq. (2.1b) into (A2), using the formula in Ref. 10 [Eq. (3.944)], and after a change of variables, we obtain

$$\begin{aligned} \pi I(r,c) &= e^{\mu(c-r)} \int_{c-r}^{\infty} e^{-\mu s} \frac{\sin s}{s} ds \\ &\quad - e^{\mu(c-r)} \int_{c+r}^{\infty} e^{-\mu s} \frac{\sin s}{s} ds \\ &= (ie^{\mu(c-r)}/2) \{\Gamma[0,(\mu+i)(c-r)] \\ &\quad - \Gamma[0,(\mu-i)(c-r)]\} - (ie^{\mu(c+r)}/2) \\ &\quad \times \{\Gamma[0,(\mu+i)(c+r)] \\ &\quad - \Gamma[0,(\mu-i)(c+r)]\} , \end{aligned} \tag{A3}$$

where $\Gamma(\cdot,\cdot)$ is the incomplete gamma function. Expressing $\Gamma(0,z)$ in terms of the exponential integral

$$E_1(z) = \int_z^{\infty} \frac{e^{-t}}{t} dt, \quad |\arg(z)| < \pi , \tag{A4}$$

viz. $\Gamma(0,z) = E_1(z)$, and using the fact that, for $z \in \mathbb{C}$,

$$\mathrm{Im} f(z) = (i/2)[f^*(z) - f(z)] ,$$

(A3) becomes

$$\begin{aligned} I(r,c) = (e^{\mu c}/\pi) \, \mathrm{Im}\{&e^{-\mu r} E_1[(c-r)(\mu-i)] \\ &- e^{\mu r} E_1[(c+r)(\mu-i)]\}, \quad \mu > 1 . \end{aligned} \tag{A5}$$

Thus from (A1) we have

$$\mathrm{II} \equiv \int_{-1}^{1} I(ct,c) f_n(t) dt = \frac{2}{\pi} \int_0^1 I(ct,c) f_n(t) dt, \quad n \text{ odd} , \tag{A6}$$

where we have used the parity of $f_n(t)$ and the fact that $I(ct,c)$ is odd in t. Now insert (A5) into (A6) and again use the parity conditions to obtain

$$\begin{aligned} \mathrm{II} = \frac{e^{\mu c}}{\pi} \int_{-1}^{1} &e^{-\mu c t} \, \mathrm{Im}\{E_1[(\mu-i)c(1-t)]\} \\ &\times f_n(t) dt, \quad n \text{ odd} . \end{aligned} \tag{A7}$$

In Appendix B, an approximation of the integral in Eq. (A7) for small c will be presented.

APPENDIX B: APPROXIMATION OF EQ. (A7) FOR SMALL c

Unfortunately, the integral in (A7) cannot be integrated analytically in closed form. Although in principle one could use the expansion for the exponential integral,[11]

$$E_1(z) = -\gamma - \ln z - \sum_{n=1}^{\infty} \frac{(-1)^n z^n}{n n!}, \quad |\arg(z)| < \pi , \tag{B1}$$

and obtain an approximation of (A7) for small c, the calculations are too cumbersome to warrant any merit. Also, the contribution of Eq. (5.8d) is quite small. Specifically, using the results of the Theorem one can show that

$$B_n^* \sim O(c^9) ,$$

which implies that as far as the quantity III* in Eq. (5.8d) is concerned, the integral in (A7) is not very significant.

Nevertheless, one could use the crude approximation in Eq. (5.3) to calculate the integral II in (A7). Instead of (A7), consider (A1) with (A2) approximated by Eq. (5.3). This gives

$$\begin{aligned} \mathrm{II} &\equiv \int_{-1}^{1} I(ct,c) f_n(t) dt \\ &\sim -2 (\lambda_n(c)/u_n^2(c))^{1/2} e^{\mu c} \tan^{-1}(1/\mu) \\ &\quad \times \int_{-1}^{1} \sinh(\mu t) P_n(t) dt . \end{aligned} \tag{B2}$$

The integral in Eq. (B2) becomes

$$\begin{aligned} &\int_{-1}^{1} \sinh(\mu t) P_n(t) dt \\ &= \frac{1}{2} [(-1)^n - 1] \int_{-1}^{1} e^{-\mu t} P_n(t) \\ &= \begin{cases} \displaystyle\int_{-1}^{1} e^{-\mu t} P_n(t) dt, & n \text{ odd}, \\ 0, & n \text{ even}. \end{cases} \end{aligned} \tag{B3}$$

Using the integrals (see Refs. 12 and 10 [Eq. (3.387)], respectively)

$$\int_{-1}^{1} f(t) P_n(t) dt = \frac{(-1)^n}{2^n n!} \int_{-1}^{1} (t^2-1)^n f^{(n)}(t) dt \tag{B4}$$

and

$$\begin{aligned} &\int_{-1}^{1} (1-t^2)^n e^{-\mu t} dt \\ &= (\pi)^{1/2} (2/\mu)^{n+1/2} \Gamma(n+1) I_{n+1/2}(\mu), \\ &\quad n \geqslant -1 , \end{aligned} \tag{B5}$$

where $I_{n+1/2}(\mu)$ is the modified Bessel function of the first kind, Eq. (B2), together with (B3) becomes

$$\mathrm{II} \sim (-1)^{n+1} 2(\lambda_n(c)/u_n^2(c))^{1/2} e^{\mu c} \times \tan^{-1}(1/\mu)(2\pi/\mu)^{1/2} I_{n+1/2}(\mu), \quad n \text{ odd only}. \tag{B6}$$

This completes the calculation of Eq. (4.5c). In spite of the crudeness in using the approximation Eq. (5.8d), the above result is a very reasonable estimate.

APPENDIX C: EVALUATION OF EQ. (6.4)

We want to show that

$$\lim_{r \to +\infty} \int_0^r \chi(s,r') \sin Ks \, ds = \sin Kr', \quad 0 \leqslant K < 1. \tag{C1}$$

The integral in (C1), after a suitable change of variables, is

$$\int_0^r \chi(s,r') \sin Ks \, ds = \frac{1}{\pi} \int_0^r \left[\frac{\sin(s-r')}{s-r'} - \frac{\sin(s+r')}{s+r'} \right] \sin Ks \, ds = \frac{1}{\pi} \int_{-r-r'}^{r-r'} \frac{\sin t}{t} \sin K(t+r') dt. \tag{C2}$$

In the limit as $r \to +\infty$, (C2) becomes

$$\frac{2 \sin Kr'}{\pi} \int_0^\infty \frac{\sin t \cos Kt}{t} dt = \frac{2 \sin Kr'}{\pi} \begin{cases} \pi/2, & 0 \leqslant K < 1, \\ \pi/4, & K = 1, \\ 0, & K > 1, \end{cases} \tag{C3}$$

where the integral can be found in Ref. 10. Thus, for relative momentum inside the Fermi sphere, i.e., $K \equiv k/k_F < 1$, the integral is $\sin Kr'$. This proves (C1).

[1] M. K. Ong, J. Math. Phys. **27**, 1154 (1986).
[2] H. A. Bethe and J. Goldstone, Proc. R. Soc. London Ser. A **238**, 551 (1957).
[3] S. A. Moszkowski and B. L. Scott, Ann. Phys. (NY) **11**, 65 (1960).
[4] L. C. Gomes, J. D. Walecka, and V. F. Weisskopf, Ann. Phys. (NY) **3**, 241 (1958).
[5] P. Ring and P. Schuck, *The Nuclear Many-Body Problem* (Springer, New York, 1980).
[6] B. D. Day, Rev. Mod. Phys. **39**, 719 (1967).
[7] A. M. Krall, *Applied Analysis* (Reidel, Holland, 1986).
[8] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971).
[9] T. M. Apostol, *Mathematical Analysis* (Addison-Wesley, Reading, 1974).
[10] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1980).
[11] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegun (Wiley, New York, 1972).
[12] I. S. Sneddon, *Special Functions of Mathematical Physics and Chemistry* (Interscience, New York, 1961).

Variational nature of dispersion equations revisited

J. L. Schwarzmeier[a] and H. R. Lewis
Los Alamos National Laboratory, MS-F642, Los Alamos, New Mexico 87545

(Received 21 April 1987; accepted for publication 30 December 1987)

Corrections and clarifications are made of some past treatments of the variational nature of the eigenfrequency calculation for dispersion equations and new results are presented. The main conclusions are the following: (1) Any relation between a normal mode and its dual must be consistent with the fact that the boundary conditions satisfied by the normal mode may differ from the adjoint boundary conditions satisfied by the dual. This will affect whether or not a given bilinear form will yield a variational result for the eigenfrequency. (2) If a dispersion matrix is constructed from the dispersion operator by using left and right basis functions that satisfy homogeneous boundary conditions on the dual eigenfunction and the eigenfunction, respectively, then generally a second-order accurate eigenfrequency is obtained by solving the matrix form of the dispersion equation. (3) When solving for the normal modes in terms of perturbation potentials, the adjoint boundary conditions are gauge dependent. For cases where the adjoint boundary conditions allow only the trivial solution for the dual eigenfunction, it may be possible to obtain variational results for the eigenfrequency by requiring that the trial functions for the normal mode and its dual satisfy *variational* boundary conditions.

I. INTRODUCTION

The variational nature of the eigenvalue problem for the linearized Vlasov–Maxwell (VM) equations and related systems has been studied by many authors (for example, see Refs. 1–8). The equation for a normal mode can be expressed as the requirement that the dispersion operator acting on the eigenfunction equal zero. Similarly, the *dual* eigenfunction is an eigenfunction of the *adjoint* dispersion operator corresponding to zero eigenvalue. By the variational nature of the eigenvalue problem we mean that, if an eigenfunction and its dual are known to first order, then the eigenfrequency generally will be accurate to second order. Such a property obviously is desirable. The variational result in its most general form appears to have been proved first in Ref. 4. The main purposes of the current paper are to clarify mathematical aspects of this problem that do not seem to have been appreciated by some plasma physicists who do variational calculations, and to prove a theorem extending the validity of the variational result to the solution of a finite-dimensional matrix problem which approximates the dispersion equation. Much of our discussion will be valid for the variational nature of dispersion equations in general, not just those that arise in plasma physics. However, some of our results may not apply when dealing with singular eigenfunctions, such as continuum modes in magnetohydrodynamics (MHD).

A dispersion functional, which is a bilinear form, can be defined as the inner product of a function that we call the "left function" with the dispersion operator acting on a function that we call the "right function." The right function must lie in the same Hilbert space as the eigenfunction, which is specified by requiring that the eigenfunction satisfy certain boundary conditions and have certain differentiability and integrability properties. A variational result for an eigenfrequency calculated as a root of the dispersion functional will hold if the left function lies in the Hilbert space of the dual eigenfunction. That is, if the dispersion functional is calculated with a right function that is a first-order accurate approximation to the eigenfunction and with a left function that is a first-order accurate approximation to the dual eigenfunction, then a second-order accurate eigenfrequency will be a root of the dispersion functional. This is the result of Theorem I of Ref. 4. The variational result generally will not hold if the eigenfrequency is calculated as a root of a dispersion functional in which the left function is in the space of the eigenfunction itself. This is because the dispersion operator generally is not Hermitian. Even for real omega the dispersion operator generally is not Hermitian unless there are no particle resonances. To use a dispersion functional in which the left function is in the space of the dual eigenfucntion requires finding approximations to the eigenfunction *and* its dual. For a certain class of equilibria, Berk *et al.*[6,7] were able to circumvent this difficulty by defining a nonstandard inner product and adjoint operator in such a way that the dual eigenfunction is simply proportional to the eigenfunction. In a separate publication,[9] it is shown that the same simple relation between the normal mode and its dual can be derived using standard definitions of the inner product and adjoint operator, even though the dispersion operator is not Hermitian.

For proving stability theorems it can be useful to define a dispersion functional for which both the left and right functions are the eigenfunction itself (see, for example, Refs. 10 and 11). Furthermore, although such a dispersion functional may not give variational results for eigenfrequencies, nevertheless approximate eigenfrequencies can be computed by using the dispersion functional.

An element of a dispersion *matrix* is the inner product of a "left basis function" with the dispersion operator acting on a "right basis function." We assume that the dispersion problem has been formulated such that the eigenfunction

[a] Current address: Cray Research Inc., 1090 Industrial Boulevard, Chippewa Falls, Wisconsin 54729.

 0022-2488/88/061486-12$02.50

satisfies *homogeneous* boundary conditions, and we assume that each right basis function satisfies those boundary conditions. For problems in which the adjoint boundary conditions are the *same* as the boundary conditions on the eigenfunction, we can choose the left and right bases to be the same. In this case Theorem II of Ref. 4 shows that eigenfrequencies calculated from this dispersion matrix generally will be variational. The significance of this theorem derives from the fact that it is not necessary to have an approximation to the eigenfunction or its dual in order to construct a dispersion matrix. We merely need that the expansion functions used in constructing the dispersion matrix be a suitable basis for constructing first-order accurate expansions of both the eigenfunction and its dual. Later in this paper we prove a theorem stating that the variational result holds for a matrix calculation even if the adjoint boundary conditions force the left and right bases to be different.

The main conclusions of this work are the following:

(1) Apparent relations between a normal mode and its dual, such as those suggested in Refs. 4–8, may not be valid because the normal mode and its dual may satisfy *different* boundary conditions. This is because integrodifferential operators, like plasma dispersion operators, generally are unbounded. The adjoint operator will always exist, but the adjoint boundary conditions may restrict the domain of the adjoint operator to be *different* than the domain of the original operator. If an approximation to the dual eigenfunction does not satisfy appropriate adjoint boundary conditions, then the eigenfrequency calculation will not be variational.

(2) When solving a matrix problem that approximates the dispersion equation, a variational result for the eigenfrequency will hold even when the adjoint boundary conditions are different than the boundary conditions on the normal mode as long as the left basis functions used in constructing the dispersion matrix form a suitable basis for the dual.

(3) When solving the dispersion equation in terms of perturbation potentials, the adjoint boundary conditions are gauge dependent. We illustrate this with the Vlasov–Maxwell (VM) system. We show that the VM system is a formally self-adjoint function of ω, but that the adjoint boundary conditions depend on the gauge choice. If the adjoint boundary conditions allow only the trivial solution for the dual eigenfunction, it may still be possible to obtain a variational result if the left and right functions are made to satisfy *variational* adjoint boundary conditions and *variational* boundary conditions, respectively.

In another publication[9] the following is shown:

(1) The kinetic part of the general multidimensional dispersion operator[12] is a "Hermitian function of ω,"[8] where ω is the complex Laplace transform frequency. The result is obtained without using Liouville eigenfunctions,[8] which are not well-defined mathematically for multidimensional equilibria.[12] Loosely speaking a Hermitian function of ω means that the dispersion operator would be Hermitian in the usual sense if ω were treated as though it were real when complex conjugating. If the dispersion operator is a "self-adjoint function of ω," then relation (19) between the eigenfunction and its dual is valid.

(2) If the equilibrium admits "conjugate orbits "[6,7] and if the "conjugate-orbit parity condition" [9] is satisfied, then the kinetic part of the dispersion matrix will be symmetric.

(3) If the dispersion matrix is symmetric and if the adjoint boundary conditions are the same as the boundary conditions on the normal mode, then the dual eigenfunction is simply proportional to the complex conjugate of the eigenfunction.

(4) The kinetic part of the dispersion functional for multidimensional VM equilibria can be analytically continued into the lower half of the ω plane. The analytic continuation of the dispersion operator onto the real ω axis shows that relation (19) is valid even for real ω.

The outline for this paper is as follows. Mathematical aspects of adjoint operators are considered in Sec. II. These ideas are illustrated with a simple mathematical example. Section III reviews the theorems of Ref. 4 with emphasis on the generality of the results. An additional theorem is presented that discusses the variational nature of the dispersion matrix problem for the case where the left and right bases are different from one another. This theorem covers the case not treated by Theorem II of Ref. 4. In Sec. IV we discuss formal self-adjointness of the VM system and the gauge dependence of the adjoint boundary conditions. We summarize in Sec. V by presenting a flow chart showing how variational results can arise for dispersion operators with a multitude of different properties. Theorem III is provied in the Appendix.

II. MATHEMATICAL CONSIDERATIONS

A. Adjoints

Let us review the relevant aspects of operator theory that pertain to adjoints. Reference 13 is suggested as a source of information on linear operator theory.

The essential mathematical features of adjoint operators that we wish to discuss here can be revealed by considering the case of scalar equations in one spatial variable. Let x be the independent spatial variable, chosen to be in the range $[0,1]$, and ω be the Laplace transform variable for functions that are Laplace transformed in time. We assume that we are dealing with linear operators defined on some Hilbert space $\mathscr{H}$, and the domain of an operator A will be denoted $\mathscr{D}(A)$. The domain $\mathscr{D}(A)$ is either the entire Hilbert space or a subset thereof. It is the space of functions on which A operates; in a boundary value problem, $\mathscr{D}(A)$ is determined at least in part by the boundary conditions. For example, we might choose the Hilbert space such that any element u in $\mathscr{H}$ satisfies

$$u(x,\omega) \text{ absolutely continuous; } u, u', \text{ and } u'' \text{ in } \mathscr{L}_2^{(c)}(0,1) . \tag{1}$$

Here $\mathscr{L}_2^{(c)}(0,1)$ denotes the space of complex functions that are square integrable on the domain $[0,1]$. Restrictions that determine $\mathscr{D}(A)$ would be in the nature of boundary conditions, which vary from problem to problem.

In this paper we are particularly concerned with quantities that depend on a complex parameter. For example, in the case of the VM equations, the dispersion operator $D(x,\omega)$ depends on the Laplace-transform variable ω as well as on the spatial coordinate x. It will be necessary to distinguish between two kinds of complex conjugates and between

two kinds of adjoints along the lines introduced in Appendix E of Ref. 8. Suppose tht a quantity $A(\omega)$ depends on a complex parameter ω. Here $A(\omega)$ may be a scalar, vector, operator, or matrix. We wish to allow for functions $A(\omega)$ that have a branch cut along the real ω axis and appropriate definitions in the upper and lower half ω planes. Let $A^{(+)}(\omega)$ and $A^{(-)}(\omega)$ denote the branches of $A(\omega)$ in the upper and lower half ω planes, respectively. Then by the symbol $A(\omega)$ we mean

$$A(\omega)=\begin{cases}A^{(+)}(\omega), & \operatorname{Im}\omega>0,\\ A^{(-)}(\omega), & \operatorname{Im}\omega<0.\end{cases} \tag{2}$$

Note that $A(\omega)$ does *not* denote the function obtained by analytically continuing $A^{(+)}(\omega)$ into the lower half of the ω plane. In this paper we do not make explicit use of functions having branch cuts along the real ω axis, but in another publication[9] we have. We define $A^*(\omega)$, the "conjugate function" of ω associated with $A(\omega)$, by

$$A^*(\omega)=[A(\omega^*)]^*, \tag{3}$$

where the quantity on the right is the complex conjugate of $A(\omega^*)$. Similarly, we define $A^\dagger(\omega)$, the "adjoint function" of ω associated with $A(\omega)$, by

$$A^\dagger(\omega)=[A(\omega^*)]^\dagger, \tag{4}$$

where the quantity on the right is the adjoint of $A(\omega^*)$. If $A(\omega)$ is an analytic function of ω in some domain (for example, $\operatorname{Im}\omega>0$) and is analytically continued outside that domain, then the resulting $A^*(\omega)$ and $A^\dagger(\omega)$ are analytic functions of ω everywhere. Because of the necessity of introducing the notion of conjugate and adjoint functions of a parameter, there is a possibility of confusion when indicating the operations of complex conjugating or taking the adjoint. In order to avoid that confusion, we shall generally indicate complex conjugates with brackets and an asterisk and indicate adjoints with brackets and a dagger, as on the right-hand sides of (3) and (4). For example, to indicate the complex conjugate of $A(\omega)$, we shall write $[A(\omega)]^*$ or $[A]^*$; and for the adjoint of $A(\omega)$, we shall write $[A(\omega)]^\dagger$ or $[A]^\dagger$. It is important to remember that, in general, the conjugate function $A^*(\omega)$ is not equal to the complex conjugate $[A(\omega)]^*$ and that the adjoint function $A^\dagger(\omega)$ is not equal to the adjoint $[A(\omega)]^\dagger$.

Mathematicians distinguish between three related concepts. An operator A is called *formally self-adjoint* [14] if the operators A and $[A]^\dagger$ are identical, $[A]^\dagger=A$. An operator A is called *Hermitian*[15] if for all v and u in $\mathscr{D}(A)$, we have $(v,Au)=(Av,u)$. An operator A is called *self-adjoint* [15] if it is Hermitian and if $\mathscr{D}(A)=\mathscr{D}([A]^\dagger)$. A dispersion operator $D(x,\omega)$ may be formally self-adjoint, Hermitian, or self-adjoint *except* for its dependence on the complex parameter ω. That is, it might have one of these properties if ω were treated as a real parameter. In these cases we would refer to $D(x,\omega)$ as a formally self-adjoint, Hermitian, or self-adjoint *function of* ω.[8] It is useful to make this distinction, because certain properties of the solution of the dispersion equation follow if the dispersion operator is a formally self-adjoint function of ω.

An operator A is *bounded* if for all u in its domain there exists a real positive constant c such that

$$\|Au\|\leqslant c\|u\|,$$

where $\|u\|$ is an appropriately defined norm of an element u. If we are considering a *bounded* linear operator A, its domain and the domain of its adjoint can be chosen to be the entire Hilbert space $\mathscr{H}$. This is because when A is bounded we are guaranteed that for all v and u in $\mathscr{H}$ there exists a bounded linear operator $[A]^\dagger$, called the *adjoint* of A, such that

$$(v,Au)=([A]^\dagger v,u), \tag{5}$$

where the inner product of v and u is defined by

$$(v,u)=\int dx[v]^*u. \tag{6}$$

The definition of an integral operator A acting on u in $\mathscr{D}(A)$ is

$$Au=\int dx'\,a(x,x')u(x'). \tag{7}$$

The operator A in (7) will be bounded if the kernel $a(x,x')$ satisfies

$$\int dx\int dx'|a^2(x,x')|<\infty, \tag{8}$$

although condition (8) is not necessary for boundedness. If A is a bounded integral operator, then the kernel of its adjoint, $\tilde{a}(x,x')$, is defined by

$$\tilde{a}(x,x')=[a(x',x)]^*. \tag{9}$$

If the kernel of the adjoint equals the kernel of A, then A is self-adjoint.

If A is an *unbounded* operator, like the derivative operator, the situation is different. Unbounded linear operators can be defined only with domains that are subsets of the Hilbert space. The subset of elements v in $\mathscr{H}$ for which (5) is valid for any u in $\mathscr{D}(A)$ is called the domain of the adjoint, $\mathscr{D}([A]^\dagger)$. In general, $\mathscr{D}(A)\neq\mathscr{D}([A]^\dagger)$. For example, if A is a differential operator, in the process of converting the left-hand side of (5) into the right-hand side of (5), integrations by parts are performed that lead to boundary terms. The form (5) will result only if the boundary terms are made to vanish. The conditions on v that make the boundary terms vanish, for given boundary conditions on u, are the so-called adjoint boundary conditions. Thus, for given boundary conditions on the elements u in $\mathscr{D}(A)$, the elements v in $\mathscr{D}([A]^\dagger)$ are the subset of elements in $\mathscr{H}$ that satisfy the adjoint boundary conditions. For unbounded operators, like differential and integrodifferential operators, the adjoint operator exists only if we restrict the domain of the adjoint operator to satisfy the adjoint boundary conditions.

B. An example

The example in this subsection illustrates several points. First, it shows with a realistic problem how the adjoint boundary conditions can be different than the original boundary conditions even though the operator is formally self-adjoint. Second, the example shows that even though an operator may be formally self-adjoint (or a formally self-adjoint function of ω), when boundary conditions are taken into account the operator may *not* be Hermitian or self-adjoint (or a Hermitian or self-adjoint function of ω). Thus,

when determining Hermiticity or self-adjointness of an operator one must consider the boundary conditions as well. Third, the example shows how apparently general relations between an eigenfunction and its dual may not be valid after boundary conditions are applied. This illustrates why it is important to determine if an eigenfunction and its dual lie in the same space.

Consider the simple boundary value problem

$$D(x,\omega)\phi(x,\omega)=0, \quad B(\phi)\equiv\begin{pmatrix}\phi'(0,\omega)\\ \phi(1,\omega)\end{pmatrix}=0\,, \tag{10}$$

where the operator D is

$$D(x,\omega)=\frac{d^2}{dx^2}+2i\nu\frac{d}{dx}+K^2(\omega)\,, \tag{11}$$

where ν is a real constant and ω is a complex number to be determined from the dispersion relation. In (10) we have symbolically written the set of boundary conditions on ϕ as $B(\phi)=0$. If the function $K(\omega)$ is real except for its dependence on ω (Lewis and Symon term this a "real function of ω" [8]), then

$$[K(\omega)]^*=K(\omega^*)\,. \tag{12}$$

The adjoint operator $[D(x,\omega)]^\dagger$ is that operator for which

$$([D]^\dagger v,u)=(v,Du) \tag{13}$$

for all v in $\mathscr{D}([D]^\dagger)$ and u in $\mathscr{D}(D)$. If we substitute (11) for D in (13) and do some integrations by parts we find

$$(v,Du)=([D]^\dagger v,u)+J(v,u)\,, \tag{14}$$

where

$$[D(x,\omega)]^\dagger=\frac{d^2}{dx^2}+2i\nu\frac{d}{dx}+[K^2(\omega)]^* \tag{15}$$

and

$$\begin{aligned}J(v,u)=&[v(1,\omega)]^*u'(1,\omega)-[v(0,\omega)]^*u'(0,\omega)\\&-[v'(1,\omega)]^*u(1,\omega)+[v'(0,\omega)]^*u(0,\omega)\\&+2i\nu\{[v(1,\omega)]^*u(1,\omega)\\&-[v(0,\omega)]^*u(0,\omega)]\}=0\,.\end{aligned} \tag{16}$$

The adjoint dispersion equation is

$$[D(x,\omega)]^\dagger\tilde{\phi}(x,\omega)=0, \quad \tilde{B}(\tilde{\phi})+0\,, \tag{17}$$

where the adjoint boundary conditions $\tilde{B}(\tilde{\phi})=0$ have not yet been specified.

Comparing (11) and (15) we see that if $K(\omega)$ were a real constant, independent of ω, then the operator (11) would be formally self-adjoint. If we were to allow $K(\omega)$ to be a real function of ω, as in (12), then $D(x,\omega)$ would be a formally self-adjoint function of ω. That is, (11), (12), and (15) would imply that the adjoint operator at the point ω were the operator itself with ω replaced by ω^*:

$$[D(x,\omega)]^\dagger=D^\dagger(x,w^*)=D(x,\omega^*)\,. \tag{18}$$

Equations (17) and (18) imply

$$D(x,\omega^*)\tilde{\phi}(x,\omega)=0\,.$$

If we substitute ω^* for ω in the last equation we obtain

$$D(x,w)\tilde{\phi}(x,\omega^*)=0\,.$$

Comparing this equation to (10) we see that the dual eigenfunction would be related to the eigenfunction by

$$\tilde{\phi}(x,\omega)=\phi(x,\omega^*) \tag{19}$$

if the adjoint boundary conditions satisfied by the dual eigenfunction were the same as the boundary conditions satisfied by the eigenfunction. This is a result quoted in Refs. 4 and 8, and a related but even sharper result is derived in Refs. 6 and 7. However, as we shall see, both of these results are invalid if the adjoint boundary conditions are different than the boundary conditions on the eigenfunction.

