Irreducible *-representations of Lie superalgebras $B(0,n)$ with finite-degenerated vacuum

J. Blank and M. Havlíček
Nuclear Centre of the Charles University, Prague, V Holešovičkách 2, 180 00 Praha 8, Czechoslovakia

(Received 3 September 1985; accepted for publication 14 August 1986)

The problem of getting irreducible *-representations π of Lie superalgebras $B(0,n)$, $n = 1,2$, is studied, starting with a recently constructed family of linear representations in terms of differential operators on the space C_N^∞ of $\mathbb{C}^N$-valued C^∞-functions. Equivalent formulation via creation-annihilation operators of a para-Bose system with n degrees of freedom is used, and the domain $\mathscr{D}$ of any π is shown to be a subset of C_N^∞ containing a nonzero vacuum subspace. By assuming its dimension finite, the necessary conditions for existence of π are derived. The method is applied to the superalgebra $B(0,1)$ and a one-parameter family Π of nonequivalent irreducible *-representations in terms of unbounded linear operators on $L^2(\mathbb{R}^+)\otimes\mathbb{C}^2$ is obtained. Each representation $\pi\in\Pi$ has a nondegenerated vacuum and for all $z\in B(0,1)$ satisfying $z = z^*$, the operators $\pi(z)$ are essentially self-adjoint.

I. INTRODUCTION

Recently, we have constructed for the Lie superalgebras $\mathrm{osp}(1,2n)$, $n = 1,2$, families of infinite-dimensional linear representations $z\mapsto\Omega^{(n)}(z)$.[1–3] The representations $\Omega^{(1)}$ form a family depending on one parameter that can assume any real value. The family of representations $\Omega^{(2)}$ is labeled by parameters N and κ: $N = 2,4,\ldots$, and κ takes values in some $\mathscr{K}_N\subset\mathbb{R}$. For each $z\in\mathrm{osp}(1,2n)$, the operator $\Omega^{(n)}(z)$ is a linear differential operator on $C_N^\infty(M_n)$—the space of C^∞ vector functions Φ: $M_n\ni x\mapsto\Phi(x)\in\mathbb{C}^N$, where $M_1 := \mathbb{R}^+$ ($N = 2$ for $n = 1$), and M_2 is some open subset of $\mathbb{R}^+\times\mathbb{R}^2$.

Under $\mathrm{osp}(1,2n)$, we understand the unique real form of the complex Lie superalgebra (LSA) $B(0,n)$. This LSA is generated by $2n$ odd elements y_l, $l = \pm 1,\ldots,\pm n$; their symmetric products determine $n(2n+1)$ independent even elements

$$x_{jk} := \tfrac{1}{2}\langle y_j, y_k\rangle = x_{kj}, \tag{1.1a}$$

and the law of multiplication reads

$$\langle x_{jk}, y_l\rangle := g_{jl}\,y_k + g_{kl}\,y_j,\quad g_{jl} := \mathrm{sgn}(j)\delta_{j+l}. \tag{1.1b}$$

By using the Jacobi identity, one gets

$$\langle x_{jk}, x_{lm}\rangle = g_{jl}x_{km} + g_{jm}x_{kl} + g_{kl}x_{jm} + g_{km}x_{jl}. \tag{1.1c}$$

The basis $\{x_{jk}, y_l\colon j,k,l = \pm 1,\ldots,\pm n\}$ will be called Racah as its even part is identical with the basis of $\mathrm{sp}(2n,\mathbb{R})$.[4]

For discussing properties of representations Ω, it is convenient to regard the space Λ_N of linear differential operators on $C_N^\infty(M)$ as an associative *-algebra equipped with adjoint operation "#",[5] and introduce the standard LSA structure[6] on its linear subspace $\mathscr{A} := \mathscr{A}_0\oplus\mathscr{A}_1$ which is determined via the Racah basis as

$$\mathscr{A}_0 := \{\Omega(x_{jk})\colon j,k = \pm 1,\ldots,\pm n\}_{\mathrm{lin}},$$
$$\mathscr{A}_1 := \{\Omega(y_l)\colon l = \pm 1,\