The general solutions of (10) and (17) for the normal mode and its dual can be found easily:

$$\begin{aligned}\phi(x,\omega)=&a_1\exp(i\{-\nu+[\nu^2+K^2(\omega)]^{1/2}\}x)\\&+a_2\exp(-i\{\nu+[\nu^2+K^2(\omega)]^{1/2}\}x)\end{aligned} \tag{20}$$

and

$$\begin{aligned}\tilde{\phi}(x,\omega)=&\tilde{a}_1\exp[i(-\nu+\{\nu^2+[K^2(\omega)]^*\}^{1/2})x]\\&+\tilde{a}_2\exp[-i(\nu+\{\nu^2+[K^2(\omega)]^*\}^{1/2})x]\,.\end{aligned} \tag{21}$$

When the property (12) is used we see that the general solutions of the equations for $\tilde{\phi}(x,\omega)$ and $\phi(x,\omega)$ can indeed be related according to (19) by taking $\tilde{a}_1=a_1$ and $\tilde{a}_2=a_2$. When the boundary conditions in (10) are applied to the general solution (20) we obtain

$$\begin{aligned}\phi(x,\omega)=a_1\Big[&\exp(i\{-\nu+[\nu^2+K^2(\omega)]^{1/2}\}x)\\&+\left(\frac{-\nu+[\nu^2+K^2(\omega)]^{1/2}}{\nu+[\nu^2+K^2(\omega)]^{1/2}}\right)\\&\times\exp(-i\{\nu+[\nu^2+K^2(\omega)]^{1/2}\}x)\Big]\,,\end{aligned} \tag{22}$$

with the accompanying dispersion relation

$$\begin{aligned}d(\omega)=&\{\nu+[\nu^2+K^2(\omega)]^{1/2}\}\exp\{i[\nu^2+K^2(\omega)]^{1/2}\}\\&+\{-\nu+[\nu^2+K^2(\omega)]^{1/2}\}\\&\times\exp\{-i[\nu^2+K^2(\omega)]^{1/2}\}=0\,.\end{aligned} \tag{23}$$

To determine the adjoint boundary conditions we return to (14) and (16) and require that $J(v,u)$ vanish. Because of the boundary conditions on ϕ, we see that the adjoint boundary conditions are

$$[\tilde{\phi}(1,\omega)]^*=[\tilde{\phi}'(0,\omega)]^*-2i\nu[\tilde{\phi}(0,\omega)]^*=0,$$

or

$$\tilde{B}(\tilde{\phi})\equiv\begin{pmatrix}\tilde{\phi}(1,\omega)\\ \tilde{\phi}'(0,\omega)+2i\nu\tilde{\phi}(0,\omega)\end{pmatrix}=0\,. \tag{24}$$

Notice that if $\nu\neq0$ the adjoint boundary conditions are *different* than those in (10), while if $\nu=0$ they are the *same*. Also it is evident from (14) that if the boundary conditions on ϕ were that $\phi(0,\omega)=\phi(1,\omega)=0$, then the $\tilde{\phi}$ would satisfy the same boundary conditions, independent of ν. Thus, whether or not the adjoint boundary conditions are different from the boundary conditions on the normal mode depends in detail on the latter boundary conditions and the differential operator. For $\nu\neq0$, since $\mathscr{D}([D]^\dagger)\neq\mathscr{D}(D)$, the problem is not self-adjoint. Moreover, the problem is not even Hermitian, since (14) is not satisfied for all u and v in $\mathscr{D}(D)$.

For $\nu \neq 0$, the solution (21) that satisfies the adjoint boundary conditions is

$$\tilde{\phi}(x,\omega) = \tilde{a}_1 \Big[\exp(i\{-\nu + [\nu^2 + K^2(\omega^*)]^{1/2}\}x) + \left(\frac{\nu + [\nu^2 + K^2(\omega^*)]^{1/2}}{-\nu + [\nu^2 + K^2(\omega^*)]^{1/2}} \right) \times \exp(-i\{\nu + [\nu^2 + K^2(\omega^*)]^{1/2}\}x) \Big], \quad (25)$$

where we have used (12). The adjoint dispersion relation is

$$\tilde{d}(\omega) = (-\nu + \{\nu^2 + [K^2(\omega)]^*\}^{1/2}) \times \exp(i\{\nu^2 + [K^2(\omega)]^*\}^{1/2}) + (\nu + \{\nu^2 + [K^2(\omega)]^*\}^{1/2}) \times \exp(-i\{\nu^2 + [K^2(\omega)]^*\}^{1/2}) = 0. \quad (26)$$

Notice that (26) is the complex conjugate of (22). Comparing (25) and (22) we see that relation (19) is no longer satisfied! The reason for this is that relation (19) is valid only when the solution of the dispersion equation and the solution of the adjoint equation are in the same function space. When boundary conditions are imposed, relation (21) is not necessarily true. It depends on whether or not the adjoint boundary conditions are different from the boundary conditions on the normal mode itself.

The discussion of this section relates to earlier works on the variational nature of dispersion equations in the following way. Some dispersion operators, like the integrodifferential dispersion operator of the linearized VM system, are unbounded. This is because derivative operators are generally unbounded. Berk *et al.*[6,7] express the differential terms as an "integral operator" by multiplying the derivatives by delta functions and integrating over space. However, the kernel of such an "integral operator" will not satisfy condition (8), so boundedness of such an integral representation of the differential terms is not assured. (In fact, we know that we are really dealing with differential operator terms, and therefore adjoint boundary conditions must be considered.) Furthermore, self-adjointness of an integral operator can be established by examining whether or not (9) is true only if the integral operator is in fact bounded.[16] Therefore, by not mentioning boundary conditions and merely examining whether (9) is satisfied, Berk *et al.*[6,7] were not justified in claiming self-adjointness of the dispersion operator. [Merely examining (9) is equivalent to testing for formal self-adjointness of the dispersion operator, not self-adjointness of the dispersion operator.] The impact of the adjoint boundary conditions on the dual eigenfunction has often not been considered. We have seen that, unless the boundary conditions are such that $\mathscr{D}(A) = \mathscr{D}([A]^\dagger)$, apparently general relations between an eigenfunction and its dual, such as relation (19), may *not* be valid. Use of relation (19) to construct a dispersion functional may not lead to a variational calculation of eigenfrequencies.

III. VARIATIONAL NATURE OF DISPERSION EQUATIONS

In this section we state and discuss three theorems that deal with the variational nature of eigenvalue calculations that involve a complex parameter ω. The first two theorems are Theorems I and II of Ref. 4. Suppose we are solving an eigenvalue problem in which the operator, eigenfunctions, and eigenvalues all depend on a complex parameter ω. In Ref. 4 it is shown that the eigenvalues of an operator are equal to the value of a dispersion functional constructed as the inner product of the normalized dual eigenfunction with the operator acting on the normalized eigenfunction. The specific question is, how accurately can the ω be calculated for which the eigenvalue (which depends on ω) equals a given complex number? This question can be asked about any operator, and no assumption must be made about the operator's being self-adjoint according to any definition. This is a more general problem than the problem of finding normal modes in plasma physics. The connection that the work of Ref. 4 has with solving the dispersion equation in plasma physics is that, in the latter problem, we seek to vary the complex parameter ω until the dispersion operator has a zero eigenvalue (normal mode). (Thus we refer to ω only as an eigenfrequency, not an eigenvalue.)

Let the operator be $D(x,\omega)$, where x stands for the set of independent spatial variables. Then we have the following theorem from Ref. 4 as applied to a dispersion operator.

Theorem I: Let ψ and $\tilde{\psi}$ approximate to order ϵ an eigenfunction and its dual, respectively, of the operator $D(x,\omega)$ corresponding to the eigenfrequency ω_0. Let ω' be an approximate eigenfrequency obtained as a root of a dispersion functional constructed with ψ and $\tilde{\psi}$ as trial functions. If $D(x,\omega)$ is an analytic function of ω in the neighborhood of ω_0, then ω' will approximate ω_0 to order ϵ^2, except in special circumstances. (These special circumstances are exhibited in Ref. 4.)

Theorem I is useful when the eigenfrequency is calculated as a root of a dispersion functional that is constructed with known forms for the eigenfunction and its dual. However, the variational result can also hold when one solves a finite-dimensional matrix problem that approximates the dispersion equation. In this case we do not need *a priori* approximations for the eigenfunction or the dual. We only need to choose appropriately the right and left basis functions that are used for constructing the dispersion matrix. The right basis functions must be such that *some* linear combination of them would be a first-order approximation to the eigenfunction. The left basis functions must be such that *some* linear combination of them would be a first-order approximation to the dual. Theorem II of Ref. 4 deals with the matrix case in which the adjoint boundary conditions are the same as the boundary conditions on the eigenfunction, so that the eigenfunction and its dual are in the same function space. Theorem III, which we shall prove in Appendix B, deals with the matrix case in which the adjoint boundary conditions are different than the boundary conditions on the eigenfunction. In the remainder of this section we suppress the dependence of operators and functions on x.

Theorem II: Let ω_0 yield a zero eigenvalue of the operator $D(\omega)$, and let ϕ be the corresponding eigenvector normalized to unity:

$$D(\omega_0)\phi = 0, \quad \|\phi\| = 1. \quad (27)$$

Let S be a subspace within which ϕ can be approximated to order ϵ:

$$\phi = P\phi + \epsilon\xi, \tag{28}$$

where P is a projection operator from the Hilbert space $\mathscr{H}$ onto S, and ξ is a vector normalized to unity. Assume *likewise* that the dual $\tilde{\phi}$ can be approximated to order ϵ in S. Let $\underline{D}(\omega)$ be the operator $D(\omega)$ restricted to the subspace S:

$$\underline{D}(\omega) = PD(\omega)P. \tag{29}$$

Let ω' yield a zero eigenvalue of $\underline{D}(\omega)$:

$$\underline{D}(\omega')\underline{\phi} = 0, \tag{30}$$

where $\underline{\phi}$ is a vector in S normalized to unity,

$$P\underline{\phi} = \underline{\phi}. \tag{31}$$

Then the eigenvector $\underline{\phi}$ approximates ϕ to order ϵ and, if ω_0 is a simple root of the dispersion relation associated with (27), ω' approximates ω_0 to order ϵ^2.

The proof of this theorem is given in Ref. 4, but the proof is not valid if the adjoint boundary conditions on the dual are different than the boundary conditions on the eigenfunction. For example, suppose that a boundary condition were that the slope of the eigenfunction vanish at $x = 0$, while an adjoint boundary condition required that a linear combination of the dual and its derivative vanish at $x = 0$ (as in the example of Sec. II B). We would expand the eigenfunction in a basis whose elements had zero slope at $x = 0$. However, taking a linear combination of these basis functions could only satisfy the boundary conditions of the dual in the special case that the dual also *happened* to vanish at the origin. Hence the dual could not generally be approximated to first order in that basis. We can extend Theorem II by showing that the variational result again holds if we assume that the left basis functions used to form the dispersion matrix are a basis with which the dual can be approximated to first order. We state the theorem here and prove it in Appendix B. (In our treatment of Theorem III we assume explicitly that the approximate eigenfrequency ω' is in the neighborhood of the exact eigenfrequency ω_0 [see Eq. (38)]. This assumption also was made for Theorems I and II of Ref. 4 [see Eqs. (15), (38), and (44) of Ref. 4], although in Ref. 4 we did not include the assumption as part of the statement of the theorems. The statement and proof given here of Theorem III are preferable to the treatment of Theorem II in Ref. 4.)

Theorem III: Let ω_0 be a simple zero of one, and only one, eigenvalue of the operator $D(\omega)$, and let ϕ be the corresponding eigenvector normalized to unity:

$$D(\omega_0)\phi = 0, \quad \|\phi\| = 1. \tag{32}$$

Let S be a subspace within which ϕ can be approximated to order ϵ:

$$\phi = P\phi + \epsilon\xi, \tag{33}$$

where P is a projection operator from the Hilbert space $\mathscr{H}$ onto S, and ξ is a vector normalized to unity. Assume that the dual $\tilde{\phi}$ can be approximated to order ϵ in a subspace $\tilde{S}$:

$$\tilde{\phi} = \tilde{P}\tilde{\phi} + \epsilon\tilde{\xi}, \tag{34}$$

where $\tilde{P}$ is the projection operator from $\mathscr{D}([D]^\dagger)$ onto $\tilde{S}$, and $\tilde{\phi}$ and $\tilde{\xi}$ are vectors normalized to unity. Both ξ and $\tilde{\xi}$ are assumed to have bounded derivatives as required. Let $\underline{D}(\omega)$ be the operator $D(\omega)$ restricted to the subspaces S *and* $\tilde{S}$:

$$\underline{D}(\omega) = \tilde{P}D(\omega)P. \tag{35}$$

Let ω' yield a zero eigenvalue of $\underline{D}(\omega)$:

$$\underline{D}(\omega')\underline{\phi} = 0, \tag{36}$$

where $\underline{\phi}$ is a vector in S normalized to unity,

$$P\underline{\phi} = \underline{\phi}. \tag{37}$$

Assume that ω' is near ω_0 in the sense that

$$|\omega' - \omega_0| < \epsilon^\alpha, \quad \text{for some } \alpha \text{ satisfying } 0 < \alpha \leqslant 1. \tag{38}$$

Then the eigenvector $\underline{\phi}$ approximates ϕ to order ϵ^β, where $\alpha \leqslant \beta \leqslant 1$. Also, since ω_0 is a simple root of the dispersion relation associated with (32), the dispersion relation associated with (36) will have a root ω' that approximates ω_0 to order $\epsilon^{2\beta}$. This theorem is proved in Appendix A.

IV. VLASOV–MAXWELL SYSTEM

In this section we discuss the formal self-adjointness of the VM system, and the gauge dependence of the adjoint boundary conditions. This will all be done in the context of *conducting* boundary conditions for the normal modes, and for cases where the plasma density vanishes at the conducting wall. Under these circumstances we will conclude the following: (a) the VM dispersion operator is a formally self-adjoint function of ω; (b) for the gauge choice $\phi_1 \equiv 0$, the VM dispersion operator is a self-adjoint function of ω; and (c) for the gauge choice $A_{1r} \equiv 0$, the VM dispersion operator is not even a Hermitian function of ω. For this case, the adjoint boundary conditions are *different* than the boundary conditions on the eigenfunction, and the adjoint boundary conditions allow only the *trivial* solution of the adjoint dispersion equation. However, we are able to construct *variational* boundary conditions to be satisfied by the eigenfunction and its dual such that a variational result still follows. The variational boundary conditions are consistent with the physical conducting boundary conditions.

Our emphasis has been on boundary conditions that are imposed at physical interfaces in a plasma system. There also may be regularity conditions imposed in the interior of the plasma. Regularity conditions are imposed to eliminate singular solutions that might otherwise arise due to a change in the coordinate system used to solve the problem. Just as there are regularity conditions on normal modes, there are adjoint regularity conditions on duals. (The example in Sec. II B contains boundary conditions at the origin, which could be interpreted as regularity conditions.) For dispersion operators that include a kinetic contribution, the determination of regularity conditions inside the plasma *may* require knowing how the kinetic species contributes asymptotically to the dispersion operator in the vicinity of the singular point. For the VM system this would involve solving the Vlasov equation asymptotically near the singular point. By computing appropriate moments of the perturbation distribution function, we could determine how the charge density and current density in Maxwell's equations affect the dispersion equation at the singular point.

The linearized VM system can be written in the form

$$\mathbf{\Lambda}\cdot\mathbf{\Phi}=\sum_s\int d\mathbf{\rho}\,\mathscr{J}_s f_{s1}\,, \tag{39}$$

where the array of perturbation potentials is

$$\mathbf{\Phi}=\begin{pmatrix}\phi_1\\ \mathbf{A}_1\end{pmatrix}, \tag{40}$$

and the charge-current kernel is

$$\mathscr{J}_s=q_s\begin{pmatrix}1\\ -\mathbf{v}/c\end{pmatrix}. \tag{41}$$

Here m_s and q_s are the mass and charge of particles of species s, respectively, and f_{s1} denotes the corresponding perturbation distribution function. The canonical momentum vector $\mathbf{\rho}$ is defined by[8]

$$\mathbf{\rho}=m_s\mathbf{v}+q_s\mathbf{A}_0/c\,, \tag{42}$$

where $\mathbf{A}_0$ is the equilibrium vector potential. The field operator of the VM system is defined by

$$\mathbf{\Lambda}(x,\omega)\cdot\mathbf{\Phi}=\frac{1}{4\pi}\begin{pmatrix}-\nabla^2\phi+\dfrac{i\omega}{c}\nabla\cdot\mathbf{A}\\ \dfrac{i\omega}{c}\nabla\phi-\nabla\times(\nabla\times\mathbf{A})+\dfrac{\omega^2}{c^2}\mathbf{A}-\sum\limits_s\dfrac{{\omega_{ps}}^2}{c^2}\mathbf{A}\end{pmatrix}. \tag{43}$$

The plasma frequency for species s is defined by

$${\omega_{ps}}^2=\frac{4\pi {q_s}^2}{m_s}\int d\mathbf{\rho}\,f_{s0}\,,$$

where f_{s0} is the equilibrium distribution function for species s.

By choosing to write the VM equations with the kernel $\mathscr{J}_s$ defined by (41), we have guaranteed that the VM system will be "completely Hamiltonian."[8] For completely Hamiltonian systems the kinetic part of the dispersion operator is a formally self-adjoint function of ω,[8] so it only remains to show that $\mathbf{\Lambda}(x,\omega)$ is a formally self-adjoint function of ω for the VM system. Also, for cases where the plasma density vanishes at the conducting wall, the adjoint boundary conditions are determined by the field part of the dispersion operator alone.

Define the vectors $\mathbf{u}$ and $\mathbf{v}$ by

$$\mathbf{u}=\begin{pmatrix}\xi\\ \mathbf{a}\end{pmatrix},\quad \mathbf{v}=\begin{pmatrix}\tilde{\xi}\\ \tilde{\mathbf{a}}\end{pmatrix}. \tag{44}$$

By straightforward use of vector identities we can cast $(\mathbf{v},\mathbf{\Lambda}(x,\omega)\cdot\mathbf{u})$ into the form

$$(\mathbf{v},\mathbf{\Lambda}(x,\omega)\cdot\mathbf{u})=(\mathbf{\Lambda}(x,\omega^*)\cdot\mathbf{v},\mathbf{u})+J(\mathbf{v},\mathbf{u})/4\pi\,, \tag{45}$$

where J is given by

$$J(\mathbf{v},\mathbf{u})=\int d\boldsymbol{\sigma}\cdot[\,(\xi\nabla\tilde{\xi}^*-\tilde{\xi}^*\nabla\xi)+(i\omega/c)(\tilde{\xi}^*\mathbf{a}+\xi\tilde{\mathbf{a}}^*)+(\tilde{\mathbf{a}}^*\times\nabla\times\mathbf{a}-\mathbf{a}\times\nabla\times\tilde{\mathbf{a}}^*)\,] \tag{46}$$

and the integral in (46) is taken over the surface of the conductor. The adjoint boundary conditions are chosen to make $J(\mathbf{v},\mathbf{u})$ vanish. Thus, from (5), we see that indeed $\mathbf{\Lambda}(x,\omega)$ is a formally self-adjoint function of ω:

$$\mathbf{\Lambda}(x,\omega)=\mathbf{\Lambda}^\dagger(x,\omega)\equiv[\mathbf{\Lambda}(x,\omega^*)]^\dagger\,. \tag{47}$$

In what follows, we assume that the potential functions $\mathbf{u}$ and $\mathbf{v}$ defined in (44) satisfy the same gauge condition. Our formalism does not require us to make this choice. However, we have not investigated possible advantages of choosing different gauges for $\mathbf{u}$ and $\mathbf{v}$.

For the gauge choice $\phi_1\equiv 0$, both ξ and $\tilde{\xi}$ are identically zero, so that (47) reduces to

$$J(\mathbf{v},\mathbf{u})=\int d\sigma[\,(\hat{\mathbf{n}}\times\tilde{\mathbf{a}}^*)\cdot(\nabla\times\mathbf{a})-(\hat{\mathbf{n}}\times\mathbf{a})\cdot(\nabla\times\tilde{\mathbf{a}}^*)\,]\,, \tag{48}$$

where $d\boldsymbol{\sigma}=d\sigma\,\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}$ is the outward unit normal from the conducting surface. For the present gauge choice, the condition that the tangential components of the electric field vanish on the conductor yields the following boundary condition on the normal mode:

$$B(\mathbf{u})\equiv\hat{\mathbf{n}}\times\mathbf{a}=0. \tag{49}$$

Equation (49) makes the second term in (48) vanish. The only way that the first term in (48) can vanish for all vectors $\mathbf{a}$ that satisfy (49) is for $\tilde{\mathbf{a}}$ to satisfy

$$\tilde{B}(\mathbf{v})\equiv\hat{\mathbf{n}}\times\tilde{\mathbf{a}}=0. \tag{50}$$

Thus the adjoint boundary conditions (50) are the *same* as the boundary conditions on the normal mode (49) [i.e., $\mathscr{D}(\mathbf{\Lambda})=\mathscr{D}([\mathbf{\Lambda}]^\dagger)$]. Since (45) holds with $J=0$ for all $\mathbf{v}$ and $\mathbf{u}$ that satisfy (49), we have established that $\mathbf{\Lambda}$ is a Hermitian function of ω. Since $\mathscr{D}(\mathbf{\Lambda})=\mathscr{D}([\mathbf{\Lambda}]^\dagger)$, we also have established that $\mathbf{\Lambda}$ is a self-adjoint function of ω.

For examining the gauge choice $A_{1r}=0$, we specialize to a conducting cylinder of radius a. In this case we can Fourier analyze all perturbation quantitites in θ and z and consider each mode separately. Accordingly, we assume that the θ and z dependence of the perturbations is $\exp[i(m\theta+kz)]$. Then the integral over $d\boldsymbol{\sigma}$ in (46) gives a factor $(2\pi)^2/k$, where the period in the z direction is $2\pi/k$. Since $A_{1r}=0$, both a_{1r} and $\tilde{a}_{1r}$ are identically zero and we have from (47)

$$\begin{aligned}J(\mathbf{v},\mathbf{u})\propto&-\tilde{\xi}^*(a)\partial_r\xi(a)+\xi(a)\partial_r\tilde{\xi}^*(a)\\&+\tilde{a}_\theta^*(a)\partial_r a_\theta(a)-a_\theta(a)\partial_r\tilde{a}_\theta(a)\\&+a_z^*(a)\partial_r a_z(a)-a_z(a)\partial_r\tilde{a}_z^*(a)\,,\end{aligned} \tag{51}$$

where $\partial_r=\partial/\partial r$ and, for any $f(r)$, $\partial_r f(a)$ means $(\partial f/\partial r)|_{r=a}$. Now impose that $\mathbf{u}$ satisfy the conducting boundary conditions:

$$B(\mathbf{u})\equiv\begin{pmatrix}-m\xi(a)/a+\omega a_\theta(a)/c\\ -k\xi(a)+\omega a_z(a)/c\end{pmatrix}=0\,. \tag{52}$$

If we use (52) to eliminate $a_\theta(a)$ and $a_z(a)$ from (51), (51) becomes

$$\begin{aligned}J(\mathbf{v},\mathbf{u})\propto\xi(a)[&\partial_r\tilde{\xi}^*(a)-(mc/a\omega)\partial_r\tilde{a}_\theta^*(a)\\&-(kc/\omega)\partial_r\tilde{a}_z^*(a)]-\tilde{\xi}^*(a)\partial_r\xi(a)\\&+\tilde{a}_\theta^*(a)\partial_r a_\theta(a)+\tilde{a}_z^*(a)\partial_r a_z(a)\,.\end{aligned} \tag{53}$$

If ξ and $\mathbf{a}$ merely satisfy the conducting boundary conditions (52), then $\xi(a)$, $\partial_r\xi(a)$, $\partial_r a_\theta(a)$, and $\partial_r a_z(a)$ are arbitrary, independent quantitites on the boundary. Therefore, (53) can vanish only if the adjoint boundary conditions are

$$\tilde{B}(\mathbf{v}) \equiv \begin{pmatrix} \partial_r \tilde{\xi}^*(a) - (mc/a\omega)\,\partial_r \tilde{a}_\theta^*(a) - (kc/\omega)\,\partial_r \tilde{a}_z^*(a) \\ \tilde{\xi}(a) \\ \tilde{a}_\theta(a) \\ \tilde{a}_z(a) \end{pmatrix} = 0 . \tag{54}$$

Since the boundary conditions on the normal mode are different than the adjoint boundary conditions ($B \neq \tilde{B}$), the domain of the dispersion operator is different than the domain of the adjoint dispersion operator. Therefore, the dispersion operator for the gauge choice $A_{l_r} = 0$ is not a self-adjoint function of ω. Furthermore, $J(\mathbf{v},\mathbf{u})$ will *not* be zero for $\mathbf{v}$ and $\mathbf{u}$ both in $\mathscr{D}(\Lambda)$. Thus the dispersion operator for the gauge choice $A_{l_r} = 0$ is also not a Hermitian function of ω.

Equation (54) specifies *four* independent conditions to be satisfied by the dual eigenfunction at the conducting wall. Comparing this to the *two* boundary conditions of (52) imposed on the normal mode at the conductor, we see that the adjoint boundary conditions (54) will allow only the *trivial* solution for the dual eigenfunction. Obviously we cannot calculate eigenfrequencies by using a left function or a left basis which is identically zero. However, we show now that we can *still* obtain variational results by restricting the class of functions $\mathbf{u}$ and $\mathbf{v}$ to satisfy what we term *variational* boundary conditions.

Theorems II and III both invoke Theorem I of Ref. 4, so let us recall in the proof of Theorem I the key step where the question of boundary conditions enters. In Theorem I we assumed we had first-order approximations to an eigenfunction and its dual:

$$\psi = \nu_0 + \epsilon \nu_1, \tag{55}$$

$$\tilde{\psi} = \tilde{\nu}_0 + \epsilon \tilde{\nu}_1, \tag{56}$$

where ν_0 is the exact normal mode and $\tilde{\nu}_0$ is a solution of the adjoint dispersion equation:

$$D(\omega_0)\nu_0 = 0, \quad B(\nu_0) = 0, \tag{57}$$

$$[D(\omega_0)]^\dagger \tilde{\nu}_0 = 0 . \tag{58}$$

We require that ν_0 satisfy the physical boundary conditions, but we have not yet specified the boundary conditions on $\tilde{\nu}_0$. Here ν_1 and $\tilde{\nu}_1$ are suitably normalized vectors that satisfy the same boundary conditions as ν_0 and $\tilde{\nu}_0$, respectively.

The question of boundary conditions enters in Theorem I in going from Eq. (11) to Eq. (12) of Ref. 4. There we assumed that the boundary conditions were such that

$$\{\tilde{\nu}_0, D(\omega_0)\nu_1\} = 0. \tag{59}$$

By the definition of the adjoint operator, the left-hand side of (59) has the form

$$\{\tilde{\nu}_0, D(\omega_0)\nu_1\} = \{[D(\omega_0)]^\dagger \tilde{\nu}_0, \nu_1\} + J(\tilde{\nu}_0, \nu_1) . \tag{60}$$

Therefore, because of (58), (59) will hold if the boundary conditions on $\tilde{\nu}_0$ and ν_1 are such that

$$J(\tilde{\nu}_0, \nu_1) = 0. \tag{61}$$

The *variational boundary conditions* alluded to earlier are those boundary conditions on $\tilde{\nu}_0$ and ν_1, denoted by $\tilde{B}'(\tilde{\nu}_0) = 0$ and $B'(\nu_1) = 0$, respectively, that make (61) valid for nontrivial $\tilde{\nu}_0$ that are solutions of (58) and for nontrivial ν_1 that already satisfy the physical boundary conditions,

$$B(\nu_1) = 0. \tag{62}$$

Note that, for our purpose here of making (59) valid, $\tilde{\nu}_0$ does *not* represent any member of a general class of functions, but rather $\tilde{\nu}_0$ is simply a solution of the adjoint dispersion equation. To avoid that $\tilde{\nu}_0$ be trivial, we require that the number of variational boundary conditions on $\tilde{\nu}_0$ be the *same* as the number of physical boundary conditions on ν_0. On the other hand, ν_1 in (61) *does* represent any element of a class of functions that satisfy the physical boundary conditions (62). We will want to require that ν_1 satisfy *additional* boundary conditions that are consistent with any solution of the dispersion equation.

Now let us leave this general discussion of variational boundary conditions and, as an example, derive specific variational boundary conditions for the VM system in cylindrical geometry with conducting boundary conditions on the normal mode and with the gauge choice $A_{l_r} \equiv 0$. We consider the form of $J(\mathbf{v},\mathbf{u})$ given by (53), where now $\mathbf{v}$ represents a solution of the adjoint dispersion equation and where $\mathbf{u}$ satisfies the conducting boundary conditions (52). Since $\mathbf{v}$ is a solution of the adjoint dispersion equation, and since the dispersion operator is a formally self-adjoint function of ω at the conducting wall (where the plasma source terms vanish in the VM system), $\mathbf{v}$ must satisfy the r component of the adjoint Ampère equation,

$$(\omega^*/c)\partial_r \tilde{\xi}(a) - (m/a)(1/a + \partial_r)\tilde{a}_\theta(a) - k\,\partial_r \tilde{a}_z(a) = 0 . \tag{63}$$

Using (63) in (53) we obtain

$$J(\mathbf{v},\mathbf{u}) \propto (cm/\omega a^2)\xi(a)\tilde{a}_\theta^*(a) - \tilde{\xi}^*(a)\partial_r \xi(a) + \tilde{a}_\theta^*(a)\partial_r a_\theta(a) + \tilde{a}_z^*(a)\partial_r a_z(a) . \tag{64}$$

As one of our variational boundary conditions on $\mathbf{u}$, let us require that $\mathbf{u}$ satisfy the r component of Ampère's equation at the wall,

$$(\omega/c)\partial_r \xi(a) - (m/a)(1/a + \partial_r)a_\theta(a) - k\,\partial_r a_z(a) = 0 . \tag{65}$$

If we use (63) and (65) in (64) we obtain

$$J(\mathbf{v},\mathbf{u}) \propto \partial_r \xi(a)[-(m/a)\tilde{\xi}(a) + (\omega^*/c)\tilde{a}_\theta(a)]^* + \partial_r a_z(a)[(m/a)\tilde{a}_z(a) - k\tilde{a}_\theta(a)]^* . \tag{66}$$

Since $\partial_r \xi(a)$ and $\partial_r a_\theta(a)$ are arbitrary quantities on the boundary, the only way that (66) can be made to vanish is by requiring $\tilde{\xi}$ and $\tilde{\mathbf{a}}$ to satisfy the conditions

$$-(m/a)\tilde{\xi}(a) + (\omega^*/c)\tilde{a}_\theta(a) = 0 \tag{67}$$

and

$$(m/a)\tilde{a}_z(a) - k\tilde{a}_\theta(a) = 0 . \tag{68}$$

Equations (67) and (68) imply

$$-k\tilde{\xi}(a)+(\omega^*/c)a_z(a)=0\,. \tag{69}$$

Thus we will obtain a variational result for the eigenfrequency if we choose left and right functions that satisfy the following variational boundary conditions:

$$B'(\mathbf{u})\equiv\begin{pmatrix}(\omega/c)\partial_r\xi(a)-(m/a)(1/a+\partial_r)a_\theta(a)-k\,\partial_r a_z(a)\\ -(m/a)\xi(a)+(\omega/c)a_\theta(a)\\ -k\xi(a)+(\omega/c)a_z(a)\end{pmatrix}=0 \tag{70}$$

and

$$\tilde{B}'(\mathbf{v})\equiv\begin{pmatrix}(\omega^*/c)\,\partial_r\tilde{\xi}(a)-(m/a)(1/a+\partial_r)\tilde{a}_\theta(a)-k\,\partial_r\tilde{a}_z(a)\\ -(m/a)\tilde{\xi}(a)+(\omega^*/c)\tilde{a}_\theta(a)\\ -k\tilde{\xi}(a)+(\omega^*/c)\tilde{a}_z(a)\end{pmatrix}=0\,. \tag{71}$$

Notice from (70) and (71) the interesting result that the variational adjoint boundary conditions are the *same* as the variational boundary conditions with ω replaced by ω^*! Notice also that this implies that the dual of an eigenfunction $\Phi(\omega)$ equals $\Phi(\omega^*)$, a result already obtained in Refs. 4, 8, and 9. Since (71) is meant to be used only with left functions $\mathbf{v}$ that are solutions of the adjoint dispersion equation, the first element of (71) will be satisfied *automatically*, since it is just one of the adjoint dispersion equations at the wall. [At the wall the plasma terms in the VM system vanish and the full VM dispersion operator just becomes the field operator $\mathbf{A}$ of (43) with $\omega_{ps}=0$.] Thus we really impose only *two* new conditions on $\mathbf{v}$, which is the correct number of boundary conditions to guarantee that a nontrivial $\mathbf{v}$ exist. Also, there is no difficulty in requiring the right functions $\mathbf{u}$ to satisfy conducting boundary conditions as well as the r component of Ampère's equation at the wall.

We reiterate that it may be possible to obtain variational results for eigenfrequencies even though the adjoint boundary conditions allow only the trivial solution for the dual eigenfunction. When obtaining eigenfrequencies as roots of a dispersion functional, a variational result will follow if the left and right functions satisfy appropriate variational boundary conditions. Similarly, when obtaining eigenfrequencies by solving a dispersion matrix problem, a variational result will follow if the left and right bases satisfy appropriate variational boundary conditions.

V. SUMMARY

We have seen that, for unbounded operators, such as integrodifferential dispersion operators, the dual eigenfunctions satisfy adjoint boundary conditions that may or may not be the same as the boundary conditions imposed on the eigenfunction itself. If the two sets of boundary conditions are different, then a dispersion functional constructed assuming that the dual and the eigenfunction lie in the same function space generally cannot be used as the basis for a variational calculation of the eigenfrequency. A variational result follows for eigenfrequencies calculated from solving the finite-dimensional matrix problem that approximates the dispersion equation if the dispersion matrix is constructed from left and right bases that satisfy the adjoint boundary conditions and boundary conditions on the normal mode, respectively. If the adjoint boundary conditions and boundary conditions on the normal mode are the same, then the left and right bases should be the same. When solving for the normal modes using potentials, the adjoint boundary conditions are gauge dependent, even for conducting boundaries. If the adjoint boundary conditions only allow for the trivial solution of the adjoint dispersion equation, it still may be possible to obtain variational results if the left and right functions satisfy appropriate variational boundary conditions.

We close with a flow chart in Fig. 1 that summarizes the courses of action that must be taken to obtain variational results for a dispersion operator which may have any of a wide spectrum of symmetry properties. For the purpose of this chart we use the notation of Eqs. (55)–(62). When computing eigenfrequencies with a dispersion functional, we seek solutions of

$$\{\tilde{\psi},D(\omega')\psi\}=0, \tag{72}$$

where the eigenfrequency ω' is only approximate, since the right and left functions are only first-order approximations to the eigenfunction and its dual [see (55) and (56)]. When computing eigenfrequencies from a dispersion matrix, we seek solutions of

$$\det D_{nn'}(\omega')=0\,, \tag{73}$$

where the dispersion matrix is defined by

$$D_{nn'}(\omega)=\{\tilde{\eta}_n,D(\omega)\eta_{n'}\}\,. \tag{74}$$

Equation (74) is obtained by expanding ψ and $\tilde{\psi}$ in appropriate right and left bases, respectively:

$$\psi(x;\omega')=\sum_{n=1}^{N}a_n(\omega')\eta_n(x) \tag{75}$$

and

$$\tilde{\psi}(x;\omega')=\sum_{n'=1}^{N}\tilde{a}_{n'}(\omega')\tilde{\eta}_{n'}(x)\,. \tag{76}$$

The boundary conditions on ψ (or the right basis) will be denoted by $B_1=0$, where B_1 is either B, which represents the physical boundary conditions on the normal mode, or B', which represents the variational boundary conditions on the normal mode. The boundary conditions on $\tilde{\psi}$ (or the left basis) will be denoted by B_2, where B_2 is either B (if the boundary conditions on the eigenfunction and its dual are the same), $\tilde{B}$, which represents the adjoint boundary condi-

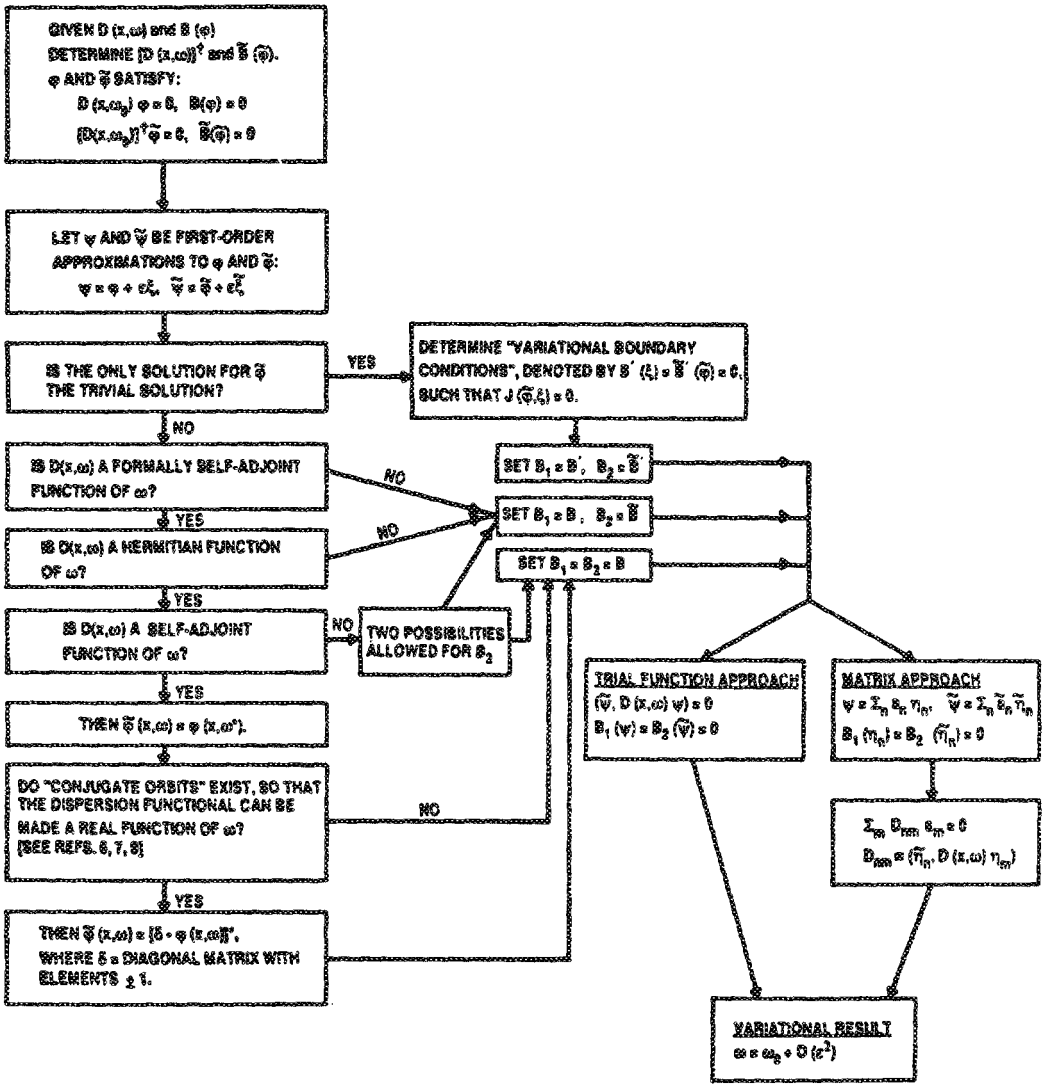


FIG. 1. Flow chart showing what must be done to achieve variational results for dispersion operators with a variety of symmetry properties.

tions, or $\tilde{B}'$, which represents the variational adjoint boundary conditions. If ω_0 is the exact eigenfrequency, the calculation of the approximate eigenfrequency ω' is said to be variational if

$$\omega' = \omega_0 + O(\epsilon^2), \tag{77}$$

where ϵ is the small number appearing in (55) and (56). There is one possibility which aids in obtaining variational results that was not discussed in detail in this paper, but which is included in the flow chart for completeness. This possibility is discussed in detail in Ref. 9 but was discovered first by Berk *et al.* in Ref. 6. The basic idea is this: If the dispersion operator and the equilibrium possess certain properties, it is possible to conclude that the dispersion functional or dispersion matrix is a real function of ω [see Eq. (12)]. Then it follows that the complex conjugate of the exact dual eigenfunction $\tilde{\nu}_0$ is merely proportional to the exact eigenfunction ν_0. In this case, if ψ is a first-order accurate approximation to ν_0, then $\tilde{\psi}$, a first-order accurate approximation to $\tilde{\nu}_0$, follows immediately.

To use the flow chart, begin with the first box. The rules are that one can exit a box only along a path that goes away from the box, and one can exit an intersection of paths only along a path that goes away from the intersection.

The flow chart illustrates that it is possible to obtain variational results for a wide range of operators and that, contrary to the common belief, having Hermitian or self-adjoint operators is *not* necessary.

ACKNOWLEDGMENTS

We wish to thank Professor Howard Conner for useful discussions on this work. We thank Dr. Daniel C. Barnes and Robert B. Webster for reading the manuscript and offering valuable suggestions. We thank Dr. Herbert L. Berk for comments he has made on this work. We also acknowledge the careful and constructive reading of our manuscript by the referee.

This work was performed under the auspices of the U.S. Department of Energy.

APPENDIX: PROOF OF THEOREM III

In this Appendix we state and prove Theorem III of Sec. III. Again we suppress the x dependence of the operators and eigenfunctions.

Theorem III: Assume all eigenvectors in this proof satisfy appropriate boundary conditions. Let ω_0 be a simple zero of one, and only one, eigenvalue of the operator $D(\omega)$, and let ϕ be the corresponding eigenvector normalized to unity:

$$D(\omega_0)\phi = 0, \quad \|\phi\| = 1. \tag{A1}$$

Let S be a subspace within which ϕ can be approximated to order ϵ:

$$\phi = P\phi + \epsilon\xi, \tag{A2}$$

where P is a projection operator from the Hilbert space $\mathscr{H}$ onto S, and ξ is a vector normalized to unity. Assume that the dual $\tilde{\phi}$ can be approximated to order ϵ in a subspace $\tilde{S}$:

$$\tilde{\phi} = \tilde{P}\tilde{\phi} + \epsilon\tilde{\xi}, \tag{A3}$$

where $\tilde{P}$ is the projection operator from $\mathscr{D}([D]^\dagger)$ onto $\tilde{S}$, and $\tilde{\phi}$ and $\tilde{\xi}$ are vectors normalized to unity. Both ξ and $\tilde{\xi}$ are assumed to have bounded derivatives, as required. Let $\underline{D}(\omega)$ be the operator $D(\omega)$ restricted to the subspaces S and $\tilde{S}$:

$$\underline{D}(\omega) = \tilde{P}D(\omega)P. \tag{A4}$$

Let ω' yield a zero eigenvalue of $\underline{D}(\omega)$,

$$\underline{D}(\omega')\underline{\phi} = 0, \tag{A5}$$

where $\underline{\phi}$ is the corresponding eigenvector in S normalized to unity,

$$P\underline{\phi} = \underline{\phi}. \tag{A6}$$

Assume that ω' is near ω_0 in the sense that

$$|\omega' - \omega_0| < \epsilon^\alpha, \quad \text{for some } \alpha \text{ satisfying } 0 < \alpha \leqslant 1. \tag{A7}$$

Then, the eigenvector $\underline{\phi}$ approximates ϕ to order ϵ^β, where $\alpha \leqslant \beta \leqslant 1$. Also, since ω_0 is a simple root of the dispersion relation associated with (A1), the dispersion relation associated with (A5) will have a root ω' that approximates ω_0 to order $\epsilon^{2\beta}$.

Proof: By the assumed analyticity of $D(\omega)$ as in Theorem I, expand $D(\omega')$ in a Taylor series,

$$D(\omega') = D(\omega_0) + (\omega' - \omega_0)D'(\omega_0) + \cdots, \tag{A8}$$

where we assume that

$$D'(\omega_0)\phi \equiv \left.\frac{\partial D(\omega)}{\partial\omega}\right|_{\omega=\omega_0}\phi \neq 0.$$

Operate on (A1) with $\tilde{P}$ and use (A2), (A3), and (A8):

$$0 = \tilde{P}D(\omega_0)\phi$$
$$= \tilde{P}[D(\omega') - (\omega' - \omega_0)D'(\omega_0) + \cdots](P\phi + \epsilon\xi),$$

or, to first order by using (A4),

$$-\underline{D}(\omega')\phi$$
$$= \epsilon\tilde{P}D(\omega_0)\xi - (\omega' - \omega_0)\tilde{P}D'(\omega_0)P\phi + \cdots, \tag{A9}$$

where dots represent higher-order terms in $(\omega' - \omega_0)$. Add (A9) to (A5):

$$\underline{D}(\omega')(\underline{\phi} - \phi)$$
$$= \epsilon\tilde{P}D(\omega_0)\xi - (\omega' - \omega_0)\tilde{P}D'(\omega'_0)P\phi + \cdots. \tag{A10}$$

In view of definition (A3) and since $P^2 = P$, we can replace ϕ by $P\phi$ on the left in (A10),

$$\underline{D}(\omega')(\underline{\phi} - P\phi)$$
$$= \epsilon\tilde{P}D(\omega_0)\xi - (\omega' - \omega_0)\tilde{P}D'(\omega_0)P\phi + \cdots. \tag{A11}$$

In analogy with definition (A4), let $[\underline{D}(\omega)]^\dagger$ be the operator $[D(\omega)]^\dagger$ restricted to the subspaces S and $\tilde{S}$ (projection operators are always Hermitian):

$$[\underline{D}(\omega)]^\dagger = P[D(\omega)]^\dagger\tilde{P}. \tag{A12}$$

The right-hand side of (A12) is indeed the adjoint of (A4). Let N be the projection operator onto the null space of $\underline{D}(\omega')$, so that for any f in S we have

$$\underline{D}(\omega')Nf = 0. \tag{A13}$$

The vector f has the unique decomposition

$$f = Nf + (1 - N)f. \tag{A14}$$

Since $\underline{\phi} - P\phi$ is a vector in S, using (A12) and (A13) we see that

$$\underline{D}(\omega')(\underline{\phi} - P\phi) = \underline{D}(\omega')(1 - N)(\underline{\phi} - P\phi). \tag{A15}$$

If $\tilde{N}$ is the projection operator onto the null space of $[\underline{D}(\omega')]^\dagger$, define

$$\underline{D}_N(\omega') = (1 - \tilde{N})\underline{D}(\omega')(1 - N). \tag{A16}$$

Within the subspace orthogonal to the null space of $\underline{D}(\omega')$, D_N has an inverse, D_N^{-1}. Operate with $(1 - \tilde{N})$ on (A11), use (A15), and solve for $\underline{\phi} - P\phi$. In view of (A2) we can write the solution in the form

$$\underline{\phi} = \phi - \epsilon\xi + D_N^{-1}(1 - \tilde{N})$$
$$\times[\epsilon\tilde{P}D(\omega_0)\xi - (\omega' - \omega_0)\tilde{P}D'(\omega_0)P\phi + \cdots]. \tag{A17}$$

By virtue of (A7) we have

$$\underline{\phi} = \phi + O(\epsilon^\beta), \tag{A18}$$

where

$$\alpha \leqslant \beta \leqslant 1. \tag{A19}$$

This is the first conclusion of the theorem. A similar argument shows that $\underline{\tilde{\phi}}$ approximates $\tilde{\phi}$ to order ϵ^β. In concluding the result (A18) we are assuming that ϵ is small enough that there are no other roots of the dispersion relation in the neighborhood of ω_0 that is specified by (A7). A simple extension of Theorem I shows that if $\underline{\phi}$ and $\underline{\tilde{\phi}}$ approximate ϕ and $\tilde{\phi}$, respectively, to order ϵ^β, then the eigenfrequency calculated as a root of the dispersion functional will approximate ω_0 to order $\epsilon^{2\beta}$. Thus the second conclusion of Theorem III is proved. Furthermore, it is easy to show that the eigenfrequency ω' calculated from the finite-dimensional approximation to the dispersion equation is exactly the *same* value of ω that would be calculated using the approximations $\underline{\tilde{\phi}}$ and $\underline{\phi}$ as trial functions in a dispersion functional. That is, since $\underline{\phi}$ lies in S and $\underline{\tilde{\phi}}$ lies in $\tilde{S}$,

$$(\underline{\tilde{\phi}}, \underline{D}(\omega')\underline{\phi}) = (\tilde{P}\underline{\tilde{\phi}}, D(\omega')P\underline{\phi}) = (\underline{\tilde{\phi}}, D(\omega')\underline{\phi}). \tag{A20}$$

The root of the leftmost member of (A20) is the eigenfrequency calculated from the finite-dimensional matrix problem, and the root of the rightmost member of (A20) is the eigenfrequency calculated as a root of a dispersion functional.

[1] H. L. Berk and R. R. Dominguez, J. Plasma Phys. **18**, 37 (1977).
[2] R. D. Hazeltine and D. W. Ross, Phys. Fluids **27**, 1140 (1978).

[3]D. W. Ross and S. M. Mahajan, Phys. Rev. Lett. **40**, 324 (1978).
[4]J. L. Schwarzmeier and K. R. Symon, J. Math. Phys. **20**, 979 (1979).
[5]J. Johner and E. K. Maschke, Plasma Phys. **22**, 679 (1980).
[6]H. L. Berk, R. R. Dominguez, and E. K. Maschke, Phys. Fluids **24**, 2245 (1981).
[7]R. R. Dominguez and H. L. Berk, Phys. Fluids **27**, 1147 (1984).
[8]H. R. Lewis and K. R. Symon, J. Math. Phys. **20**, 413, 2372(E) (1979).
[9]J. L. Schwarzmeier, D. C. Barnes, H. R. Lewis, and C. E. Seyler, Phys. Fluids **30**, 2414 (1987).
[10]C. E. Seyler and J. P. Freidberg, Phys. Fluids **23**, 331 (1980).
[11]C. E. Seyler and H. R. Lewis, J. Plasma Phys. **27**, 37 (1982).
[12]H. R. Lewis, D. C. Barnes, J. L. Schwarzmeier, and C. E. Seyler, Phys. Fluids **28**, 3546 (1985).
[13]I. Stakgold, *Boundary Value Problems of Mathematical Physics* (Macmillan, New York, 1969), Vol. I, p. 170.
[14]See Ref. 13, p. 70.
[15]See Ref. 13, p. 172.
[16]See Ref. 13, p. 212.

Spectral properties of an optical polaron in a magnetic field

H. Löwen
Institut für Physik der Universität Dortmund, D-4600 Dortmund 50, West Germany

(Received 30 January 1987; accepted for publication 13 January 1988)

An optical polaron, which is exposed to a homogeneous magnetic field, is considered. Making use of functional analytical methods of Fröhlich [Fortschr. Phys. **22**, 159 (1974)], it is proved that the ground-state energy, the magnetic polaron mass, and the number of virtual phonons in the ground state are analytical functions of the electron–phonon coupling parameter and the magnetic field strength. Consequently, a discontinuous stripping transition, which was claimed recently by several authors, does not exist. In fact, some authors have stated that the discontinuities they encounter might indeed be artifacts due to the approximation. The spectrum of the momentum-decomposed Fröhlich Hamiltonian is analyzed; bounds and smoothness properties of the ground state and the discrete excited states are derived. All results hold also for lower spatial dimensions.

I. INTRODUCTION AND STATEMENT OF THE PROBLEM

In the present paper we discuss spectral properties of the momentum-decomposed Fröhlich Hamiltonian of an optical polaron in a constant uniform magnetic field.

The standard (three-dimensional) polaron model is defined by the well-known Hamiltonian H_F, proposed by Fröhlich, Pelzer, and Zienau,[1]

$$H_F = \frac{1}{2m}(\mathbf{p} + |e|\mathbf{A}(\mathbf{x}))^2 + \int d^3k\, \hbar\omega(\mathbf{k}) a_{\mathbf{k}}^+ a_{\mathbf{k}} + \int d^3k\, \alpha^{1/2}(g(\mathbf{k}) a_{\mathbf{k}} e^{i\mathbf{k}\mathbf{x}} + \text{H.c.}) \tag{1}$$

with

$$\omega(\mathbf{k}) = \omega_0 > 0 \tag{2}$$

and

$$g(\mathbf{k}) = \hbar\omega_0(\hbar/2m\omega_0)^{1/4}(4\pi)^{1/2}(2\pi)^{-3/2}|k|^{-1}$$

EL4

$$\equiv \Omega|k|^{-1}. \tag{3}$$

Here, m, $\mathbf{x}$, $\mathbf{p}$ are the mass, the position, and momentum operator of the (spinless) electron; $\mathbf{k}$, $\omega(\mathbf{k})$ are the wave vector and frequency of the phonons (i.e., spinless bosons); $g(\mathbf{k})$ is the electron–phonon coupling, α being the dimensionless electron–phonon coupling parameter, and $|e|$ the elementary charge. As usual, we set henceforth $\hbar = \omega_0 = m = |e| = 1$ and keep α and B as the only parameters. Let the magnetic field $\mathbf{B} = (0,0,B)$, $B > 0$, be along the x_3 axis. Then, in the Landau gauge, the vector potential $\mathbf{A}$ may be written as $\mathbf{A}(\mathbf{x}) = (0,Bx_1,0)$.

In the case of free optical polarons ($B = 0$), the analytical properties of the ground-state energy were unclear for a long time, until Spohn[2] applied the beautiful functional analytical work of Fröhlich[3] directly to prove the analyticity of the ground-state energy and the polaron mass as a function of the coupling parameter. In this paper, we want to generalize this result to arbitrary magnetic fields. Making use of operator methods developed by Fröhlich,[3] we do show that the ground-state energy, the ground-state wave function, and expectation values of the ground state as well as the magnetic polaron mass are analytical functions in the coupling parameter α and the magnetic field strength B ($B > 0$). The same holds true for the energies and wave functions of the momentum-decomposed discrete excited states, i.e., the Landau levels below the one-phonon continuum.

This paper represents the first rigorous study of analytical properties in optical polaron systems for arbitrary coupling and arbitrary magnetic field strength at zero temperature. Only for very small α, Alvarez-Estrada[4] has established analytical properties by using several perturbation approaches. We note that Gerlach and the author have proved in a previous work[5] under rather general conditions that the (formal) free energy of an acoustical or optical polaron system, exposed to a homogeneous magnetic field, is analytic in the temperature T ($0 < T < \infty$), coupling parameter α, and magnetic field strength B ($0 < B$). But the limit $T \to 0$ was not studied there.

The most important consequence of our results concerns a so-called stripping transition, which was first studied by Peeters and Devreese.[6,7] In a series of papers, Peeters and Devreese have calculated the ground-state energy,[6,7] the polaron mass,[7,8] the polaron radius[9] as well as the number of virtual phonons in the ground state[9] and the magnetoabsorption spectrum[10] within the anisotropic Feynman approximation. They do find nonanalytical behavior of these quantities at certain critical values of α and B. They indicate that this might be an artifact of their approximation. We note that Gorshkov, Zabrodin, Rodriguez, and Fedyanin[11] have already questioned these discontinuous transitions. A similar nonanalytical behavior is found for a two-dimensional polaron (see the recent work of Wu Xiaoguang, Peeters, and Devreese[12]). Within the Fock approximation, Lépine and Matz[13] and Lépine[14] get discontinuous transitions, too. In fact, there may be large changes in the polaron quantities as a function of α or B, but these changes are continuous. All discontinuities reported in the references quoted above are artifacts of the the approximations rather than intrinsic properties of the Fröhlich Hamiltonian. In fact, in Refs. 6–9, Peeters and Devreese carefully stated that the discontinuities they encounter could be artifacts of their approximation.

The basic steps of the proof are as follows: In Sec. II, we introduce the corresponding momentum decomposed Hamiltonian $H(\mathbf{Q})$ whose spectral properties are under study. Two different cutoffs are successively introduced which

 0022-2488/88/061498-07$02.50

clearly have to be removed later: a UV cutoff in the coupling and a lattice cutoff which leads to a discrete phonon momentum space. Then, it is proved that the ground state energy of the momentum decomposed cutoff Hamiltonian belongs to the discrete part of the spectrum, if the momentum $\mathbf{Q}$ fulfills a simple inequality. After that, we show that the same result is also valid if the lattice cutoff is removed. In Sec. III, it is proved that this result even holds, if the UV cutoff is removed, using a dressing transformation. After having shown that the ground state of $H(\mathbf{Q})$ is nondegenerate (see Sec. IV), we derive bounds on the ground state energy of $H(\mathbf{Q})$ (see Sec. V) that guarantee that the inequality mentioned above is fulfilled. Since this finally implies that the ground state of $H(\mathbf{Q})$ is discrete and nondegenerate, we are able to apply analytic perturbation theory in Sec. VI which establishes smoothness properties of the ground state and the discrete excited states. Finally, in Sec. VII we give other examples, to which our methods are applicable.

II. SPECTRAL PROPERTIES OF THE HAMILTONIAN WITH UV CUTOFF

First, we perform a Lee–Low–Pines transformation. Defining

$$U = \exp(-i(P_2 x_2 + P_3 x_3)), \tag{4}$$

where

$$\mathbf{P} = \int d^3k\, \mathbf{k} a_{\mathbf{k}}^+ a_{\mathbf{k}} \tag{5}$$

is the phonon momentum, we obtain

$$\tilde{H}_F := U^{-1} H_F U = \mathbf{G}^2/2 + H_{0ph} + H_I \tag{6a}$$

with

$$H_{0ph} = \int d^3k\, a_{\mathbf{k}}^+ a_{\mathbf{k}}, \tag{6b}$$

$$H_I = \alpha^{1/2} \int d^3k\, (g(\mathbf{k}) a_{\mathbf{k}} \exp(ik_1 x_1) + \text{H.c.}) \tag{6c}$$

and

$$\mathbf{G} = (p_1, Bx_1 + p_2 - P_2, p_3 - P_3). \tag{6d}$$

Furthermore,

$$U^{-1} p_i U = p_i - P_i, \quad i = 2,3. \tag{7}$$

Clearly, $\tilde{H}_F$ does not depend on x_2 and x_3, i.e., p_2 and p_3, now playing the role of the total momentum [see (7)], are conserved quantities which may be replaced by c numbers Q_2 and Q_3. Mathematically, this means that $\tilde{H}_F$ admits a direct integral decomposition as follows:

$$\tilde{H}_F = \int_{\oplus} dQ_2\, dQ_3\, H(\mathbf{Q}), \quad \mathbf{Q} = (0, Q_2, Q_3). \tag{8}$$

Here, $H(\mathbf{Q})$, being the Hamiltonian corresponding to fixed total momentum Q_2 and Q_3, is given by

$$H(\mathbf{Q}) = H_0(\mathbf{Q}) + H_I, \tag{9a}$$

$$H_0(\mathbf{Q}) = H_{0ph} + \mathbf{G}(\mathbf{Q})^2/2, \tag{9b}$$

$$\mathbf{G}(\mathbf{Q}) = (p_1, Bx_1 + Q_2 - P_2, Q_3 - P_3). \tag{9c}$$

It is well-known that, for $B > 0$, the spectrum of $H(\mathbf{Q})$ is independent of Q_2 (see, e.g., Devreese[15]). Nevertheless we retain the trivial Q_2 dependence.

For the underlying Hilbert space $\mathscr{H}$, it is convenient to take

$$\mathscr{H} = F \otimes L^2(\mathbb{R}), \tag{10}$$

where

$$F = \bigoplus_{m=0}^{\infty} L^2(\mathbb{R}^3)^{\textcircled{S} m} \tag{11}$$

is the usual Fock space of the phonons, $\textcircled{S}$ denoting the symmetrical tensor product. We define UV cutoff Hamiltonians $H_r(\mathbf{Q})$, H_{Ir} by replacing $g(\mathbf{k})$ in (9a) and (6c) by $g_r(\mathbf{k}) \equiv g(\mathbf{k}) \cdot \theta(r - k)$, $0 < r < \infty$. Then, according to a result of Nelson,[16] we may state: For all $\epsilon > 0$ there exists a number $b(r,\epsilon) < \infty$ such that

$$\|H_{Ir}\psi\| \leqslant \epsilon \|H_{0ph}\psi\| + b\,\|\psi\| \leqslant \epsilon \|H_0(\mathbf{Q})\psi\| + b\,\|\psi\|, \quad \text{for all } |\psi\rangle \in D(H_0(\mathbf{Q})). \tag{12}$$

Clearly, $H_0(\mathbf{Q})$ is self-adjoint and bounded from below. Consequently, the Kato–Rellich theorem[17] ensures us that $H_r(\mathbf{Q})$ is self-adjoint and bounded from below.

Now we introduce a second cutoff: We replace the phonon momentum space $\mathbb{R}^3$ by a momentum lattice Γ_d (see Refs. 3 and 18 for a detailed discussion),

$$\Gamma_d = \{\mathbf{k} \in \mathbb{R}^3 | k_j = n_j/\Lambda_d,\ n_j \in \mathbb{Z},\ \Lambda_d = 2^d \Lambda_0, \Lambda_0 \in \mathbb{R}^+,\ j = 1,2,3\}. \tag{13}$$

To each $\mathbf{k} \in \mathbb{R}^3$ we associate a $\mathbf{k}|_d \in \Gamma_d$, namely,

$$\mathbf{k}|_d = (n_1, n_2, n_3)/\Lambda_d, \quad n_j = \langle k_j \Lambda_d \rangle, \tag{14}$$

where

$$\langle a \rangle \equiv \begin{cases} \text{largest integer} \leqslant a, & \text{if } a < 0, \\ \text{smallest integer} > a, & \text{if } a \geqslant 0. \end{cases}$$

The continuum limit is obtained by taking the limit $d \to \infty$. We now define a subspace $S_d \subseteq L^2(\mathbb{R}^3)$ of step functions,

$$f \in S_d \Leftrightarrow f(\mathbf{k}) = f(\mathbf{k}|_d). \tag{15}$$

For $g \in L^2(\mathbb{R}^3)$ let $g|_d$ denote the orthogonal projection of g onto S_d. We need some further definitions,

$$F_d \equiv \bigoplus_{m=0}^{\infty} S_d^{\textcircled{S} m}, \tag{16}$$

$$F_d^{\perp} \equiv \left(\bigoplus_{m=1}^{\infty} S_d^{\perp \textcircled{S} m} \right) \textcircled{S} F_d. \tag{17}$$

Then we have

$$F = F_d \oplus F_d^{\perp}. \tag{18}$$

We introduce the d cutoff in the Hamiltonian in the following way:

$$H_{dr}(\mathbf{Q}) = H_{0d} + H_{Idr}, \tag{19}$$

$$H_{0d} = \int d^3k\, a_{\mathbf{k}}^+ a_{\mathbf{k}} + \frac{p_1^2}{2} + \frac{(Bx_1 + Q_2 - P_2|_d)^2}{2} + \frac{(Q_3 - P_3|_d)^2}{2}, \tag{20}$$

$$H_{Idr} = \alpha^{1/2} \int d^3k\, ((g_r(\mathbf{k})|_d \exp(ik_1|_d \cdot x_1) a_{\mathbf{k}} + \text{H.c.}), \tag{21}$$

where now

$$\mathbf{P}|_d = \int d^3k\, \mathbf{k}|_d a_{\mathbf{k}}^+ a_{\mathbf{k}}.$$

One easily verifies that (12) remains true for the new d cutoff Hamiltonian $H_{dr}(\mathbf{Q})$. Consequently, $H_{dr}(\mathbf{Q})$ is self-adjoint and bounded from below, too. Let $E(r,\mathbf{Q}), E(d,r,\mathbf{Q})$ be the ground-state energy of the Hamiltonians $H_r(\mathbf{Q})$, $H_{dr}(\mathbf{Q})$.

Lemma 2.1: Suppose that the momentum $\mathbf{Q}$ is such that

$$\inf_{\mathbf{k}} (E(r,d\,\mathbf{Q} - (0,k_2,k_3)) + 1 - E(r,d,\mathbf{Q})) \equiv \Delta(r,d,\mathbf{Q}) > 0. \tag{22}$$

Then the interval $[E(r,d,\mathbf{Q}), E(r,d,\mathbf{Q}) + \Delta(r,d,\mathbf{Q})[$ belongs to the discrete part of $\mathrm{spec}(H_{dr}(\mathbf{Q}) \upharpoonright F_d \otimes L^2(\mathbb{R}))$, where $\upharpoonright$ denotes (as usual) the restriction.

Proof: First, we define a new subspace $J \subseteq S_d$ by

$$f \in J \Leftrightarrow f(\mathbf{k}) \equiv 0, \quad \text{for } |\mathbf{k}|_d > r + 3/\Delta_d. \tag{23}$$

Moreover, let $\tilde{\Gamma}_d \equiv \{\mathbf{k} \in \Gamma_d \,||\mathbf{k}| \leqslant r + 3/\Lambda_d\}$ and

$$W \equiv \overset{\infty}{\underset{m=0}{\oplus}} J^{\circledS m}, \quad W_\perp \equiv \left(\overset{\infty}{\underset{m=1}{\oplus}} J^{\perp \circledS m} \right),$$
$$W^\perp = W_\perp \circledS W. \tag{24}$$

Clearly, $F_d = W \oplus W^\perp$ and $H_{dr}(\mathbf{Q})$, H_{Idr} leave W invariant.

We consider $(\vartheta - H_{dr})^{-1} \upharpoonright W \otimes L^2(\mathbb{R})$. Since H_{Idr} is a Kato potential with respect to H_{0d} the following von Neumann resolvent expansion converges in norm:

$$(\vartheta - H_{dr})^{-1} = (\vartheta - H_{0d})^{-1} \left(\sum_{n=0}^{\infty} [H_{Idr}(\vartheta - H_{0d})^{-1}]^n \right). \tag{25}$$

For Re ϑ sufficiently small, the second factor of (25) defines a bounded operator. According to Glimm and Jaffe[18] there exists an isomorphism between J and the square summable complex functions of the momentum lattice $\tilde{\Gamma}_d$ consisting of a finite number of lattice vectors. Hence the eigenvalue problem of $H_{0d} \upharpoonright W \otimes L^2(\mathbb{R})$ is completely solvable and the eigenvalues E_n can be numbered by natural numbes n, where $E_n \to \infty$ as $n \to \infty$. Therefore (see, e.g. Ref. 19), $(\vartheta - H_{0d})^{-1}$ is compact and, because of (25), $(\vartheta - H_{dr})^{-1} \upharpoonright W \otimes L^2(\mathbb{R})$ is compact for all $\vartheta \notin \mathrm{spec}(H_{dr} \upharpoonright W \otimes L^2(\mathbb{R}))$.

In a second step, we estimate $\inf \mathrm{spec}(H_{dr}(\mathbf{Q}) \upharpoonright W^\perp \otimes L^2(\mathbb{R}))$. Let us consider a vector $|\chi\rangle \in W^\perp \otimes L^2(\mathbb{R})$, which has the form $|\chi\rangle = |\theta\rangle \circledS |\varphi\rangle$, where $|\varphi\rangle \in W \otimes L^2(\mathbb{R})$, $|\theta\rangle \in W_\perp$, $|\theta\rangle \equiv |\theta(\tilde{k}^N)\rangle$, $\tilde{k}^N = (\mathbf{k}^1,\ldots,\mathbf{k}_N)$. Let $|\theta\rangle$ be an eigenstate of the phonon number operator with eigenvalue $N \geqslant 1$ and let $|\theta\rangle$ fulfill

$$\mathbf{P}|_d |\theta\rangle = \sum_{j=1}^{N} \mathbf{k}^j|_d |\theta\rangle, \quad \text{where} \quad \mathbf{k}^j|_d \in \Gamma_d \setminus \tilde{\Gamma}_d.$$

Then, $\|\chi\| = \|\theta\| \cdot \|\varphi\|$ and

$$\langle \chi | H_{Idr} | \chi \rangle = \langle \varphi | H_{Idr} | \varphi \rangle \langle \theta | \theta \rangle$$

and therefore

$$\langle \chi | H_{dr}(\mathbf{Q}) | \chi \rangle = \sum_{j=1}^{N} \langle \chi | \chi \rangle + \langle \varphi | H_{dr}\left((\mathbf{Q} - \sum_{j=1}^{N} (0,k_2^j|_d, k_3^j|_d) \right) | \varphi \rangle \langle \theta | \theta \rangle$$
$$\geqslant \left(1 + E\left(d,r,\mathbf{Q} - \sum_{j=1}^{N} (0,k_2^j|_d, k_3^j|_d) \right) \right) \langle \chi | \chi \rangle \geqslant \left(1 + \inf_{\mathbf{k}} E(d,r,\mathbf{Q} - (0,k_2,k_3)) \right) \langle \chi | \chi \rangle.$$

The same inequalities are valid for vectors which are finite linear combinations of pairwise orthogonal vectors of the form $|\theta(\tilde{k}^N)\rangle \circledS |\varphi\rangle$, $N \in \mathbb{N}$. Since these vectors are dense in $W^\perp \otimes L^2(\mathbb{R})$ we conclude

$$\inf \mathrm{spec}(H_{dr}(\mathbf{Q}) \upharpoonright W^\perp \otimes L^2(\mathbb{R}))$$
$$> 1 + \inf_{\mathbf{k}} E(d,r,\mathbf{Q} - (0,k_2,k_3)). \tag{26}$$

Third, let f be a positive C^∞ function on $\mathbb{R}$ such that $f(0) = 1, f(x) = 0$ if $x \geqslant \Delta(r,d,\mathbf{Q}) > 0$. Then we know from (26),

$$f(H_{dr}(\mathbf{Q}) - E(d,r,\mathbf{Q})) \upharpoonright W^\perp \otimes L^2(\mathbb{R}) \equiv 0.$$

On the other hand, the compactness of $(\vartheta - H_{dr})^{-1} \upharpoonright W \otimes L^2(\mathbb{R})$ implies that $f(H_{dr}(\mathbf{Q}) - E(d,r,\mathbf{Q})) \upharpoonright W \otimes L^2(\mathbb{R})$ is compact.

Since $F_d \otimes L^2(\mathbb{R}) = (W \otimes L^2(\mathbb{R})) \oplus (W^\perp \otimes L^2(\mathbb{R}))$, it follows that $f(H_{dr}(\mathbf{Q}) - E(d,r,\mathbf{Q})) \upharpoonright F_d \otimes L^2(\mathbb{R})$ is compact. This immediately implies Lemma 2.1. □

Now we can proceed along similar lines as Fröhlich does (see Theorem 2.3 in Ref. 3). The only difference is that our Hilbert space is $F_d \otimes L^2(\mathbb{R})$ (instead of F_d) and that in our case *a priori* $f(H_{dr}(\mathbf{Q}) - E(d,r,\mathbf{Q})) \upharpoonright F_d \otimes L^2(\mathbb{R})$ is compact, whereas in Ref. 3 the *total* spectrum is discrete. Nevertheless, Fröhlich's proof can directly be mimicked. As a consequence, we arrive at the following theorem, where the d cutoff is removed and a phonon gap in the spectrum is established.

Therorem 2.2: Suppose that the momentum $\mathbf{Q}$ is such that

$$\inf_{\mathbf{k}} E(r,\mathbf{Q} - (0,k_2,k_3)) + 1 - E(r,\mathbf{Q}) \equiv \Delta(r,\mathbf{Q}) > 0. \tag{27}$$

Then the interval $[E(r,\mathbf{Q}), E(r,\mathbf{Q}) + \Delta(r,\mathbf{Q})[$ belongs to the discrete part of spec $(H_r(\mathbf{Q}) \upharpoonright F \otimes L^2(\mathbb{R}))$.

III. REMOVING THE UV CUTOFF

To remove the UV cutoff, we use a canonical transformation, which was proposed by Gross[20] and mathematically studied by Nelson.[16] We define

$$H_r^T(\mathbf{Q}) = e^T H_r(\mathbf{Q}) e^{-T}, \tag{28}$$

where

$$T \equiv T_{r\Lambda} = \int d^3k (\beta_{r\Lambda}(\mathbf{k}) a_{\mathbf{k}} \exp(ik_1 x_1) - \mathrm{H.c.}) \tag{29}$$

and

$$\beta_{r\Lambda}(\mathbf{k}) \equiv \beta(\mathbf{k}) = -\alpha^{1/2} 2 g_r(\mathbf{k}) \theta(k - \Lambda)/(2 + k^2),$$
$$1 < \Lambda < r. \tag{30}$$

A lengthy but straightforward calculation, similar to those in Refs. 16 and 20, yields

$$H_r{}^T(\mathbf{Q}) = H_0(\mathbf{Q}) + \alpha^{1/2}\int d^3k\,(g_\Lambda(\mathbf{k})e^{ik_1x_1}a_\mathbf{k} + \text{H.c.}) + (\boldsymbol{\Phi}+\boldsymbol{\Phi}^*)^2/2 - \mathbf{G}\cdot\boldsymbol{\Phi} - \boldsymbol{\Phi}^*\cdot\mathbf{G} + \Sigma, \quad (31)$$

where we have used the abbreviation

$$\boldsymbol{\Phi} = \int d^3k\,\mathbf{k}\,\beta_{r\Lambda}(\mathbf{k})\exp(ik_1x_1)a_\mathbf{k}. \quad (32)$$

Here Σ is a finite self-energy, given by

$$\Sigma = \int d^3k\,[|\beta(\mathbf{k})|^2 + \alpha^{1/2}g^*(\mathbf{k})\beta(\mathbf{k}) + \alpha^{1/2}g(\mathbf{k})\beta^*(\mathbf{k})]. \quad (33)$$

We have to estimate each term in (31). As an example, we discuss the term involving the magnetic field. For all $|\psi\rangle\in D(H_0^{1/2})$ we have

$$|\langle\psi|(\mathbf{G}\cdot\boldsymbol{\Phi} + \boldsymbol{\Phi}^+\cdot\mathbf{G}|\psi\rangle| \leqslant 2\sum_{i=1}^{3}\|G_i\psi\|\cdot\|\Phi_i\psi\| \leqslant \|H_0^{1/2}\psi\|\cdot C(\Lambda)\cdot\|H_{0ph}^{1/2}\psi\| \leqslant C(\Lambda)\|H_0^{1/2}\psi\|^2,$$

where $C(\Lambda)\to 0$ as $\Lambda\to\infty$.

Hence

$$|\mathbf{G}\cdot\boldsymbol{\Phi} + \boldsymbol{\Phi}^+\cdot\mathbf{G}| \leqslant C(\Lambda)H_0. \quad (34)$$

Estimating the remaining terms in analogy to Ref. 3 (see Sec. 2.2 in Ref. 3), it follows that for all $\epsilon>0$ there exists a $\Lambda<\infty$ such that

$$|H_\Gamma^T(\mathbf{Q}) - H_0| \leqslant \epsilon H_0 + b(\Lambda), \quad (35)$$

where $b(\Lambda)$ is uniform in $r<\infty$ and Q. Mimicking Fröhlich's proof (see Theorem 2.4 in Ref. 3) we get the following theorem.

Theorem 3.1: Let Λ be fixed and $r\to\infty$.

$$\operatorname*{norm\text{-}lim}_{r\to\infty}(\vartheta - H_r{}^T(\mathbf{Q}))^{-1} = (\vartheta - H^T(\mathbf{Q}))^{-1}$$

exists, where $(\vartheta - H^T(\mathbf{Q}))$ is the resolvent of a unique s.a. operator $H^T(\mathbf{Q})$ bounded from below. Here $H^T(\mathbf{Q})$ can be related to the Hermitian forms induced by (35) by a variant of Friedrich's extension theorem (see Nelson[16]).

$$\operatorname*{s\text{-}lim}_{r\to\infty}\exp(T_{r\Lambda}) \equiv \exp(T_{\infty\Lambda})$$

exists. Therefore

$$H(\mathbf{Q}) \equiv \exp(-T_{\infty\Lambda})H^T(\mathbf{Q})\exp(T_{\infty\Lambda})$$

is self-adjoint and bounded below. □

Again, we follow Fröhlich (Theorem 2.7 in Ref. 3) and obtain that Theorem 2.2 is even valid in the limit $r\to\infty$, i.e., the following lemma.

Lemma 3.2: Let $E(\mathbf{Q})$ denote the ground-state energy of $H(\mathbf{Q})$. Suppose that the momentum $\mathbf{Q}$ is such that

$$\inf_{\mathbf{k}}(E(\mathbf{Q} - (0,k_2,k_3)) + 1 - E(\mathbf{Q})) \equiv \Delta(\mathbf{Q}) > 0. \quad (36)$$

Then the interval $[E(\mathbf{Q}), E(\mathbf{Q}) + \Delta(\mathbf{Q})[$ belongs to the discrete part of spec $H(\mathbf{Q})$.

IV. NONDEGENERACY OF THE MOMENTUM DECOMPOSED GROUND STATE

Keeping in mind that we intend to apply a generalized version of the Perron–Frobenius theorem, it is useful to consider a slightly different cutoff Hamiltonian $H_n'(\mathbf{Q})$. (Of course, the cutoff is removed later.) In doing so, we now replace in (6c) and (9a) the coupling $g(\mathbf{k})$ by

$$g_n(\mathbf{k}) \equiv -g(\mathbf{k})\mu_n(k^\perp) \equiv -g(\mathbf{k})(\theta(n-k^\perp) + \theta(k^\perp - n)\exp(-(k^\perp - n))), \quad (37)$$

where

$$k^{\perp 2} = k_2^2 + k_3^2. \quad (38)$$

Note that

$$g_n(\mathbf{k}) \in L^2(\mathbb{R}^3), \quad \text{for} \quad n<\infty. \quad (39)$$

Additionally, for the first component of the phonon Fock space we now choose the q representation (Schrödinger representation) instead of the momentum representation. In this new representation the Hamiltonians read

$$H_n'(\mathbf{Q}) = H_0' + H_{IN}', \quad H_0' = \int d^3l\,b_l{}^+b_l + \mathbf{G}'^2/2, \quad (40)$$

$$H_{IN}' = \left(\frac{2\alpha}{\pi}\right)^{1/2}\Omega\int d^3l\,\mu_n(k^\perp) \times K_0(k^\perp(x_1-q))(b_l + b_l{}^+), \quad (41)$$

where now

$$\mathbf{G}' = (p_1, Bx_1 + Q_2 - P_2', Q_3 - P_3'),$$

$$b_l = (2\pi)^{-1/2}\int dk_1 a_\mathbf{k}\exp(ik_1q), \quad (42)$$

and

$$P_i' = \int d^3l\,k_i b_l{}^+b_l, \quad i=2,3, \quad \mathbf{l}=(q,k_2,k_3), \quad (43)$$

and where $K_0(x)$ denotes a strictly positive Bessel function of imaginary argument. By a canonical transformation, analogous to that in Sec. III, one proves for all $|\varphi\rangle, |\Phi\rangle\in D(H_0')$ and for $\vartheta < \inf\text{ spec } H(\mathbf{Q})$,

$$\lim_{n\to\infty}\langle\varphi|(H_n'(\mathbf{Q}) - \vartheta)^{-1}|\Phi\rangle = \langle\varphi|(H'(\mathbf{Q}) - \xi)^{-1}|\Phi\rangle, \quad (44)$$

where $H'(\mathbf{Q})$ is s.a., bounded below, and has the same spectrum as $H(\mathbf{Q})$. Because of (39) the following expansion converges in norm:

$$(H_n'(\mathbf{Q}) - \vartheta)^{-1} = (H_0' - \vartheta)^{-1}\sum_{m=0}^{\infty}[(-1)H_{In}'(H_0' - \vartheta)^{-1}]^m. \quad (45)$$

Computing the kernel

$$\langle y|\otimes\langle 0|b_{l_1'}\cdots b_{l_{u'}'}|(H_n'(\mathbf{Q}) - \vartheta)^{-1}|b_{l_1}^+\cdots b_{l_u}^+|0\rangle\otimes|x\rangle, \quad u,u'\in IN_0, \quad (46)$$

we see by inspection of (45) that (46) is strictly positive for $\alpha>0$: With respect to the electron coordinate in the Schrö-

dinger representation (i.e., with fixed positive phonon wave function), $(H'_0 - \vartheta)^{-1}$ is positivity improving as resolvent of a harmonic oscillator. With respect to the phonon coordinate, $(H'_0 - \vartheta)^{-1}$ is positivity preserving and it preserves the support. Furthermore, H'_{In} is positivity preserving with respect to the electron coordinate because of the positivity of K_0. Moreover, for a suitable choice of m in (45) it can be achieved that

$$\langle y| \otimes \langle 0|b_{l'_1} \cdots b_{l'_{l'}}|(-H'_{In}(H'_0 - \vartheta)^{-1})^m|b_{l_1}^+ \cdots b_{l_u}^+|0\rangle \otimes |x\rangle > 0. \quad (47)$$

Consequently, $(H'_n(\mathbf{Q}) - \vartheta)^{-1}$ is positivity improving in the chosen representation. Since (46) is monotonically increasing with n, we get with (44) that $(H' - \vartheta)^{-1}$ is positivity improving. From Sec. III we know that under condition (36) $\inf \operatorname{spec} H'(\mathbf{Q}) = E(\mathbf{Q})$ is an eigenvalue of $H'(\mathbf{Q})$. Therefore (see, e.g., Ref. 19) $E(\mathbf{Q})$ is a simple eigenvalue, or, equivalently, the ground state is nondegenerate.

V. BOUNDS ON THE GROUND-STATE ENERGY

Lemma 5.1: For the ground-state energy $E(\mathbf{Q})$ we have the bounds

$$\text{(i) } E(\mathbf{Q}) \leqslant E(\mathbf{0}) + Q_3^2/2, \quad (48)$$

$$\text{(ii) } E(\mathbf{Q}) \geqslant E(\mathbf{0}). \quad (49)$$

Proof: (i) follows from the fact that $E(\mathbf{Q}) - Q_3^2/2$ is a concave symmetrical function of Q_3, since the Hamiltonian $H(\mathbf{Q}) - Q_3^2/2$ couples linear to Q_3. (ii) We use the same procedure as in (37) and (40)–(43) transforming now the third component of the phonon coordinate into the Schrödinger representation, i.e.,

$$k^{12} = k_1^2 + k_2^2, \quad \mathbf{l} = (k_1, k_2, q).$$

Thereby we obtain new Hamiltonians $\tilde{\tilde{H}}_n(\mathbf{Q})$, $\tilde{\tilde{H}}_{0ph}$, $\tilde{\tilde{\mathbf{G}}}(\mathbf{Q})^2/2$, $\tilde{\tilde{H}}_{In}$, and the phonon momentum operator $\tilde{\tilde{\mathbf{P}}}$. Because of Theorem 3.1, it suffices to show (49) for $n < \infty$. One easily sees (e.g., by a Dyson expansion) that

$$\exp(-tL) \equiv \exp(-t(\tilde{\tilde{H}}_{0ph} + \tilde{\tilde{H}}_{In} + \tilde{\tilde{G}}_1^2/2 + \tilde{\tilde{G}}_2^2/2)) \quad (50)$$

is positivity preserving for $t > 0$. Now, we follow an idea of Gross[21] and write

$$\exp\left(-\frac{t(Q_3 - \tilde{\tilde{P}}_3)^2}{2}\right) = (2\pi t)^{-1/2} \int dy \exp\left(-\frac{y^2}{2t}\right) \exp(iy(Q_3 - \tilde{\tilde{P}}_3)). \quad (51)$$

Hence, we have, since $\exp(-iy\tilde{\tilde{P}}_3)$ preserves positivity,

$$\begin{aligned} &|\exp(-t\tilde{\tilde{G}}_3(\mathbf{Q})^2/2)\exp(-tL)|\varphi\rangle| \\ &\leqslant (2\pi t)^{-1/2} \int dy \exp\left(-\frac{y^2}{2t}\right) \\ &\quad \times \exp(iy\tilde{\tilde{P}}_3)\exp(-tL)\|\varphi|\rangle \\ &\leqslant \exp(-t\tilde{\tilde{G}}_3(\mathbf{0})^2/2)\exp(-tL)\|\varphi|\rangle, \\ &\quad \text{for } |\varphi\rangle \in D(\tilde{\tilde{H}}_n). \end{aligned} \quad (52)$$

One proceeds by induction to obtain

$$\begin{aligned} &|(\exp(-t\tilde{\tilde{G}}_3(\mathbf{Q})^2/(2k))\exp(-tL/k))^k|\varphi\rangle| \\ &\leqslant (\exp(-t\tilde{\tilde{G}}_3(\mathbf{0})^2/(2k))\exp(-tL/k))^k\|\varphi|\rangle, \end{aligned} \quad (53)$$

which—because of the Trotter product formula—implies finally

$$\|\exp(-t\tilde{\tilde{H}}_n(\mathbf{Q}))|\varphi\rangle\| \leqslant \|\exp(-t\tilde{\tilde{H}}_n(\mathbf{0}))\| \cdot \|\varphi\|. \quad (54)$$

Hence (49) is established.

From Lemma 5.1 it follows immediately that (36) is fulfilled, if

$$Q_3^2 < 2. \quad (55)$$

We note that we can prove, using a new functional integral method developed by Gerlach *et al.*,[22] that the continuum edge of $H(\mathbf{Q})$ begins exactly at the point $E(\mathbf{0}) + 1$ involving scattering states with one real phonon (see Devreese[23] for a review). This yields the bound

$$E(\mathbf{Q}) \leqslant E(\mathbf{0}) + 1. \quad (56)$$

VI. ANALYTICAL PERTURBATION THEORY

To establish analytical properties in $\mathbf{Q}$, α, and B, we start from the canonically transformed Hamiltonian $H_r^T(\mathbf{Q}) \equiv H_r^T(Q, \sqrt{\alpha}, B)$, see (28). Let now $\mathbf{Q}, \alpha, B$ be fixed, where $Q_3 < 2$, $\alpha \geqslant 0$, $B > 0$. We consider small deviations around these fixed parameters. The Q_2 dependence is trivial. Concerning the Q_3 dependence we have

$$\begin{aligned} &H_r^T((0, Q_2, Q_3 + \kappa), \sqrt{\alpha}, B) \\ &\quad = H_r^T(\mathbf{Q}, \sqrt{\alpha}, B) + \kappa(Q_3 - P_3 - \Phi_3 - \Phi_3^*) + \kappa^2/2. \end{aligned} \quad (57)$$

It is easily seen that P_3 is form bounded:

$$|P_3| \leqslant aH_r^T(\mathbf{Q}, \sqrt{\alpha}, B) + b, \quad \text{for} \quad r \leqslant \infty. \quad (58)$$

Therefore, the associated operators $H^T((0, Q_2, Q_3 + \kappa), \sqrt{\alpha}, B)$ are a holomorphic family of s.a. operators of type (B) in κ in the sense of Kato.[17] The α dependence can be treated in a similar way:

$$H_r^T(\mathbf{Q}, \sqrt{\alpha} + \gamma, B) = H_r^T(\mathbf{Q}\sqrt{\alpha}, B) + \gamma H''_r. \quad (59)$$

The estimations in Ref. 3 used in Sec. III show that H''_r is form bounded with constants independent of r ($r \leqslant \infty$). Hence $H^T(\mathbf{Q}, \sqrt{\alpha} + \gamma, B)$ forms a holomorphic family of type (B) in the sense of Kato in γ, too.

The dependence on the magnetic field strength B is more difficult. We use a scaling transformation

$$\tilde{x} \equiv B^{1/2} x_1, \quad \tilde{p} \equiv B^{-1/2} p_1, \quad (60)$$

$$\tilde{a}_{\mathbf{k}} \equiv B^{3/4} a_{B^{1/2}\mathbf{k}}, \quad \tilde{a}_{\mathbf{k}}^+ \equiv B^{3/4} a_{B^{1/2}\mathbf{k}}^+. \quad (61)$$

Written in these new operators the resulting Hamiltonian $\tilde{H}_r^T(\mathbf{Q})$ has the form

$$\begin{aligned} \tilde{H}_r^T(\mathbf{Q}) = &\frac{B\tilde{G}^2}{2} + \int d^3k\, \tilde{a}_{\mathbf{k}}^+ \tilde{a}_{\mathbf{k}} + B^{1/2}\alpha^{1/2} \\ &\times \int d^3k\, g_{\tilde{\Lambda}}(\mathbf{k})(\exp(ik_1\tilde{x})\tilde{a}_{\mathbf{k}} + \text{H.c.}) \\ &+ B^{3/2}(\tilde{\mathbf{\Phi}} + \tilde{\mathbf{\Phi}}^+)^2/2 \\ &- B^{5/4}(\tilde{\mathbf{G}} \cdot \tilde{\mathbf{\Phi}} + \tilde{\mathbf{\Phi}}^+ \cdot \tilde{\mathbf{G}}) + \Sigma, \end{aligned} \quad (62)$$

where $\tilde{\mathbf{G}}$ and $\tilde{\Phi}$ are given as in (8), (32) replacing the old operators with the new ones and the old quantities $\beta(\mathbf{k})$, Λ, r, $\mathbf{Q}$ with

$$\tilde{\beta}(\mathbf{k}) = -2\alpha^{1/2} g_{\tilde{r}}(\mathbf{k})\theta(k-\tilde{\Lambda})/(2+Bk^2), \quad (63)$$

$$\tilde{\Lambda} = B^{-1/2}\Lambda, \quad \tilde{r} = B^{-1/2} r, \quad \tilde{\mathbf{Q}} = B^{-1/2}\mathbf{Q}. \quad (64)$$

Note that the operators $\tilde{x}, \tilde{p}, \tilde{a}_{\mathbf{k}}, \tilde{a}_{\mathbf{k}}^{+}$ fulfill the same commutator relations as $x_1, p_1, a_{\mathbf{k}}, a_{\mathbf{k}}^{+}$. Therefore, $\tilde{H}_r^{\,T}(\mathbf{Q})$ has the same spectrum as $H_r^{\,T}(\mathbf{Q})$. More properly, the Hamiltonian $\tilde{H}_r^{\,T}(\mathbf{Q})$ is obtained from $H_r^{\,T}(\mathbf{Q})$ by a canonical transformation, what is easily seen using Wigner's theorem (see Bargmann[24]). Now, the B dependence manifests itself mainly as simple prefactors before the single parts of the Hamiltonian $\tilde{H}_r^{\,T}(\mathbf{Q})$. It is easily seen by developing (63) in its power series that

$$\tilde{H}_r^{\,T}(\mathbf{Q},\sqrt{\alpha},B+\epsilon) = \tilde{H}_r^{\,T}(\mathbf{Q},\sqrt{\alpha},B) + \sum_{n=1}^{\infty} \epsilon^n G_r^{(n)}(\mathbf{Q},\sqrt{\alpha},B), \quad |\epsilon| < B, \quad (65)$$

where in the sense of forms

$$|G_r^{(n)}(\mathbf{Q},\sqrt{\alpha},B)| \leqslant c'^n (a'\tilde{H}_r^{\,T}(\mathbf{Q},\sqrt{\alpha},B) + b'), \quad \text{for} \quad r \leqslant \infty. \quad (66)$$

Therefore we can repeat our statement that we are dealing with a holomorphic family of type (B) in ϵ. It now follows from standard perturbation theory (see Kato[17]) and from the fact that $E(\mathbf{Q})$ is an isolated simple eigenvalue that the ground-state energy $E(\mathbf{Q})$ is jointly real analytic in α,[25] B, and $\mathbf{Q}$ for $\alpha \geqslant 0$, $B > 0$, $Q_3^2 < 2$. The same holds true for the energies of the discrete excited states lying in the spectral interval $[E(\mathbf{0}), E(\mathbf{0}) + 1[$, where we have to exclude possible degenerate points.[26] Furthermore, the associated wave functions are analytic in α, B, and $\mathbf{Q}$. This, in turn, has immediate consequences on expectation values of operators which are independent of α, B, $\mathbf{Q}$ like the number of virtual phonons or the polaron radius, etc. (see Peeters and Devreese[9]). Again, all these quantities are analytic in α, B, and $\mathbf{Q}$. From Lemma 5.1 we know that the ground-state energy $E_F(\alpha,B)$ of the original Hamiltonian H_F is obtained by taking $E(\mathbf{0})$. Especially, $E_F(\alpha,B)$ is analytic in α and B.

Another quantity, which is of interest, is the magnetic polaron mass. Peeters and Devreese[7] have defined parallel and perpendicular magnetic polaron masses in the anisotropic Feynman approximation. One way to define a parallel magnetic polaron mass m'' *a priori* without using any approximation is

$$\frac{1}{m''} \equiv \frac{\partial^2 E(\mathbf{Q})}{\partial Q_3^2}\bigg|_{Q=0}. \quad (67)$$

Another possibility to define a cyclotron mass m^* (depending on α and B) at weak or intermediate magnetic fields is

$$E_1(\mathbf{0}) - E(\mathbf{0}) = B/(2m^*), \quad (68)$$

where $E_1(\mathbf{Q})$ is the energy of the first excited state, i.e., the second Landau level. It follows immediately that both masses m'' and m^* are analytic in α and B.

VII. EXTENSION TO OTHER PROBLEMS

First, the dispersions and the coupling can be generalized to

$$\omega(\mathbf{k}) = \omega(k) \geqslant \omega_0 > 0, \quad (69)$$

$$g(\mathbf{k}) = g(k), \quad \int d^3k \frac{|g(k)|^2}{c+k^2} < \infty, \quad c > 0. \quad (70)$$

Additionally, performing the scaling transformation (60), (61), we have to assume that $g(k)$ is representable as a finite or infinite linear combination of powers $k^p (p \in \mathbb{R})$ in a domain of $\mathbb{R}^3$. Then, the same proof goes through with two exceptions: The uniqueness proof of the ground state and Lemma 5.1 (ii) have to be modified. The condition $\omega_0 > 0$ cannot be weakened with our methods, since the gap in the spectrum, which makes perturbation theory possible, does not remain. For more singular couplings $g(\mathbf{k})$ one has to renormalize in a well-known way (see Nelson[16]).

Furthermore, the space dimension d is not relevant for our proof, if we take

$$|g(\mathbf{k})|^2 \sim k^{1-d}. \quad (71)$$

Several branches of optical phonons can easily be included in the proof. Whether or not we consider a discrete $\mathbf{k}$ summation or a $\mathbf{k}$ integration has no influence on the phase transition problem. The methods worked out in Secs. II and III are applicable, if the unperturbed Hamiltonian with discrete, cutoff $\mathbf{k}$ sums has a compact resolvent, where conserved components of the total momentum are replaced by C numbers. For example, the problem of a polaron in an external potential $V(\mathbf{x})$, where $V(\mathbf{x}) \to \infty$ as $|\mathbf{x}| \to \infty$ is tractable. Another example concerns the polaron in an external uniform electric field. Since the resulting Hamiltonian is unbounded, it has to be renormalized. We cut off the potential as follows:

$$V(\mathbf{x}) = \begin{cases} |e|Ex_1, & \text{for } x_1 \geqslant L, \\ \infty, & \text{for } x_1 < L. \end{cases} \quad (72)$$

Then, all results concerning the ground-state energy, etc., hold. Especially, the ground-state energy is analytic in the coupling parameter α and the electric field strength $E (E > 0)$.

ACKNOWLEDGMENTS

I am most grateful to Professor H. Spohn for valuable discussions, which have initiated this work, and for helpful information concerning Secs. IV and V. Furthermore, I thank Professor B. Gerlach for numerous clarifying discussions and Professor K. Veselic and Professor W. Hazod for explanation of some mathematical details. Moreover, I am indebted to Professor J. Fröhlich for sending me a detailed manuscript of his methods.

Financial support for this part of my thesis by the Studienstiftung des deutschen Volkes is gratefully acknowledged, too.

[1] H. Fröhlich, H. Pelzer, and S. Zienau, Philos. Mag. **41**, 221 (1950).
[2] H. Spohn, Ann. Phys. (NY) **175**, 278 (1987).
[3] J. Fröhlich, Fortschr. Phys. **22**, 159 (1974).
[4] R. F. Alvarez-Estrada, Ann. Inst. H. Poincaré A **31**, 141 (1979).
[5] B. Gerlach and H. Löwen, Phys. Rev. B **35**, 4291, 4297 (1987).

[6]F. M. Peeters and J. T. Devreese, Solid State Commun. **39**, 445 (1981); Phys. Status Solidi B **110**, 631 (1982).
[7]F. M. Peeters and J. T. Devreese, Phys. Rev. B **25**, 7281, 7302 (1982).
[8]F. M. Peeters and J. T. Devreese, Phys. Status Solidi B **115**, 285 (1983); Physica B **117**, 567 (1983).
[9]F. M. Peeters and J. T. Devreese, Phys. Rev. B **31**, 4890 (1985).
[10]F. M. Peeters and J. T. Devreese, *Lecture Notes in Physics*, Vol. 177 (Springer, Berlin, 1983), p. 406.
[11]S. N. Gorshkov, A. V. Zabrodin, C. Rodriguez, and V. K. Fedyanin, Teor. Mat. Fiz. **62**, 304 (1985) [Theor. Math. Phys. **62**, 205 (1985)].
[12]Wu Xiaoguang, F. M. Peeters, and J. T. Devreese, Phys. Rev. B **32**, 7964 (1985).
[13]Y. Lépine and D. Matz, Can. J. Phys. **54**, 1979 (1976).
[14]Y. Lépine, J. Phys. C **18**, 1817 (1985).
[15]J. T. Devreese, in *Theoretical Aspects and New Developments in Magneto Optics*, edited by J. T. Devreese (Plenum, New York, 1980), p. 217.
[16]E. Nelson, J. Math. Phys. **5**, 1190 (1964).
[17]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).
[18]J. Glimm and A. Jaffe, Ann. Math. **91**, 362 (1970).
[19]M. Reed and B. Simon, *Analysis of Operators* (Academic, New York, 1978).
[20]E. P. Gross, Ann. Phys. (NY) **19**, 219 (1962).
[21]L. Gross, J. Funct. Anal. **10**, 52 (1972).
[22]B. Gerlach, H. Löwen, and H. Schliffke, Phys. Rev. B **36**, 6320 (1987).
[23]J. T. Devreese, in *Polarons in Ionic Crystals and Polar Semiconductors*, edited by J. T. Devreese (North-Holland, Amsterdam, 1972), pp. 83–159.
[24]V. Bargmann, J. Math. Phys. **5**, 862 (1964).
[25]We remark that analytical properties in $\sqrt{\alpha}$ are simultaneously analytical properties in α, since, around $\alpha = 0$, the perturbation series has only powers α^n.
[26]Degeneration of the Landau levels, however, is not expected for $\alpha > 0$ because of the Wigner–von Neumann noncrossing theorem.

Analytical behavior of the ground-state energy and pinning transitions for a bound polaron

H. Löwen
Institut für Physik der Universität Dortmund, D-4600 Dortmund 50, West Germany

(Received 6 August 1987; accepted for publication 17 February 1988)

Spectral properties of a three-dimensional optical polaron, bound in an external potential, are studied. If the associated one-particle Hamiltonian has a bound state, it is proved that the ground-state energy of the polaron and expectation values of the ground state are analytical functions of the coupling parameter α and the potential strength β. Especially in the case of a Coulomb potential, all changes in the polaron state are continuous, disproving claims of several variational calculations. If, on the other hand, the one-particle Hamiltonian has no bound state, the existence of a pinning transition is shown for the polaron. As physically relevant potentials for the pinning transition, a spherical square well and a screened Coulomb potential are considered. Their phase diagrams are given in the effective-mass approximation.

I. INTRODUCTION

An electron, bound to a defect in polar semiconductors (such as an impurity or a vacancy) and interacting with the longitudinal optical phonons, is called a bound polaron. Since the earlier papers of Buimistrov and Pekar[1] and Platzman[2] and Larsen,[3] this important problem has received considerable attention, as recent publications show (see, e.g., Adamowski[4] and the references therein, as well as Mason and Das Sarma[5] and Degani and Hipolito[6]). The present paper is concerned with analytical and spectral properties of a polaron in a generalized external potential $V(\mathbf{r})$.

The bound (three-dimensional) polaron is described by the well-known Fröhlich Hamiltonian,[7] which reads as follows:

$$H_F(\alpha,\beta) = H_{0ph} + \mathbf{p}^2/2 - \beta V(\mathbf{r}) + \alpha^{1/2} H_{IF}, \tag{1}$$

where

$$H_{0ph} = \int d^3k\, \omega(\mathbf{k}) a^+(\mathbf{k}) a(\mathbf{k}) \tag{2}$$

and

$$H_{IF} = \int d^3k\, [g(\mathbf{k}) a(\mathbf{k}) \exp(i\,\mathbf{kr}) + \text{H.c.}]. \tag{3}$$

Here, $\mathbf{r}$ and $\mathbf{p}$ are the position and momentum operator of the electron, respectively, and $\mathbf{k}$, $\omega(\mathbf{k})$, $a^+(\mathbf{k})$, and $a(\mathbf{k})$ are the wave vector, frequency, creation, and annihilation operators of the phonons, respectively (i.e., a scalar Bose field); $g(\mathbf{k})$ denotes the electron–phonon coupling, α being the coupling parameter. Setting $m = \hbar = 1$, we keep α and β as the only parameters $(\alpha,\beta \geqslant 0)$.

Henceforth, the following conditions (4)–(6) on $\omega(\mathbf{k})$ and $g(\mathbf{k})$ are assumed:

$$\inf_{\mathbf{k}} \omega(\mathbf{k}) \equiv \omega_0 > 0, \tag{4}$$

$\omega(\mathbf{k})$ being a continuous function of $\mathbf{k}$. Thus (4) implies that we are dealing with optical phonons. Furthermore we assume

$$\int d^3k \frac{|g(\mathbf{k})|^2}{1+k^2} < \infty \tag{5}$$

and reflection symmetry

$$\omega(-\mathbf{k}) = \omega(\mathbf{k}), \quad g(-\mathbf{k}) = g(\mathbf{k}). \tag{6}$$

For more UV-singular couplings $g(\mathbf{k})$ one has to renormalize by the scheme proposed by Nelson.[8] For the potential $V(\mathbf{r})$ we treat the case

$$V \in R + L_\epsilon^\infty(\mathbb{R}^3), \tag{7}$$

where R is the Rollnik class

$$V \in R \Leftrightarrow \int d^3x\, d^3y \frac{|V(\mathbf{x})| \cdot |V(\mathbf{y})|}{|\mathbf{x}-\mathbf{y}|^2} < \infty. \tag{8}$$

Statement (7) means that for any positive ϵ, V is representable as $V = f_\epsilon + g_\epsilon$, where $f_\epsilon \in R$ and $|g_\epsilon|$ is bounded by ϵ. Additionally, we assume that the associated one-particle Hamiltonian

$$H_e(\beta) = p^2/2 - \beta V(\mathbf{r}) \tag{9}$$

is essentially self-adjoint and bounded from below and has at least one bound state with strictly negative energy $E_0(\beta)$. The last assumption is abandoned in Sec. IV.

Physically most relevant cases are $\omega(\mathbf{k}) \equiv \omega_0 > 0$ (i.e., dispersionless optical phonons) and $g(\mathbf{k}) \sim 1/k$ for polar scattering or $g(\mathbf{k}) \sim \Theta(k_0 - k)$, $k_0 > 0$, for deformation potential scattering. Possible choices for the external potential $V(\mathbf{r})$ are a Coulomb potential $[V(\mathbf{r}) = 1/r]$, a screened Coulomb potential $[V(\mathbf{r}) = \exp(-k_0 r)/r, \quad k_0 > 0]$, or a spherical square well $[V(\mathbf{r}) = \Theta(a-r), \quad a > 0]$. The case of anisotropic bound polarons is included in our general assumptions (4)–(6), too.

The analytical properties of the ground-state energy of an optical polaron, subject to an external potential $V(\mathbf{r})$, are known only for a small class of potentials: In the case of free optical polarons ($V \equiv 0$) Spohn[9] recently proved the analyticity of the ground-state energy using the functional analytical work of Fröhlich[10] whereas Gerlach and the author[11,12] showed that the (formal) free energy is analytic in α and in the temperature T for $0 < T < \infty$. In Ref. 13, the methods of Fröhlich are generalized to an optical polaron, exposed to a homogeneous magnetic field or to an external potential $V(\mathbf{r})$ with $\lim_{r\to\infty} V(\mathbf{r}) = \infty$, implying the analyticity of the

 0022-2488/88/061505-09$02.50

ground-state energy in α and β. Apparently, this is another class of potentials as the class of impurity potentials, defined by (7). We remark that a third interesting class of potentials, which is not treated in this paper, concerns periodic potentials. They are, however, well studied in connection with quantum Brownian motion (see, e.g., Fisher and Zwerger,[14] and references therein).

The aim of the present paper is twofold.

First, we prove that the ground-state energy as well as the mean number of virtual phonons and the polaron radius of a bound polaron are analytical functions of the coupling parameter and the potential strength. A crucial assumption for this result is that the associated one-particle Hamiltonian (9) has a bound state. For long-range potentials (e.g., for a Coulomb potential) such a bound ground state exists for any potential strength.

Consequently, for long-range potentials a discontinuous "phase transition" (i.e., a nonanalyticity of the ground-state energy) does not exist. From the beginning of the polaron story up to now, the question of whether or not an optical polaron, bound in a Coulomb potential, shows up a "phase transition," was controversially discussed in the literature. It was mainly studied with the help of variational calculations yielding an upper bound on the exact ground-state energy. The physical background of such a "phase transition" becomes clear in the work of Toyozawa.[15] He gets a transition from a shallow state, formed by the external potential, to a deep self-trapped state, caused by a lattice distortion. This process is called shallow–deep instability.

On the one hand, Larsen,[3,16] Tokuda, Shoji, and Yoneya,[17] and Tokuda[18] obtain a variational bound on the ground-state energy of a bound polaron that exhibits a nonanalyticity, whereas Matsuura[19] and Mason and Das Sarma,[5] on the other hand, emphasize that their results are smooth quantities. In view of our proof, we remark that the nonanalyticities quoted above are nothing more than artifacts of the approximations made, but not intrinsic properties of the Fröhlich Hamiltonian. Takegahara and Kasuya[20] describe different states of the bound polaron by different sections in the α-β plane. However, note that, in view of our result, the properties of a bound polaron cannot be described within a phase transition concept.

The situation becomes quite different for an attractive three-dimensional short-range potential. This is the second concern of the present paper. In this case, the potential strength must exceed a critical value, to generate a bound ground state of the one-particle Hamiltonian (9). This phenomenon is well understood in atomic physics; we refer, for instance, to Glaser *et al.*,[21] Reed and Simon,[22] and Simon.[23] As the potential strengh β increases, a bound ground state arises from the continuum edge; the ground state undergoes a localization transition, the associated ground state energy being nonanalytic in β.

It is an interesting task to study the influence of the phonon interaction on this transition. In the framework of a discrete model for an exciton, this was examined by Shinozuka and Toyozawa.[24] In the adiabatic approximation, they found a localization transition of the ground state, which they called impurity assisted self-trapping. It is connected with a nonanalyticity of the ground-state energy. Shinozuka and Toyozawa also give the phase diagram in the α-β plane [see Fig. 2(b) in Ref. 24] Note, however, that they get even a discontinuous transition for $\beta = 0$, which was shown[25] to be an artifact of their approximation.

It was Spohn[26] who described the polaron approximatively as a single particle with phonon-induced altered mass. In this so-called effective-mass approximation, it turns out that the localization transition persists for phonon coupling $\alpha > 0$. Its critical line in the α-β plane, however, does not intersect the α axis. Spohn called this transition pinning transition; we shall use this term, too. In this paper, we prove that the exact ground state undergoes a pinning transition for any $\alpha \geqslant 0$. Consequently, the effective mass approximation reflects the right qualitative behavior of the analyticity of the ground-state energy. Furthermore, we discuss the relationship of the exact critical line to the one obtained in the effective mass approximation.

The organization of the present paper is as follows: In Sec. II, we show that $H(\alpha,\beta)$ is a well-defined self-adjoint operator. If the one-particle Hamiltonian has a bound state, we prove that the ground-state energy belongs to the discrete part of the spectrum of $H(\alpha,\beta)$ and is nondegenerate. In doing so, we make use of functional analytical methods of Fröhlich,[10] which clarified spectral properties of the free optical polaron. Moreover, we determine the continuum edge of $H(\alpha,\beta)$ and show the stability of bound states under the influence of the phonon interaction. The consequences (like analyticity properties of the ground state), following from Sec. II, and extensions of our theory are pointed out in Sec. III. Section IV is devoted to a discussion on the pinning transition. Applying our methods of Sec. II, we prove the existence of a pinning transition and discuss further properties of the pinning transition and the effective-mass approximation. In particular, a spherical square well and a screened Coulomb potential are considered. In Sec. V, we conclude our results.

II. SPECTRAL PROPERTIES OF A BOUND POLARON

It will be profitable to transform the Fröhlich Hamiltonian (1) by a Lee–Low–Pines transformation. Defining the unitary operator

$$U = \exp(-i\,\mathbf{P}\mathbf{r}), \tag{10}$$

where

$$\mathbf{P} = \int d^3k\, \mathbf{k} a^+(\mathbf{k}) a(\mathbf{k}) \tag{11}$$

is the phonon momentum, we shall discuss hereafter the unitarily equivalent Hamiltonian

$$H(\alpha,\beta) \equiv U^{-1} H_F(\alpha,\beta) U = H_{0ph} + \tfrac{1}{2}(\mathbf{p}-\mathbf{P})^2 - \beta V(\mathbf{r}) + \alpha^{1/2} H_I. \tag{12}$$

In (12), H_I is given by

$$H_I = \int d^3k\, [g(\mathbf{k}) a(\mathbf{k}) + g^*(\mathbf{k}) a^+(\mathbf{k})]. \tag{13}$$

To begin with, we pose the Fröhlich Hamiltonian on a mathematically rigorous level. We first specify the underlying Hilbert space. It is taken to be $\mathscr{H} = F \otimes L^2(\mathbb{R}^3)$, where

$$F = \bigoplus_{m=0}^{\infty} (L^2(\mathbb{R}^3))^{\textcircled{s}m} \tag{14}$$

is the usual Fock space for the phonons, $\textcircled{s}$ denoting the symmetrical tensor product.

In (13), we replace the coupling $g(\mathbf{k})$ by $g(\mathbf{k})\Theta(\rho - k)$, where $\rho < \infty$ is a UV cutoff, which makes $g(\mathbf{k})$ square integrable and which is to be removed later. Thereby we obtain the Hamiltonian $H(\alpha,\beta,\rho)$ [resp. $H_I(\rho)$]. Following Nelson,[8] it is easily proved that $H_I(\rho)$ is a Kato potential with respect to H_{0ph} with relative bound zero. Since $H_{0ph} + (\mathbf{p} - \mathbf{P})^2/2 - \beta V(\mathbf{r})$ is essentially self-adjoint, the Kato–Rellich theorem assures us that $H(\alpha, \beta, \rho)$ is bounded from below and essentially self-adjoint, too.

We now construct a discrete momentum lattice Γ_d for the phonon momentum space $\mathbb{R}^3$ in analogy to Fröhlich[10] and Glimm and Jaffe[27]:

$$\Gamma_d = \{\mathbf{k}\in\mathbb{R}^3 | \mathbf{k}_j = n_j/\Lambda_d,\ n_j\in\mathbb{Z},\ \Lambda_d = 2^d\Lambda_0,\ \Lambda_0\in\mathbb{R}^+,\ j = 1,2,3\}. \tag{15}$$

To each $\mathbf{k}\in\mathbb{R}^3$ we associate a $\mathbf{k}|_d\in\Gamma_d$, namely

$$\mathbf{k}|_d = (n_1,n_2,n_3)/\Lambda_d, \quad n_j = [k_j\Lambda_d], \tag{16}$$

where

$$[a] \equiv \begin{cases} \text{largest integer} \leqslant a, & \text{if } a < 0, \\ \text{smallest integer} > a, & \text{if } a \geqslant 0. \end{cases}$$

Furthermore, we define a subspace $S_d \subseteq L^2(\mathbb{R}^3)$ of step functions:

$$F\in S_d \Leftrightarrow f(\mathbf{k}) = f(\mathbf{k}|_d). \tag{17}$$

For $g\in L^2(\mathbb{R}^3)$ let $g|_d$ denote the orthogonal projection of g onto S_d. This notation is readily generalized for locally integrable g. Then, let

$$F_d \equiv \bigotimes_{m=0}^{\infty} S_d^{\textcircled{s}m} \tag{18}$$

and

$$F_d^{\perp} \equiv \left(\bigotimes_{m=1}^{\infty} (S_d^{\perp})^{\textcircled{s}m} \right) \textcircled{s}\, F_d. \tag{19}$$

Clearly,

$$F = F_d \otimes F_d^{\perp}. \tag{20}$$

Now we are able to define a new d cutoff Hamiltonian

$$H_d(\alpha,\beta,\rho) = H_{0phd} + (\mathbf{p} - \mathbf{P}|_d)^2/2 - \beta V(\mathbf{r}) + \alpha^{1/2} H_{Id}(\rho), \tag{21}$$

with

$$H_{0phd} = \int d^3k\, \omega(\mathbf{k})|_d a^+(\mathbf{k})a(\mathbf{k}), \tag{22}$$

$$\mathbf{P}|_d = \int d^3k\, \mathbf{k}|_d a^+(\mathbf{k})a(\mathbf{k}), \tag{23}$$

and

$$H_{Id}(\rho) = \int d^3k\, [g(\mathbf{k})|_d a(\mathbf{k}) + g^*(\mathbf{k})|_d a^+(\mathbf{k})] \times \Theta(\rho - k). \tag{24}$$

Using the same methods as for $H(\alpha,\beta,\rho)$, one easily verifies that $H_0(\alpha,\beta,\rho)$ is self-adjoint and bounded from below, too.

Moreover, we define a new subspace $J\subseteq S_d$ by

$$f\in J \Leftrightarrow F(\mathbf{k}) \equiv 0, \text{ for } |\mathbf{k}|_d > \rho + 3/\Lambda_d. \tag{25}$$

Let

$$\tilde{\Gamma}_d \equiv \{\mathbf{k}\in\Gamma_d |\ |\mathbf{k}| \leqslant \rho + 3/\Lambda_d\} \tag{26}$$

and

$$W \equiv \bigoplus_{m=0}^{\infty} J^{\textcircled{s}m}, \quad W_{\perp} \equiv \bigoplus_{m=0}^{\infty} J^{\perp\textcircled{s}m}, \quad W^{\perp} = W_{\perp}\ \textcircled{s}\ W. \tag{27}$$

Clearly, $F_d = W \oplus W^{\perp}$ and $H_d(\alpha,\beta,\rho)$ as well as $H_{Id}(\rho)$ leave W invariant.

We remark that there exists a canonical isomorphism between the Hamiltonian $H_d(\alpha,\beta,\rho) \restriction W\otimes L^2(\mathbb{R}^3)$ ($\restriction$ denotes, as usual, the restriction) and the Hamiltonian $H(\alpha,\beta,N)$ of the interaction of an electron with $N \equiv |\tilde{\Gamma}_d|$ phonon modes, confined to the Hilbert space $F(N)\otimes L^2(\mathbb{R}^3)$, where

$$F(N) = \bigoplus_{m=0}^{\infty} (\mathbb{C}^N)^{\textcircled{s}m}, \tag{28}$$

which was pointed out by Glimm and Jaffe.[27] Clearly, N depends on d and ρ. Therefore, for the sake of simplicity, we consider henceforth the latter Hamiltonian

$$H(\alpha,\beta,N) = H_0(N) + \alpha^{1/2} H_I(N), \tag{29}$$

with

$$H_0(N) = \sum_{j=1}^{N} \omega(\mathbf{k}_j)a^+(\mathbf{k}_j)a(\mathbf{k}_j) + \frac{(\mathbf{p} - \mathbf{P}(N))^2}{2} - \beta V(\mathbf{r}), \tag{30}$$

$$\mathbf{P}(N) = \sum_{j=1}^{N} \mathbf{k}_j a^+(\mathbf{k}_j)a(\mathbf{k}_j), \tag{31}$$

$$H_I(N) = \sum_{j=1}^{N} [g(\mathbf{k}_j)a(\mathbf{k}_j) + g^*(\mathbf{k}_j)a^+(\mathbf{k}_j)]. \tag{32}$$

The N dependence of $H(\alpha,\beta,N)$ should not be confused with the ρ dependence of $H(\alpha,\beta,\rho)$. The quantities $H(\alpha,\beta,N)$ and $H(\alpha,\beta,\rho)$ are different Hamiltonians. In (30)–(32), $\{\mathbf{k}_j | j\in\mathbf{N}_N\} = \tilde{\Gamma}_d$. At the end of this section, we remove the discrete momentum lattice. Then we come back to our original Hamiltonians $H_d(\alpha,\beta,\rho)$ and $H(\alpha,\beta,\rho)$.

In the case $\beta\equiv 0$, the spectral properties are well understood. It has been shown[10] that $H(\alpha,0,N)$ is representable as a direct integral:

$$H(\alpha,0,N) = \int_{\oplus} d^3Q\, H_{\mathbf{Q}}(\alpha,0,N), \tag{33}$$

$\mathbf{Q}$ being the "C-number" of the conserved total momentum (see also Ref. 28). In Ref. 28, it is proved that

$$\inf_{\mathbf{Q}}(\inf \text{spec } H_{\mathbf{Q}}(\alpha,0,N)) = \inf \text{spec } H_{\mathbf{Q}\equiv 0}(\alpha,0,N), \tag{34}$$

and Fröhlich[10] has shown that the normalized ground state $|\Phi_0\rangle$ of $H_{\mathbf{Q}\equiv 0}(\alpha,0,N)$, lying in $F(N)$, is nondegenerate up to an arbitrary phase factor. We are now prepared to prove the following proposition.

Proposition 1:

$$\langle\Phi_0|\mathbf{P}(N)|\Phi_0\rangle = 0. \tag{35}$$

Proof: By Wigner's theorem, there exists a unitary oper-

ator U such that

$$Ua(\mathbf{k}_j)U^{-1} = a(-\mathbf{k}_j), \tag{36}$$

$$Ua^+(\mathbf{k}_j)U^{-1} = a^+(-\mathbf{k}_j), \tag{37}$$

for all $\mathbf{k}_j \in \tilde{\Gamma}_d$. Then, we find

$$UH_{\mathbf{Q}=\mathbf{0}}(\alpha,0,N)U^{-1} = H_{\mathbf{Q}=\mathbf{0}}(\alpha,0,N) \tag{38}$$

and

$$U\mathbf{P}(N)U^{-1} = -\mathbf{P}(N). \tag{39}$$

Because of the nondegeneracy of $|\Phi_0\rangle$, (38) implies

$$U|\Phi_0\rangle = \exp(i\lambda)|\Phi_0\rangle, \quad \lambda \in \mathbb{R}. \tag{40}$$

Therefore, by (39) and (40), $\langle\Phi_0|\mathbf{P}(N)|\Phi_0\rangle = -\langle\Phi_0|\mathbf{P}(N)|\Phi_0\rangle$, which implies (35). □

We now determine the essential spectrum of $H(\alpha,\beta,N)$.

Lemma 2:

$$\sigma_{\text{ess}}(H(\alpha,\beta,N)) = \sigma_{\text{ess}}(H(\alpha,0,N)). \tag{41}$$

Proof: By Weyl's theorem (see, e.g., Ref. 22), we have to prove that V is a relative form compact perturbation, i.e., that for $\zeta \notin \text{spec } H(\alpha,\beta,N) \cup \text{spec } H(\alpha,0,N)$,

$$\begin{aligned}(H(\alpha,\beta,N)-\zeta)^{-1} &- (H_0(N)+\alpha^{1/2}H_I(N)-\zeta)^{-1} \\ &= \beta(H(\alpha,\beta,N)-\zeta)^{-1}V(H_0(N) \\ &\quad + \alpha^{1/2}H_I(N)-\zeta)^{-1}\end{aligned} \tag{42}$$

is compact. We use the norm-convergent resolvent expansions

$$\begin{aligned}&(H(\alpha,\beta,N)-\zeta)^{-1} \\ &\quad = \sum_{n=0}^{\infty}[(H_0(N)-\beta V-\zeta)^{-1}(-\alpha^{1/2}H_I(N))]^n \\ &\qquad \times (H_0(N)-\beta V-\zeta)^{-1},\end{aligned} \tag{43}$$

$$\begin{aligned}&(H_0(N)+\alpha^{1/2}H_I(N)-\zeta)^{-1} \\ &\quad = (H_0(N)-\zeta)^{-1} \\ &\qquad \times\left(\sum_{n=0}^{\infty}[-\alpha^{1/2}H_I(N)(H_0(N)-\zeta)^{-1}]^n\right).\end{aligned} \tag{44}$$

For Re ζ sufficiently small and negative, the second factor of (44) as well as the first factor of (43) define bounded operators. Therefore, to establish the compactness of (42) it is sufficient to show that

$$(H_0(N)-\beta V-\zeta)^{-1}V(H_0(N)-\zeta)^{-1} \tag{45}$$

is compact. Now, we observe that the operators $a^+(\mathbf{k}_j)$ and $a(\mathbf{k}_j)$, commute with (45). Hence we can classify the spectrum of (45) by a set of natural numbers $L \equiv (n_1,\ldots,n_N)$, $n_j \in \mathbb{N}^0$, where the spectrum of (45) tends to zero as $|\mathbf{L}| \to \infty$. Therefore, all that remains to prove is that (45), restricted to a subspace with L fixed, is compact. Choosing new momentum and position operators

$$\mathbf{p}_n = \mathbf{p} - \sum_{j=1}^{N}\mathbf{k}_j n_j, \tag{46}$$

$$\mathbf{r}_n = \mathbf{r}, \tag{47}$$

this problem is clearly the same as studying the one-particle problem

$$\begin{aligned}&\left(\sum_{j=1}^{N}\omega(\mathbf{k}_j)n_j + \frac{\mathbf{p}_n^2}{2} - \beta V(\mathbf{r}_n) - \zeta\right)^{-1}V(\mathbf{r}_n) \\ &\quad \times\left(\sum_{j=1}^{N}\omega(\mathbf{k}_j)n_j + \frac{\mathbf{p}_n^2}{2} - \zeta\right)^{-1}.\end{aligned} \tag{48}$$

The compactness of (48) was shown by Reed and Simon (see pp. 117–118 in Ref. 22). This implies that (42) is compact, too, and our proof of Lemma 2 is finished. □

We are now able to prove the existence and stability of bound states and state the following theorem.

Theorem 3: Let $N(H)$ denote the number of bound states of the Hamiltonian H, i.e., the number of states lying below the continuum edge. Then,

$$N(H(\alpha,\beta,N)) \geqslant N(H_e(\beta)). \tag{49}$$

Proof: The idea of the proof is to apply a generalization of the Rayleigh–Ritz principle. Let $|\varphi_n\rangle$ be the normalized eigenfunctions of the one-particle Hamiltonian (9):

$$H_e(\beta)|\varphi_n\rangle = E_n|\varphi_n\rangle,$$

$$\text{with } E_n < 0, \quad n \in \mathbb{N}^0, \quad n < N(H_e(\beta)). \tag{50}$$

Consider the "trial functions" $|\varphi_n\rangle \otimes |\Phi_0\rangle \in D(H(\alpha,\beta,N))$. Because of Proposition 1, we obtain

$$\begin{aligned}&\langle\varphi_n| \otimes \langle\Phi_0|H(\alpha,\beta,N)|\Phi_0\rangle \otimes |\varphi_m\rangle \\ &\quad = \delta_{nm}\langle\Phi_0|H_{\mathbf{Q}=\mathbf{0}}(\alpha,0,N)|\Phi_0\rangle \\ &\qquad + \langle\varphi_n|H_e(\beta)|\varphi_m\rangle - \langle\varphi_n|\mathbf{p}|\varphi_m\rangle\langle\Phi_0|\mathbf{P}(N)|\Phi_0\rangle \\ &\quad = \delta_{nm}(\inf\text{spec}\,H(\alpha,0,N) + E_n).\end{aligned} \tag{51}$$

Hence the Rayleigh–Ritz technique tells us that we have found upper bounds on the exact bound states, which—because of Lemma 2—belong to the discrete part of spec $H(\alpha,\beta,N)$. Consequently, (49) is established. □

Returning to our original Hamiltonian $H(\alpha,\beta,\rho)$, we state the following lemma.

Lemma 4: Let $E(\alpha,\beta,\rho) = \inf\text{spec}\,H(\alpha,\beta,\rho)$ and let f_δ be a positive C^∞ function on $\mathbb{R}$ with $f_\delta(0) = 1$ and $f_\delta(x) = 0$ for $x \geqslant \delta$. Then $f_\delta(H(\alpha,\beta,\rho) - E(\alpha,\beta,\rho))$ is compact, if

$$\delta < \Delta(\rho) \equiv \min\{\omega_0, E(\alpha,0,\rho) - E(\alpha,\beta,\rho)\} > 0. \tag{52}$$

Proof: First, via the isomorphism mentioned above all our spectral results for $H(\alpha,\beta,N)$ are directly transferable to the Hamiltonian $H_d(\alpha,\beta,\rho) \upharpoonright W \otimes L^2(\mathbb{R}^3)$.

Let

$$E_d(\alpha,\beta,\rho) = \inf\text{spec}\,H_d(\alpha,\beta,\rho) \upharpoonright F_d \otimes L^2(\mathbb{R}^3).$$

The same calculations as in Ref. 13 (the second step of the proof of lemma 2.1 in Ref. 13) yield

$$\inf\text{spec}\,(H_d(\alpha,\beta,\rho) \upharpoonright W^\perp \otimes L^2(\mathbb{R}^3)) \geqslant \omega_0 + E_d(\alpha,\beta,\rho). \tag{53}$$

Since $F_d = W \oplus W^\perp$, it follows from Lemma 2 that

$$f_\delta(H_d(\alpha,\beta,\rho) - E_d(\alpha,\beta,\rho)) \upharpoonright F_d \otimes L^2(\mathbb{R}^3)$$

is compact, if

$$\delta < \Delta_d \equiv \min\{\omega_0, E_d(\alpha,0,\rho) - E_d(\alpha,\beta,\rho)\}.$$

From Theorem 3 we know that $\Delta_d > 0$.

We apply this argumentation once more: The same calculations as in Ref. 10 (Corollary 2.2.iii) can be used to show

$$\inf \operatorname{spec} H_d(\alpha,\beta,\rho) \upharpoonright F_d^{\perp} \otimes L^2(\mathbb{R}^3)$$
$$\geqslant \omega_0 + \inf \operatorname{spec} H_d(\alpha,\beta,\rho) \upharpoonright F \otimes L^2(\mathbb{R}^3). \tag{54}$$

Since $F = F_d \oplus F_d^{\perp}$, it follows that

$$\inf \operatorname{spec} H_d(\alpha,\beta,\rho) \upharpoonright F \otimes L^2(\mathbb{R}^3) = E_d(\alpha,\beta,\rho)$$

and that

$$f_\delta(H_d(\alpha,\beta,\rho) - E_d(\alpha,\beta,\rho)) \upharpoonright F \otimes L^2(\mathbb{R}^3)$$

is compact, if $\delta < \Delta_d$. The methods used by Fröhlich (Lemma 2.1 in Ref. 10) show that $H_d(\alpha,\beta,\rho) \to H(\alpha,\beta,\rho)$ in norm resolvent convergence as $d \to \infty$. Consequently, $E_d(\alpha,\beta,\rho) \to E(\alpha,\beta,\rho)$ as $d \to \infty$ and $f_\delta(H(\alpha,\beta,\rho) - E(\alpha,\beta,\rho))$ remains compact if $\delta < \Delta(\rho)$. The same proof as in Theorem 3 results in $\Delta(\rho) > 0$, which completes the proof of Lemma 4. □

All what remains to do is to remove the UV cutoff ρ.

Theorem 5: Let $E(\alpha,\beta) = \inf \operatorname{spec} H(\alpha,\beta)$ and f_δ as in Lemma 4. Then, $f_\delta(H(\alpha,\beta) - E(\alpha,\beta))$ is compact, if $\delta < \Delta$ where

$$\Delta \equiv \min(\omega_0, E(\alpha,0) - E(\alpha,\beta)) > 0. \tag{55}$$

Proof: Transforming the Hamiltonian $H(\alpha,\beta,\rho)$ with the canonical transformation e^T, where

$$T \equiv T_{\rho\Lambda} = \int d^3k \,(C_{\rho\Lambda}(\mathbf{k}) a(\mathbf{k}) - \text{H.c.}) \tag{56}$$

with

$$C_{\rho\Lambda}(\mathbf{k}) \equiv C(\mathbf{k}) = -\alpha^{1/2} g(\mathbf{k}) \Theta(g-k)\Theta(k-\Lambda)/(\omega(\mathbf{k}) + k^2/2), \quad 0 < \Lambda < \rho, \tag{57}$$

we obtain

$$e^T H(\alpha,\beta,\rho) e^{-T} \equiv H^T(\alpha,\beta,\rho) = H_0(\beta) + \alpha^{1/2} H_I(\rho) + (\mathbf{Z} + \mathbf{Z}^+)^2/2 - (\mathbf{p} - \mathbf{P})\cdot\mathbf{Z} - \mathbf{Z}^+(\mathbf{p} - \mathbf{P}) + \Sigma, \tag{58}$$

where

$$H_0(\beta) = H_{0\text{ph}} + (\mathbf{p} - \mathbf{P})^2/2 - \beta V(\mathbf{r}), \tag{59}$$

$$\mathbf{Z} = \int d^3k \, \mathbf{k} C_{\rho\Lambda}(\mathbf{k}) a(\mathbf{k}), \tag{60}$$

and

$$\Sigma = \int d^3k \,[\omega(\mathbf{k}) |C(\mathbf{k})|^2 + \alpha^{1/2} g^*(\mathbf{k}) C(\mathbf{k}) + \alpha^{1/2} g(\mathbf{k}) C^*(\mathbf{k})] < \infty. \tag{61}$$

Again we can use Fröhlich's methods (see Sec. 2.2 in Ref. 10) to prove the following facts: For all $\epsilon > 0$ there exists a $\Lambda < \infty$ such that

$$|H^T(\alpha,\beta,\rho) - H_0(\beta)| \leqslant \epsilon H_0(\beta) + b(\Lambda), \tag{62}$$

where $b(\Lambda)$ is uniform in $\rho \leqslant \infty$. Furthermore

$$\operatorname*{norm-lim}_{\rho\to\infty} (\zeta - H^T(\alpha,\beta,\rho))^{-1} = (\zeta - H^T(\alpha,\beta))^{-1} \tag{63}$$

exists, where $H^T(\alpha,\beta)$ is a unique self-adjoint operator bounded from below. Also

$$\operatorname*{s-lim}_{\rho\to\infty} \exp(T_{\rho\Lambda}) \equiv \exp(T_{\infty\Lambda})$$

exists as a unitary operator. Therefore

$$H(\alpha,\beta) \equiv \exp(-T_{\infty\Lambda}) H^T(\alpha,\beta) \exp(T_{\infty\Lambda})$$

is self-adjoint and bounded from below, too. Finally, the norm-resolvent convergence (63) together with Lemma 4 implies that $f_\delta(H(\alpha,\beta) - E(\alpha,\beta))$ remains compact for $\delta < \Delta$. □

We now determine the continuum edge of $H(\alpha,\beta)$.

Theorem 6: Let $E(\alpha,\beta) = \inf \operatorname{spec} H(\alpha,\beta)$. The continuum edge begins exactly at the point

$$\Omega \equiv \min(E(\alpha,\beta) + \omega_0, E(\alpha,0)). \tag{64}$$

Proof: This Ω is a lower bound for the continuum edge because Theorem 5 implies that all eigenvalues smaller than Ω are discrete. Furthermore, without loss of generality we may assume that the number of eigenvalues being smaller than Ω is finite. Otherwise these infinite eigenvalues have to accumulate at Ω and (64) is trivially proved.

Suppose first

$$E(\alpha,0) \leqslant E(\alpha,\beta) + \omega_0. \tag{65}$$

Since the absolute continuous spectrum of $H_e(\beta)$ begins at zero, we can always find functions $|\psi_n\rangle \in L^2(\mathbb{R}^3)$ $(n \in \mathbb{N})$ with $\langle\psi_n|\psi_m\rangle = \delta_{nm}$ and with $\langle\psi_n|H_E(\beta)|\psi_m\rangle = \delta_{nm} E_n$, where $E_n > 0$ and $E_n \to 0$ as $n \to \infty$. Choosing the trial functions $|\Phi_0\rangle \otimes |\psi_n\rangle$, where $|\Phi_0\rangle$ denotes the ground state of the free polaron Hamiltonian subjected to total momentum $\mathbf{Q} \equiv \mathbf{0}$, we calculate, as in the proof of Theorem 3,

$$\langle\psi_n| \otimes \langle\Phi_0|\Phi_0\rangle \otimes |\psi_m\rangle = \delta_{nm}$$

and

$$\langle\psi_n| \otimes \langle\Phi_0|H(\alpha,\beta,)|\Phi_0\rangle \otimes |\psi_m\rangle = \delta_{nm}(E(\alpha,0) + E_n). \tag{66}$$

Since $E_n \to 0$ as $n \to \infty$, a modification of the mini–max principle (see, e.g., Reed and Simon,[22] Theorem XIII.1) ensures us that $E(\alpha,0)$ is the bottom of the essential spectrum of $H(\alpha,\beta)$.

In the second case

$$E(\alpha,\beta) + \omega_0 < E(\alpha,0), \tag{67}$$

we again use a trial function argument, but now with different fuctions involving one-phonon states. By (4) we know that there exists a $\mathbf{q} \in \mathbb{R}^3$ with $\omega(\mathbf{k}) \to \omega_0$ as $\mathbf{k} \to \mathbf{q}$. Without loss of generality we assume $q < \infty$, the case $q = \infty$ can be treated quite similarly. We choose $\epsilon > 0$ fixed.

First, we need some definitions. Let $U(\delta,\mathbf{q})$ denote a ball around $\mathbf{q}$ with radius δ. We construct "disks" $D(n)$ as follows:

$$D(n) = U(\epsilon 2^{-n},\mathbf{q}) \setminus U(\epsilon 2^{-n-1},\mathbf{q}). \tag{68}$$

Let $H(\mathbf{q})$ be the Hamiltonian that results if one replaces $\mathbf{p}$ in $H(\alpha,\beta,)$ by $\mathbf{p} - \mathbf{q}$, $\mathbf{q}$ being a C-number. Obviously, $H(\mathbf{q})$ is unitarily equivalent to $H(\alpha,\beta)$. The ground state of $H(\mathbf{q})$ is denoted by $\psi(\mathbf{q}) \in \mathscr{H}$. Furthermore, we define a projection operator P_A $(A \subseteq \mathbb{R}^3)$ that annihilates all phonon parts with momentum $\mathbf{k} \in A$ by

$$P_A = \sum_{n=1}^{\infty} \int d^3k_1 \cdots \int d^3k_n \, \chi_A(\mathbf{k}_1) \cdots \chi_A(\mathbf{k}_n) \times a(\mathbf{k}_1) \cdots a(\mathbf{k}_n), \tag{69}$$

where $\chi_A(\mathbf{k})$ is the characteristic function equal to 1 if $\mathbf{k} \in A$ and 0 otherwise.

We are now able to give the explicit form of our trial functions $\Phi_{n\epsilon}$, $n\in\mathbb{N}$,

$$\Phi_{n\epsilon} = [(P_{U(\epsilon,\mathbf{q})}\psi(\mathbf{q}))\,\textcircled{S}\,\varphi_{n\epsilon}]/\|(P_{U(\epsilon,\mathbf{q})}\psi(\mathbf{q}))\,\textcircled{S}\,\psi_{n\epsilon}\|, \tag{70}$$

where $\varphi_{n\epsilon}$ is a one-phonon state, $\varphi_{n\epsilon}\geqslant 0$, whose momentum distribution is explicitly given by

$$|\varphi_{n\epsilon}(\mathbf{k})|^2\, d^3k = \chi_{D(n)}(\mathbf{k}) d^3k. \tag{71}$$

For ϵ sufficiently small, $\Phi_{n\epsilon}\in\mathscr{H}$ and $\Phi_{n\epsilon}\neq 0$. Since $U(\epsilon,\mathbf{q})\supseteq D(n)$ and since the $D(n)$ are pairwise disjunct, one calculates

$$\langle\Phi_{n\epsilon}|\Phi_{m\epsilon}\rangle = \delta_{nm}, \quad \langle\Phi_{n\epsilon}|H(\alpha,\beta)|\Phi_{m\epsilon}\rangle = E_n(\epsilon)\delta_{nm}. \tag{72}$$

Further inspection of E_n shows that $E_n(\epsilon)\to\omega_0 + E(\alpha,\beta,\epsilon)$ as $n\to\infty$, where $\omega_0(\epsilon)\to\omega_0$ and $E(\alpha,\beta,\epsilon)\to E(\alpha,\beta)$ as $\epsilon\to 0$. Since $\epsilon>0$ can be chosen arbitrarily small, again the minimax principle tells us that the continuum edge has the upper bound $\omega_0 + E(\alpha,\beta)$.

Putting all facts together, we finally finish the proof of Theorem 6. □

The physical interpretation of the two possibilities (64) for the continuum edge is easily understood. In the case (67) the continuum involves scattering states with one real phonon of energy ω_0 present. On the other hand, if (65) holds, the continuum at $E(\alpha,0)$ consists of delocalized electron states.

Obviously, the Rayleigh–Ritz argument of Theorem 3 can be done for $H(\alpha,\beta)$, too. If (65) holds, this implies the existence and stability of the bound states and gives furthermore simple upper bounds on the associated energies.

Corollary 7: Let $E(\alpha,0) < E(\alpha,\beta) + \omega_0$ and let $N(H)$ be the number of states of the Hamiltonian H below the continuum edge. Then,

$$N(H(\alpha,\beta))\geqslant N(H_e(\beta)), \quad \text{for all } \alpha\geqslant 0. \tag{73}$$

□

We know from Theorem 5 that $E(\alpha,\beta)$ is an eigenvalue. Now we prove the next lemma.

Lemma 8: The ground state of $H(\alpha,\beta)$ is nondegenerate.

Proof: We represent the underlying Hilbert space now as

$$L^2(Q,d\mu)\otimes L^2(\mathbb{R}^3,d^3x), \tag{74}$$

where $L^2(Q,d\mu)$ is the phonon Q space, which is isomorphic to the Fock space F (see Simon[29] for a detailed discussion). If one takes the Schrödinger representation ($\mathbf{r}$ representation) for the electron coordinate, the operator

$$L\equiv -\beta V(\mathbf{r}) + \alpha^{1/2}H_I \tag{75}$$

acts as a multiplication operator. The operator L can be approximated by bounded multiplication operators L_n, such that $H_0 + L_n\to H(\alpha,\beta)$ and $H(\alpha,\beta) - L_n\to H_0$ in the strong resolvent sense as $n\to\infty$. This holds for arbitrary cutoff $\rho\leqslant\infty$. We know from the proof of Theorem 5 that the operators $H_0 + L_n$ and $H(\alpha,\beta) - L_n$ are uniformly bounded from below. Therefore, Theorem XIII.45 of Reed and Simon[22] is applicable (see also Ref. 28). This implies that, in order to prove Lemma 8, we have to show that $\exp(-H_{0ph} - (\mathbf{p}-\mathbf{P})^2/2)$ is positivity improving in the chosen representation. We write $\exp(-(\mathbf{p}-\mathbf{P})^2/2)$ as Fourier transform

$$\exp(-(\mathbf{p}-\mathbf{P})^2/2) = (2\pi)^{-3/2}\int d^3\lambda \exp\left(-\frac{1}{2}\lambda^2\right)e^{i\lambda\mathbf{p}}e^{-i\lambda\mathbf{P}}. \tag{76}$$

Now, $\exp(-H_{0ph})$ is positivity improving and $\exp(-i\lambda\mathbf{p})$ is positivity preserving with respect to the phonon Q space; $\exp(i\lambda\mathbf{p})$ acts as translational operator in the $\mathbf{r}$ representation of the electron coordinate. Since $\exp(-\frac{1}{2}\lambda^2)$ is strictly positive, we get that $\exp(-H_{0ph} - \frac{1}{2}(\mathbf{p}-\mathbf{P})^2)$ is positivity improving in the chosen representation. This implies that $\exp(-H(\alpha,\beta))$ is positivity improving and consequently $E(\alpha,\beta)$ is a simple eigenvalue. □

Summarizing, we have proved in this section that the Hamiltonian $H(\alpha,\beta)$ for a bound optical polaron is a well-defined self-adjoint operator, bounded from below. If the one-particle Hamiltonian (9) has a bound ground state, then also $H(\alpha,\beta)$ has a discrete bound ground state that is nondegenerate. This ground state is separated from the continuous spectrum by a gap whose magnitude was exactly determined: It is the minimum of the phonon dispersion ω_0 or the difference $E(\alpha,0) - E(\alpha,\beta)\geqslant|E_0(\beta)|$, where $E_0(\beta)$ is the ground-state energy of the one-particle Hamiltonian (9).

III. CONSEQUENCES AND GENERALIZATIONS

To begin with, we state that the associated forms of the resolvent of Eq. (63) $\{\zeta - H^T(\alpha,\beta)\}^{-1}$ are an analytic family of type (B) in the sense of Kato[30] in both parameters α (see Fröhlich[10]) and β (see Simon[31]) for $\alpha,\beta\geqslant 0$. Since Lemma 8 implies that $E(\alpha,\beta)$ is an isolated, simple eigenvalue for $\alpha\geqslant 0$ [β being such that $H_e(\beta)$ has a negative eigenvalue], the standard analytical perturbation theory (see Kato[30]) is applicable. It follows from Hartog's theorem that $E(\alpha,\beta)$ is jointly real analytic in α and β in the specified domain. The same is true for the discrete excited states, if they are not degenerate. Moreover, the associated wave functions are analytic in α and β, too.

Let $|\psi_0(\alpha,\beta)\rangle$ be the wave function of the ground state of $H(\alpha,\beta)$. Then the mean number of virtual phonons in the ground state is defined by

$$N(\alpha,\beta) = \langle\psi_0(\alpha,\beta)|\int d^3k\; a^+(\mathbf{k})a(\mathbf{k})|\psi_0(\alpha,\beta)\rangle. \tag{77}$$

Furthermore, several possibilities were proposed to define a polaron radius and a self-induced potential as quantities derived from the ground-state expectation values of H_I (see, e.g., Peeters and Devreese[32]). Clearly, $N(\alpha,\beta)$ as well as the polaron radius and the self-induced potential are analytic in α and β for $\alpha\geqslant 0$, β as above.

We conclude that all changes in the bound polaron state are not accompanied by a nonanalytical behavior, but are smooth transitions.

We now add some remarks on possible extensions of our theory. First, one may consider an optical polaron in arbitrary spatial dimension (see Peeters, Wu Xiaoguang, and Devreese[33]). The conditions (4)–(8) are readily generalized to arbitrary dimensions (see Simon[31] for an extension of the Rollnik condition). Then, the same proof is possible.

We mention two physical interesting examples. First, Sak[34] (see also Degani and Hipolito[6]) considers an electron

that couples to the optical surface phonon modes and is bound in the perpendicular direction to a Coulomb potential resulting from its image charge. The associated Hamiltonian H_s can be cast into the form

$$H_s = (Q_1 - P_1)^2/2 + (Q_2 - P_2)^2/2 + p_3/2 - \beta V(r_3) + \omega_s \int d^2k\, a^+(\mathbf{k}) a(\mathbf{k}) + \alpha^{1/2} \int d^2k\, k^{-1/2} \times \exp(-k_3 r_3)[a(\mathbf{k}) + a^+(\mathbf{k})]. \quad (78)$$

The Hilbert space $\mathscr{H}_s$ belonging to H_s is

$$\mathscr{H}_s = F \otimes L^2([0,\infty[). \quad (79)$$

The parameters Q_1 and Q_2 correspond to the conserved components of the momentum. To get inf spec H_s, one may set $Q_1 = Q_2 = 0$ (see Ref. 28). For $V(r_3)$ we do not take $1/r_3$, as Sak does, but for mathematical and physical reasons (see Cole[35]) we have to take a cutoff potential:

$$V(\mathbf{z}) = \begin{cases} 1/\mathbf{z}, & \text{for } \mathbf{z} \geqslant b, \\ 1/b, & \text{for } \mathbf{z} < b, \end{cases} \quad (80)$$

where the cutoff b is a strictly positive constant. Without going into the mathematical details, we remark that our methods are applicable to H_s. In particular, the ground-state energy is analytic in α and β. This is in a marked contrast to the work of Tokuda.[18] The above model can be extended to include bulk phonon effects, which was discussed recently by Gu and Zheng.[36]

A second example concerns a quasi-two-dimensional polaron in polar quantum wells, bound to a two-dimensional Coulomb potential, which was studied by Mason and Das Sarma.[5]

IV. THE PINNING TRANSITION

Up to now, for all potentials considered, the associated ground-state energies are analytic in α or β. One may ask the question the other way around: Which potentials lead to a ground-state energy that is nonanalytic in α or β? This brings us back directly to our condition that the one-particle Hamiltonian (9) has at least one bound state. In one or two dimensions, it is well known that an attractive potential always leads to a bound state. No so in three dimensions; the question of whether or not the one-particle Hamiltonian has a bound state depends sensitively on the mass of the particle for short-range potentials. The idea of Spohn[26] is to describe the polaron problem approximately as a one-particle problem with an effective mass $m(\alpha)$ and to study then the occurrence of bound states with increasing α. For a suitable static binding potential, at a critical coupling α_c a pinning transition is obtained, i.e., by the phonon-induced mass enhancement of the electron, a new bound state suddenly arises from the continuum.

To get a connection with our results, we consider a slightly different situation: Let α be fixed and vary β ($\beta \in \mathbb{R}$). For the sake of definiteness, let V be an element of the Rollnik class R [see (8)] and let V be negative ($V \leqslant 0$). The occurrence of bound states of $H_e(\beta)$ is well understood (see Refs 21–23). The Birman–Schwinger bound shows that for all $\beta \in \mathbb{R}$ the number of bound states $N(H_e(\beta))$ is finite and that $N(H_e(\beta)) = 0$ for $\beta < \beta_c$, where $\beta_c > 0$. Therefore, inf spec $H_e(\beta)$ is nonanalytic for $\beta = \beta_c$, corresponding to a localization transition (pinning transition) of the ground state.

We prove that such a transition is obtained even if the electron–phonon coupling is nonzero and state the following.

Theorem 4.9: Let the potential V be in the Rollnik class R for spatial dimension $d = 3$ and let $V \leqslant 0$. Let the ground-state energy of the one-particle Hamiltonian $H_e(\beta)$ be nonanalytic for $\beta = \beta_c > 0$. Then the ground-state energy $E(\alpha,\beta)$ of the bound polaron is nonanalytic for $\beta = \beta_c(\alpha)$, where $\beta_c(\alpha)$ is a unique number with

$$0 \leqslant \beta_c(\alpha) \leqslant \beta_c, \quad (81)$$

and $\beta_c(\alpha)$ is continuous in α for $0 \leqslant \alpha < \infty$.

Proof: Clearly, $E(\alpha,\beta)$ is monotone decreasing (and concave) in β. From Theorem 6, it then follows that $E(\alpha,\beta) = E(\alpha,0)$, for $\beta \leqslant 0$. On the other hand, we know from Corollary 7 that $E(\alpha,\beta) < E(\alpha,0)$ for $\beta > \beta_c$ and that $E(\alpha,\beta)$ is analytic in β for $\beta > \beta_c$. Thus $E(\alpha,\beta)$ cannot be analytic in β in the total interval $[0,\beta_c]$ because the identity theorem for holomorphic functions requires that then $E(\alpha,\beta) \equiv E(\alpha,0)$. Therefore, there exists a nonanalyticity $\beta_c(\alpha)$, with $0 \leqslant \beta_c(\alpha) \leqslant \beta_c$. At $\beta = \beta_0(\alpha)$, $E(\alpha,\beta)$ abandons the continuum edge. Because of the monotonicity of $E(\alpha,\beta)$ in β, $E(\alpha,\beta)$ is separated by a gap from the continuum for all $\beta > \beta_c(\alpha)$. Analytical perturbation theory ensures us that $E(\alpha,\beta)$ is analytic in β for all $\beta > \beta_c(\alpha)$. Therefore the nonanalyticity $\beta_c(\alpha)$ is a unique number with $0 \leqslant \beta_c(\alpha) \leqslant \beta_c$. The continuity of $\beta_c(\alpha)$ in α follows directly from analytical perturbation theory and from the monotonicity of $E(\alpha,\beta)$ in β. □

We remark that the same proof can be done to show that the energy of the nth discrete excited state is nonanalytic at the point where it is pushed out of the continuum edge.

Clearly, $\beta_c(0) \equiv \beta_c$ and we conjecture that $\beta_c(\alpha)$ is monotone decreasing in α and that $\beta_c(\alpha) \to 0$ as $\alpha \to \infty$. An estimation on $\beta_c(\alpha)$, which is better than (81), however, requires a nontrivial extension of our result. We leave this as an open problem.

We summarize our results in two figures. In Fig. 1 we sketch $E(\alpha,\beta)$ for three different fixed values of α and vary β.

In Fig. 2, we give a qualitative picture of the phase diagram of the pinning transition in the α-β plane describing the

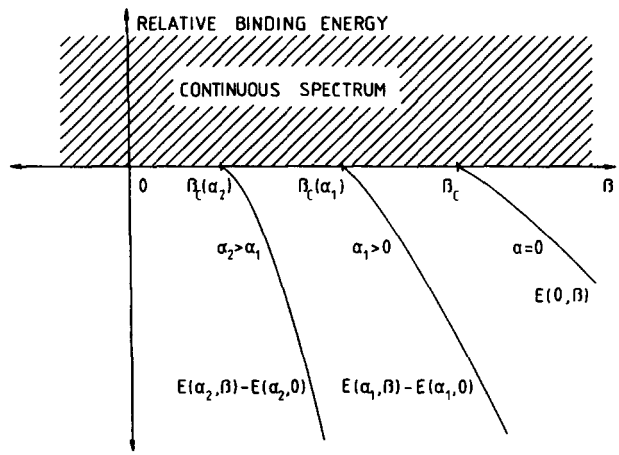


FIG. 1. Qualitative picture of the pinning transition: relative binding energy versus potential strength β for different values of the coupling α.

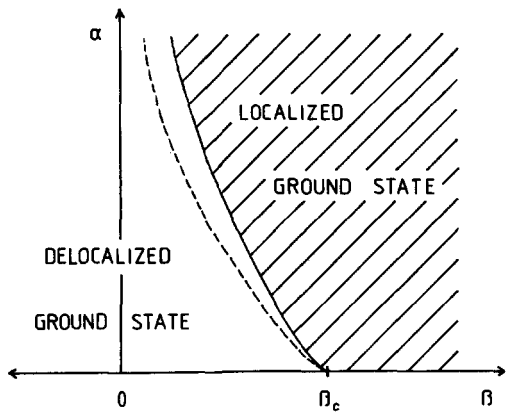


FIG. 2. Phase diagram of the pinning transition (qualitative picture). The solid line represents the exact solution, the dashed line the effective-mass approximation.

pinning transition from a delocalized to a localized state.

Let us now discuss the effective-mass approximation [i.e., the approximation of the polaron as a single-particle with mass $m(\alpha)$] in some more detail. For a Coulomb potential, Mason and Das Sarma[5] compare the ground-state energy shifts for fixed small α and varied β between the "exact" (variational) solution and the effective-mass approximation. It turns out that the effective-mass approximation yields an overestimation of the energy shift, being asymptotically correct for small β but becoming worse for intermediate and large β. Transferring this result to a short-range potential the situation is quite the same. One may conjecture that the one-particle approximation leads to a value of $\beta_c(\alpha)$ that is too small. This belief is based on the intuitive argument that a bound electron cannot use all phonons in such a way to raise its effective mass as a free electron. The one-particle approximation should only work for small α, β being small, too.

By a simple scaling argument, one finds the critical coupling strength in the effective-mass approximation by $\beta_c(\alpha)_{\text{eff}}$,

$$\beta_c(\alpha)_{\text{eff}} = \beta_c/m(\alpha), \tag{82}$$

if $m(0) = m = 1$. Therefore, critical lines for different potentials, but for the same dispersion and coupling function, are proportional in the effective-mass approximation, the potential merely determines the prefactor β_c. We have also indicated the qualitative behavior of the critical line for the effective-mass approximation in Fig. 2 (dashed line).

A finite temperature $T>0$ destroys the pinning transition. This can be seen considering the (formal) free energy (instead of the ground-state energy) in the path integral representation. The free energy is analytic in all parameters $\alpha \geqslant 0$, $\beta \geqslant 0$, and $T>0$, if the potential $V(\mathbf{r})$ is short range or if $V(\mathbf{r})$ is a long-range Coulomb potential. As for details, we refer to Ref. 25.

Finally, we give the phase diagram in the effective mass approximation for an optical Fröhlich polaron for two concrete examples: First a spherical square well

$$V(\mathbf{r}) = \Theta(1-r), \tag{83}$$

and, second, a screened Coulomb potential

$$V(\mathbf{r}) = \exp(-r)/r. \tag{84}$$

In the case of a spherical square well the eigenvalues and eigenfunctions are well-known (see, e.g., Messiah[37]). In particular, the critical potential strength turns out to be

$$\beta_c = \beta_c(0) = \pi^2/8 = 1.233\,7005\ldots. \tag{85}$$

For a screened Coulomb potential, β_c is not known analytically. Kesarwani and Varshni[38] determine β_c numerically as

$$0.839\,9032 \leqslant \beta_c \leqslant 0.839\,9039. \tag{86}$$

For the usual Fröhlich model

$$\omega(\mathbf{k}) = 1, \quad g(\mathbf{k}) = (8\pi^2)^{1/4}/k, \tag{87}$$

the polaron mass $m(\alpha)$ was calculated in Ref. 39. Consequently, all variables of (82) are known. The limiting cases (see, again, Ref. 39) are

$$\beta_c(\alpha)_{\text{eff}} = \beta_c \cdot (1 - \alpha/6) + O(\alpha^2), \quad \text{as } \alpha \to 0, \tag{88}$$

$$\beta_c(\alpha)_{\text{eff}} = 44.05 \cdot \beta_c \cdot \alpha^{-4}, \quad \text{as } \alpha \to \infty. \tag{89}$$

The effective-mass approximation of the phase diagrams for the Fröhlich polaron and a spherical square well (resp. a screened Coulomb potential) are shown in Fig. 3.

A variational calculation of $\beta_c(\alpha)$ is in progress and will be published elsewhere.

Concerning experimental consequences, we finally state that first experimental evidences of the pinning transition were observed by Dmochowski *et al.*[40] They found bound polaron states very close in energy and differing strongly in localization. Such a situation just occurs in the neighborhood of the pinning transition.

V. CONCLUSIONS

Summarizing, we have proved the analyticity of polaron quantities in the coupling parameter and the potential strength, if the potential is long range (e.g., for a Coulomb potential) or if the one-particle Hamiltonian has a bound

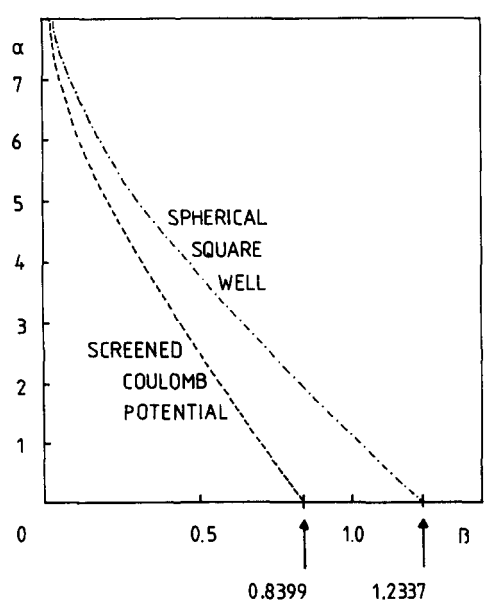


FIG. 3. Critical lines of the pinning transition in the α-β plane in the effective-mass approximation for a spherical square well and a screened Coulomb potential for a Fröhlich polaron.

state. Consequently, no "phase transitions" occur and the shallow–deep instability is continuous in this case.

For a short-range attractive potential we have shown the existence of a pinning transition, which depends on the electron–phonon coupling. This pinning transition is connected with a nonanalyticity of the ground-state energy and with a potential assisted localization transition of the ground state from a delocalized to a localized state as the potential strength increases. We have discussed this pinning transition for a spherical square well and a screened Coulomb potential, giving the phase diagram in the effective-mass approximation.

ACKNOWLEDGMENTS

I am indebted to Professor B. Gerlach, Professor H. Spohn, and Professor K. Veselic for helpful discussions.

Furthermore, financial support by the Studienstiftung des deutschen Volkes is gratefully acknowledged.

[1]V. M. Buimistrov and S. I. Pekar, Zh. Eksp. Teor. Fiz. **32**, 1193 (1957) [Sov. Phys. JETP **5**, 970 (1957)].
[2]P. M. Platzman, Phys. Rev. **125**, 1961 (1962).
[3]D. Larsen, Phys. Rev. **187**, 1147 (1969).
[4]J. Adamowski, Phys. Rev. B **32**, 2588 (1985).
[5]B. A. Mason and S. Das Sarma, Phys. Rev. B **33**, 8379 (1986).
[6]M. H. Degani and O. Hipolito, Phys. Rev. B **35**, 7717 (1987).
[7]H. Fröhlich, H. Pelzer, and S. Zienau, Philos. Mag. **41**, 221 (1950).
[8]E. Nelson, J. Math. Phys. **5**, 1190 (1964).
[9]H. Spohn, Ann. Phys. (NY) **175**, 278 (1987).
[10]J. Fröhlich, Fortschr. Phys. **22**, 159 (1974).
[11]B. Gerlach and H. Löwen, Phys. Rev. B **35**, 4291 (1987).
[12]B. Gerlach and H. Löwen, Phys. Rev. B **35**, 4297 (1987).
[13]H. Löwen, J. Math. Phys. **29**, 1498 (1988).
[14]M. P. A. Fisher and W. Zwerger, Phys. Rev. B **32**, 6190 (1985).
[15]Y. Toyozawa, Physica B **116**, 7 (1983).
[16]D. Larsen, in *Polarons in Ionic Crystals and Polar Semiconductors*, edited by J. T. Devreese (North-Holland, Amsterdam, 1972), pp. 272–287.
[17]N. Tokuda, H. Shoji, and K. Yoneya, J. Phys. C **14**, 4281 (1981).
[18]N. Tokuda, J. Phys. C **15**, 1353 (1982).
[19]M. Matsuura, Can. J. Phys. **52**, 1 (1974).
[20]K. Takegahara and T. Kasuya, J. Phys. Soc. Jpn. **39**, 1292 (1975).
[21]V. Glaser, H. Grosse, A. Martin, and W. Thirring, in *Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann*, edited by E. H. Lieb, B. Simon, and A. S. Wightman (Princeton U.P., Princeton, NJ, 1976), p. 169.
[22]M. Reed and B. Simon, *Analysis of Operators, Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.
[23]B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).
[24]Y. Shinozuka and Y. Toyozawa, J. Phys. Soc. Jpn. **46**, 505 (1979).
[25]H. Löwen, Ph.D. thesis, University of Dortmund, 1987.
[26]H. Spohn, J. Phys. A **19**, 533 (1986).
[27]J. Glimm and A. Jaffe, Ann. Math. **91**, 362 (1970).
[28]B. Gerlach and H. Löwen, "Absence of phonon-induced localization for the free optical polaron and for the corresponding Wannier exciton-phonon system," Phys. Rev. B (in press).
[29]B. Simon, *The $P(\Phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, NJ, 1974).
[30]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1966).
[31]B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U.P., Princeton, NJ, 1971).
[32]F. M. Peeters and J. T. Devreese, Phys. Rev. B **31**, 4890 (1985).
[33]F. M. Peeters, Wu Xiaoguang, and J. T. Devreese, Phys. Rev. B **33**, 3926 (1986).
[34]J. Sak, Phys. Rev. B **10**, 3981 (1972).
[35]M. W. Cole, Phys. Rev. B **2**, 4239 (1970).
[36]S. W. Gu and R. S. Zheng, Sol. State Comm. **62**, 695 (1987).
[37]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1980), Vol. 1, p. 359.
[38]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **19**, 819 (1978).
[39]B. Gerlach, H. Löwen, and H. Schliffke, Phys. Rev. B **36**, 6320 (1987).
[40]J. E. Dmochowski, J. M. Langer, Z. Kalinski, and W. Jantsch, Phys. Rev. Lett. **56**, 1735 (1986).

The Vaidya–Patel solution with Robertson–Walker metric as a rotating inflationary scenario

Öyvind Grön
Oslo College of Engineering, Cort Adelers gt. 30, Oslo, Norway and Institute of Physics, University of Oslo, Oslo, Norway

Harald H. Soleng
Institute of Theoretical Astrophysics, University of Oslo, Norway and Institute of Physics, University of Oslo, Oslo, Norway[a)]

(Received 21 April 1987; accepted for publication 17 February 1988)

The Vaidya–Patel solution of a rotating homogeneous fluid in the presence of a Maxwellian source-free electromagnetic field is interpretated as an inflationary scenario with a gauge field with local $U(1)$ symmetry, a vacuum energy, and a rotating perfect fluid. An explicit solution is found to be expressible in terms of known solutions representing the radiation filled Robertson–Walker universe with a cosmological term. In the case that the rotating fluid is radiation, the discussion of the model is considerably simplified. How the time scale of transition into a pseudo-de Sitter stage, as observed by an observer following the rotating fluid, is affected by vorticity is also studied.

I. INTRODUCTION

The spacelike part of the Robertson–Walker metric was discussed by Bianchi[1] in 1898. This metric represents the most general isotropic and homogeneous space-time geometry. The three different Robertson–Walker metrics are special cases of universes of Bianchi types I, V, and IX in which the three-space is flat, hyperbolic, or closed, respectively. These isotropic special cases of the Bianchi models have since been extensively explored, and a large number of general relativistic cosmological models with this form of the metric has been constructed. Vaidya and Patel[2] have recently shown that the closed Robertson–Walker metric can also be a solution of Einstein's field equations with a rotating homogeneous fluid in the presence of a Maxwellian source-free electromagnetic field. This solution has lately been generalized to the axisymmetric case by Patel and Pandya.[3]

In this paper we will rederive the Vaidya–Patel solution using a more conventional definition of the scale factor than that of Vaidya and Patel. In the case that the fluid satisfies the equation of state $P=\frac{1}{3}\rho$, the Lorentz transformation between the fluid-comoving system and the coordinate-comoving system involves a constant Lorentz factor, which simplifies the study of this particular case considerably. In this case we will further find that exact expressions for the evolution of the scale factor can be stated in terms of well-known solutions of the field equations of the Friedmann–Robertson–Walker equation with a closed space-time filled with radiation and a cosmological term.

The solution, which is discussed here, is a model of a rotating inflationary scenerio in which the energy-momentum density is described as an ultrarelativistic fluid, a vacuum energy, and a source-free gauge field with a local $U(1)$ gauge symmetry. We shall study how vorticity affects the time scale of transition into an inflationary era in this model.

In the case of a rotating universe with inflation, a rotating generalization of the de Sitter solution would be useful. Both Vaidya[4] and Grön[5] have presented solutions that were interpretated as rotating generalizations of the de Sitter universe. In Vaidya's model, the geodesic condition, which is natural in a true vaccum solution, restricts the parameter m to the value zero, which corresponds to zero vorticity. The model of Grön has, among other models, been criticized by Jantzen.[6] His model should not be interpreted as a rotating universe model, but rather as a universe with shear.[7,8]

In the case of true vacuum solutions there cannot be vorticity in a physical sense. The angular momentum of the vacuum fluid will always be zero. Thus there does not exist a rotating generalization of the de Sitter vacuum solutions without a nonvacuum fluid content.

II. THE DYNAMICAL EQUATIONS

The closed Robertson–Walker metric may be written[2] as

$$ds^2 = dt^2 - R^2[\cos^2\theta\, d\psi^2 + d\theta^2 + \sin^2\theta\, d\varphi^2] \tag{2.1}$$

in spheroidal polar coordinates. Following Vaidya and Patel we introduce the coordinate transformation

$$d\beta = d\varphi + d\psi, \tag{2.2a}$$

$$du = (1/R)dt - d\psi, \tag{2.2b}$$

$$\alpha = 2\theta. \tag{2.2c}$$

Then one achieves

$$\begin{aligned} ds^2 = {} & 2R\,(du + \sin^2(\alpha/2)d\beta\,)dt \\ & - R^2[(du + \sin^2(\alpha/2)d\beta\,)^2 \\ & + \tfrac{1}{4}(d\alpha^2 + \sin^2\alpha\, d\beta^2)\,]. \end{aligned} \tag{2.3}$$

A non-Lorentz tetrad basis is introduced through

$$\theta^{(1)} = \mathbf{d}t + R\sin^2(\alpha/2)\mathbf{d}\beta - R\,\mathbf{d}\psi, \tag{2.4a}$$

$$\theta^{(2)} = (R/2)\mathbf{d}\alpha, \tag{2.4b}$$

$$\theta^{(3)} = (R/2)\sin\alpha\,\mathbf{d}\beta, \tag{2.4c}$$

$$\theta^{(4)} = \tfrac{1}{2}\mathbf{d}t - (R/2)\sin^2(\alpha/2)\mathbf{d}\beta + (R/2)\mathbf{d}\psi. \tag{2.4d}$$

[a)] Present address.

 0022-2488/88/061514-04$02.50

The metric in the θ basis takes the form

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

which we call non-Lorentz because of its nondiagonal form. In the θ basis the Ricci tensor has the following components:

$$R_{(11)} = -\frac{\ddot{R}}{2R} + \frac{1}{2}\left(\frac{\dot{R}}{R}\right)^2 + \frac{1}{2R^2}, \tag{2.5a}$$

$$R_{(14)} = -\frac{2\ddot{R}}{R} - \left(\frac{\dot{R}}{R}\right)^2 - \frac{1}{R^2}, \tag{2.5b}$$

$$R_{(22)} = R_{(33)} = \frac{\ddot{R}}{R} + 2\left(\frac{\dot{R}}{R}\right)^2 + \frac{2}{R^2}, \tag{2.5c}$$

$$R_{(44)} = -\frac{2\ddot{R}}{R} + 2\left(\frac{\dot{R}}{R}\right)^2 + \frac{2}{R^2}. \tag{2.5d}$$

The parentheses mean that the basis is the θ basis. The field equations of Einstein are

$$R_{\alpha\beta} - \tfrac{1}{2}Rg_{\alpha\beta} = 8\pi T_{\alpha\beta} + \Lambda g_{\alpha\beta}, \tag{2.6}$$

where the energy-momentum tensor is the sum of a perfect fluid term $(p+\rho)\,v_\alpha v_\beta - pg_{\alpha\beta}$, the energy-momentum tensor of the $U(1)$ gauge field $E_{\alpha\beta}$, and a flowing null radiation $\sigma w_\alpha w_\beta$, where σ is the density of the flowing null radiation. The energy-momentum tensor of the gauge field is given by

$$E_{\alpha\beta} = -g^{\delta\gamma}F_{\alpha\delta}F_{\beta\gamma} + \tfrac{1}{4}g_{\alpha\beta}F_{\delta\gamma}F^{\delta\gamma}, \tag{2.7a}$$

$$F_{\alpha\beta} = A_{\alpha,\beta} - A_{\beta,\alpha}, \tag{2.7b}$$

$$F^{\alpha\beta}{}_{;\beta} = 0. \tag{2.7c}$$

The four-velocity of the fluid is denoted by v^α, and the flow vector of the following null radiation is denoted by w^α. The vectors v^α and w^α satisfy

$$v^\alpha v_\alpha = 1, \quad w^\alpha w_\alpha = 0, \quad v^\alpha w_\alpha = 1. \tag{2.8}$$

In the θ basis the four-velocity is taken to be $v_{(\alpha)} = (1/2\lambda,0,0,\lambda)$, whereas the flowing null radiation has the four-velocity $w_{(\alpha)} = (1/\lambda,0,0,0)$.

From the field equations we then find

$$F_{(12)} = F_{(13)} = F_{(24)} = F_{(34)} = 0, \tag{2.9}$$

leaving only the terms $F_{(14)}$ and $F_{(23)}$ nonzero. The energy-momentum tensor must therefore have the nonvanishing components

$$E_{(22)} = E_{(33)} = E_{(14)} = \tfrac{1}{2}((F_{(23)})^2 + (F_{(14)})^2). \tag{2.10}$$

Since the energy density of radiation should decay as R^{-4} we may write

$$E_{(33)} = b^2/(2R^4), \tag{2.11}$$

where b is a constant. Then using the Maxwell equations, we find

$$F_{(14)} = \dot{A}/R \quad \text{and} \quad F_{(23)} = -2A/R^2$$

leading to the equation of the vector potential,

$$A' = A/R, \tag{2.12a}$$

whereas the rest of the contravariant components are zero in the coordinated basis. This A satisfies

$$\dot{A}^2 + 4A^2/R^2 = b^2/R^2. \tag{2.12b}$$

The gravitational field equations in the tetrad basis may be written as

$$R_{(\alpha\beta)} = 8\pi[(\rho+p)v_{(\alpha)}v_{(\beta)} - \tfrac{1}{2}(\rho-p)g_{(\alpha\beta)}] - \Lambda g_{(\alpha\beta)} + 8\pi E_{(\alpha\beta)} + 8\pi\sigma w_{(\alpha)}w_{(\beta)}. \tag{2.13}$$

From these equations one gets the following relations deduced in the paper of Vaidya and Patel (due to the difference in the definitions of R and a conformal difference in time coordinates the equations look slightly different):

$$8\pi P = \Lambda - \frac{4\pi b^2}{R^4} - \left(2\frac{\ddot{R}}{R} + \left(\frac{\dot{R}}{R}\right)^2 + \frac{1}{R^2}\right), \tag{2.14a}$$

$$8\pi\rho = -\Lambda - \frac{12\pi b^2}{R^4} + 3\left(\frac{\dot{R}}{R}\right)^2 + \frac{3}{R^2}, \tag{2.14b}$$

$$\lambda^2 = \left(1 - \frac{8\pi b^2/R^2}{1+\dot{R}^2 - R\ddot{R}}\right)^{-1}, \tag{2.14c}$$

$$8\pi\sigma = \frac{8\pi b^2}{R^2}\left[\frac{1+\dot{R}^2 - R\ddot{R} - 4\pi b^2/R^2}{1+\dot{R}^2 - R\ddot{R} - 8\pi b^2/R^2}\right], \tag{2.14d}$$

whereas the vorticity scalar of the perfect fluid is

$$\omega = (\lambda^2 - 1)/R\lambda. \tag{2.14e}$$

The motion of the fluid is geodesic only if the vorticity is vanishing. If there is vorticity, observers comoving with the fluid will experience a pressure gradient in the $U(1)$ gauge field pressure $4\pi b^2/R^4$, causing a nongeodesic motion.

III. KINEMATICAL PROPERTIES OF THE MODEL

From the field equations one may readily derive the Friedmann equations:

$$\dot{R}^2 = \frac{\Lambda}{3}R^2 + \frac{8\pi\rho R^2}{3} + \frac{4\pi b^2}{R^2} - 1, \tag{3.1}$$

$$\ddot{R} = \frac{\Lambda}{3}R - \frac{4\pi}{3}(\rho + 3P)R - \frac{4\pi b^2}{R^3}. \tag{3.2}$$

In the rest of the paper, we shall discuss a model in which the rotating fluid is radiation with equation of state $P = \frac{1}{3}\rho$. The results obtained will depend on this assumption. The energy density of radiation will be proportional to R^{-4}. Using that $P = \frac{1}{3}\rho$ one finds that the b^2 term acts as an additional radiation term in Eqs. (3.1) and (3.2). The first-order equation (3.1) may be integrated to give solutions of the classes M_1, M_2, A_1, A_2, S_1, and O_1, which were classified and studied by Stabell and Refsdal[9] and by Harrison.[10] These solutions have been discussed in the literature, and we shall not repeat this discussion here.

The four-velocity of the fluid in the θ-tetrad basis is

$$\mathbf{v} = (1/2\lambda)\boldsymbol{\theta}^{(1)} + \lambda\,\boldsymbol{\theta}^{(4)}. \tag{3.3}$$

We may introduce a comoving tetrad basis $\boldsymbol{\sigma}^\alpha$, where $\mathbf{v} = \boldsymbol{\sigma}^0$, through the transformation

$$\boldsymbol{\sigma}^0 = (1/2\lambda)\boldsymbol{\theta}^{(1)} + \lambda\,\boldsymbol{\theta}^{(4)}, \tag{3.4a}$$

$$\boldsymbol{\sigma}^1 = \boldsymbol{\theta}^{(2)}, \tag{3.4b}$$

$$\boldsymbol{\sigma}^2 = \boldsymbol{\theta}^{(3)}, \tag{3.4c}$$

$$\boldsymbol{\sigma}^3 = (1/2\lambda)\boldsymbol{\theta}^{(1)} - \lambda\,\boldsymbol{\theta}^{(4)}. \tag{3.4d}$$

Expansion, vorticity, and shear may be expressed in terms of the structure coefficients of the comoving tetrad basis $\boldsymbol{\sigma}^\alpha$. The structure coefficients are defined through the relation

$$\mathbf{d}\boldsymbol{\sigma}^\alpha = \tfrac{1}{2}C^\alpha{}_{\mu\nu}\boldsymbol{\sigma}^\mu \wedge \boldsymbol{\sigma}^\nu.$$

To calculate the structure coefficients we will have to know the time evolution of λ. In the case of the equation of state $P = \frac{1}{3}\rho$ we find that λ is constant. This may be concluded by substituting the expressions Eqs. (3.1) and (3.2) for $\dot{R}^2, \ddot{R}$ into the expression (2.14c) for λ. Then using that $\rho \sim R^{-4}$, we find that λ has the constant value

$$\lambda = (1 + 3b^2/2\rho_0)^{1/2}, \quad \rho_0 = \rho R^4. \tag{3.5}$$

We define the expansion tensor $\theta_{\mu\nu}$ as the symmetric part of the covariant derivative of the four-velocity of the cosmic fluid[11]:

$$\theta_{\mu\nu} = v_{(\mu;\nu)}, \tag{3.6}$$

where parentheses indicate the symmetric combination.

Using that the $\boldsymbol{\sigma}$ basis is comoving with the fluid and orthonormal, we find

$$\begin{aligned}\theta_{\mu\nu} &= v_{(\mu,\nu)} - \Gamma_{0(\mu\nu)}v^0 \\ &= -\Gamma_{0(\mu\nu)} = \tfrac{1}{2}(C_{\mu 0\nu} + C_{\nu 0\mu}),\end{aligned} \tag{3.7}$$

where $\Gamma_{0(\mu\nu)}$ are the symmetricized connection coefficients of the comoving $\boldsymbol{\sigma}$ basis field.

Since the structure constants in an orthonormal basis satisfy the relations $C^\mu{}_{0\nu} = -C^\nu{}_{0\mu}$, we deduce

$$\theta^\mu{}_\mu = C^\mu{}_{0\mu}, \quad \theta^\mu{}_\nu = 0, \quad \mu \neq \nu. \tag{3.8}$$

In the present model, the expansion is equal in all three spatial directions. Thus we find

$$\theta = \theta^\mu{}_\mu = C^1_{01} + C^2_{02} + C^3_{03} = \frac{3}{2}\frac{\dot{R}}{R}\left(\frac{1}{\lambda} + \lambda\right). \tag{3.9}$$

Since the expansion is isotropic, the observers comoving with the rotating fluid will not observe any shear. This may easily be verified using the definition of the shear tensor as the traceless part of the expansion tensor. We write

$$\sigma_{\mu\nu} = \theta_{\mu\nu} - \tfrac{1}{3}\theta h_{\mu\nu}, \tag{3.10}$$

where $h_{\mu\nu}$ is the projection tensor $h_{\mu\nu} = g_{\mu\nu} - v_\mu v_\nu$. Hence the shear scalar $\sigma^2 = \frac{1}{2}\sigma_{\mu\nu}\sigma^{\mu\nu}$ is given by

$$\begin{aligned}\sigma^2 = \tfrac{1}{3}[&(C^1_{01})^2 + (C^2_{02})^2 + (C^3_{03})^2 \\ &- C^1_{01}C^2_{02} - C^1_{01}C^3_{03} - C^2_{02}C^3_{03}],\end{aligned} \tag{3.11}$$

which is zero in the present model. An analogous argument can be carried through for the vorticity tensor, defined as the antisymmetric part of the covariant derivative of the velocity vector, to find a general expression for the components of the vorticity vector in terms of the structure coefficients of the fluid-comoving basis. Thus one finds that

$$\omega_{\mu\nu} = \tfrac{1}{2}C^0{}_{\nu\mu}. \tag{3.12}$$

The structure coefficient C^0_{12} gives vorticity in the three-direction. As both C^0_{13} and C^0_{23} are zero, the vorticity scalar is

$$\omega = C^0_2 = \frac{1}{R}\left(\lambda - \frac{1}{\lambda}\right) = \frac{\lambda^2 - 1}{R\lambda}, \tag{3.13}$$

which is Eq. (2.14e). The Robertson–Walker coordinate-comoving tetrad basis $\boldsymbol{\omega}^\alpha$ (the Lorentz basis in which the time coordinate t is the proper time) is deduced with the aid of Eqs. (2.4). Demanding a Lorentz basis with $\mathbf{d}t = \boldsymbol{\omega}^0$, we find

$$\boldsymbol{\omega}^0 = \boldsymbol{\theta}^{(4)} + \tfrac{1}{2}\boldsymbol{\theta}^{(1)}, \tag{3.14a}$$

$$\boldsymbol{\omega}^1 = \boldsymbol{\theta}^{(4)} - \tfrac{1}{2}\boldsymbol{\theta}^{(1)}, \tag{3.14b}$$

$$\boldsymbol{\omega}^2 = \boldsymbol{\theta}^{(2)}, \tag{3.14c}$$

$$\boldsymbol{\omega}^3 = \boldsymbol{\theta}^{(3)}. \tag{3.14d}$$

Because both this and the $\boldsymbol{\sigma}$ basis are of the Lorentz type (the metric is locally Minkowski), there is Lorentz transformation between the two systems $\boldsymbol{\sigma}$ and $\boldsymbol{\omega}$:

$$\boldsymbol{\sigma}^0 = \gamma(\boldsymbol{\omega}^0 - u\boldsymbol{\omega}^1), \tag{3.15a}$$

$$\boldsymbol{\sigma}^3 = \gamma(-\boldsymbol{\omega}^1 + u\boldsymbol{\omega}^0). \tag{3.15b}$$

Using Eqs. (3.4a)–(3.4d) and (3.14a)–(3.14d) to express these equations in terms of the $\boldsymbol{\theta}$ basis, we find that the Lorentz factor is given by

$$\gamma = \tfrac{1}{2}(1/\lambda + \lambda). \tag{3.16}$$

This is the time dilation factor between the fluid-comoving basis and the coordinate-comoving basis. This difference should be taken into account when discussing the time scale of transition into an inflationary era.

The vacuum Λ term can be regarded as the first approximation of a homogeneous scalar field for which the potential dominates the energy-momentum tensor. Such scalar fields are called inflation fields because they can be the driving force for inflation.

Let $\varphi = \varphi(t)$ be a complex inflation field. We define a gauge-covariant derivative

$$\bar{D}_\mu = D_\mu - igA_\mu, \tag{3.17}$$

where D_μ is the ordinary covariant derivative, g is a coupling constant, and A_μ is the $U(1)$ gauge vector. The Lagrangian of this charged inflation field is then

$$\mathscr{L} = (\bar{D}^\mu\varphi)^\dagger(\bar{D}_\mu\varphi) - V(\varphi,\varphi^\dagger). \tag{3.18}$$

We define the four-velocity of the inflation fluid (false vacuum fluid) to be proportional to the current $j^\mu = -(1/g)(\partial\mathscr{L}/\partial A_\mu)$,

$$j^\mu = i[\varphi^\dagger D^\mu\varphi - (D^\mu\varphi)^\dagger\varphi] - 2gA^\mu\varphi^\dagger\varphi. \tag{3.19}$$

Since the A vector is directed in the coordinate time direction, Eq. (2.12a), we find that the current form $\mathbf{j}$ is proportional to $\mathbf{d}t$ and thus

$$\mathbf{u} = \mathbf{d}t, \tag{3.20}$$

which is a four-velocity field without vorticity.

If the inflation field is real there is no current, but we may follow Belinskii and Khalatnikov,[12] and define a four-velocity of the inflation fluid as

$$u_\mu \equiv P_\mu (P_\alpha P^\alpha)^{-1/2}, \tag{3.21}$$

where P_α is the conjugate momentum of the inflation field. Using the Lagrangian

$$\mathscr{L} = \tfrac{1}{2} D_\mu \varphi D^\mu \varphi - V(\varphi), \tag{3.22}$$

we find

$$P_\alpha = \frac{\partial \mathscr{L}}{\partial (D^\alpha \varphi)} = D_\alpha \varphi, \tag{3.23}$$

which again leads to Eq. (3.20). During the decay of the false vacuum the rotating fluid will be diluted by a huge factor by this nonrotating fluid. The coupling to a hypersurface orthogonal gauge vector field does not affect the nonrotating nature of the inflation field.

IV. ASYMPTOTIC BEHAVIOR

The parameter $A(t)$ determining the gauge vector potential through Eq. (2.12a) satisfies Eq. (2.12b). In the asymptotic region where R grows exponentially, A satisfies

$$e^{2ht}\dot{A}^2 + 4A^2 = b^2. \tag{4.1}$$

The general solution of this asymptotic equation is

$$A = \frac{b}{2} \sin\left(\arcsin\left(\frac{2A_0}{b}\right) + \frac{2}{H}(e^{-Ht} - e^{-Ht_0})\right). \tag{4.2}$$

For large times A approaches the constant value

$$A_\infty = \frac{b}{2} \sin\left(\arcsin\left(\frac{2A_0}{b}\right) - \frac{2}{H} e^{-Ht_0}\right). \tag{4.3}$$

where A_0 is the value of A at t_0; the beginning of the inflationary era. In the comoving basis of the rotating fluid, the σ basis, the field tensor $F_{\mu\alpha}$ takes the form

$$F_{\alpha\beta} = \begin{bmatrix} 0 & 0 & 0 & -\dot{A}/R \\ 0 & 0 & -2A/R^2 & 0 \\ 0 & 2A/R^2 & 0 & 0 \\ \dot{A}/R & 0 & 0 & 0 \end{bmatrix}. \tag{4.4}$$

The components of the gauge vector (2.12) as well as the components of the field tensor $F_{\mu\alpha}$ will decay exponentially during inflation.

Using Eq. (3.13) the vorticity is also seen to decay exponentially during inflation:

$$\omega = \omega_0 e^{-Ht}. \tag{4.5}$$

V. CONCLUDING REMARKS

Until the advent of the inflationary cosmological models, the very slow rate of cosmic rotation was explained by invoking Mach's principle. Ellis and Olive[13] pointed out that inflation could solve the problem of slow cosmic rotation. A general discussion of inflation in anisotropic space-times can be found in Rothman and Ellis.[14]

In the present model vorticity and the energy density of the source-free $U(1)$ gauge field are intimately connected. With this in mind, it can be understood that vorticity does not speed up expansion in this model as the energy density of the gauge field will also increase when ω is increased.

Through a time dilation effect with respect to the coordinate time, the rotating fluid observers will find a greater cosmic age than that of the coordinate time. For these observers, the transition into an inflationary era takes more time than in the nonrotating coordinate frame.

It should also be noted that the energy density of the $U(1)$ gauge field and the vorticity of the cosmic fluid will have an exponential decay during inflation. The amount of decay depends on the total expansion during inflation. To solve the flatness and horizon problems an expansion of the order 10^{28} is required. Thus the minimal decay of vorticity is of the order 10^{-28}. In addition the nonrotation of the vacuum fluid should be taken into account, as this will introduce a further diluting effect that reduces vorticity by a factor of the order 10^{-112} during the reheating period.[15]

The effect of vorticity on the microwave background was analyzed by Collins and Hawking[16] and by Barrow *et al.*,[17] who found that the present rotation period must be more than 3×10^5 T_H, where T_H is the Hubble time ($1\text{–}2\times10^{20}$ years) in the case $k = 0$. Taking into account the diluting effect we find that the present period of rotation will be much greater than the observed limit even for minimal inflation. The huge rotation period of the universe may be a test of the inflationary hypothesis, if an observed lower limit to the period could be established.

[1] L. Bianchi, Mem. Mat. Fis. Soc. Italiana Sci. **11**, 267 (1898).
[2] P. C. Vaidya and L. K. Patel, Gen. Relativ. Gravit. **16**, 355 (1984).
[3] L. K. Patel and B. Pandya, Ind. J. Pure Appl. Math. **17**, 1224 (1986).
[4] P. C. Vaidya, Pramana J. Phys. **25**, 513 (1985).
[5] Ö. Grön, Phys. Rev. D **33**, 1204 (1986).
[6] R. T. Jantzen, J. Math. Phys. **27**, 2748 (1986).
[7] D. Brill and F. Flaherty, Ann. Inst. H. Poincaré A **28**, 335 (1978).
[8] H. H. Soleng, Astrophys. Space Sci. **137**, 373 (1987).
[9] R. Stabell and S. Refsdal, Mon. Not. R. Astron. Soc. **132**, 379 (1966).
[10] E. R. Harrison, Mon. Not. R. Astron. Soc. **137**, 69 (1967).
[11] G. F. R. Ellis, "Relativistic cosmology," in *General Relativity and Cosmology*, Proceedings of the International School of Physics "Enrico Fermi" course 47, edited by R. K. Sachs (Academic, New York, 1971), pp. 106ff.
[12] V. A. Belinskii and I. M. Khalatnikov, Sov. Phys. JETP **36**, 591 (1973).
[13] J. Ellis and K. A. Olive, Nature **303**, 679 (1983).
[14] T. Rothman and G. F. R. Ellis, Phys. Lett. B **180**, 19 (1986).
[15] Ö. Grön and H. H. Soleng, Nature **328**, 501 (1987).
[16] C. B. Collins and S. W. Hawking, Mon. Not. R. Astron. Soc. **162**, 307 (1973).
[17] J. D. Barrow, R. Juszkiewicz, and D. H. Sonoda, Mon. Not. R. Astron. Soc. **213**, 917 (1985).

Erratum: Canonical formalism of dissipative fields in the thermo field dynamics [J. Math. Phys. 28, 2741 (1987)]

T. Arimitsu
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan
H. Umezawa and Y. Yamanaka
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton T6G 2J1, Canada

(Received 14 January 1988; accepted for publication 27 January 1988)

At the end of Sec. V, p. 2751, left column, it is stated that $\kappa(\mathbf{k}) = 0$ is one of the solutions of the self-consistent equations (5.11). But the solution $\kappa(\mathbf{k}) = 0$ is not allowed for the following reason. When we perform our perturbation calculation with the free fields without dissipation ($\kappa = 0$), the nonvanishing imaginary part of the self-energy is naturally created even at one-loop level through the decay process of a nontilde particle into another nontilde particle and a tilde one. What was wrong in the paper resides in the dispersion formula (5.13c) for the renormalized energy. Since the Lorentz symmetry is broken in thermal situations, we should not assume the relativistic form of the dispersion like (5.13c).

 0022-2488/88/061518-01$02.50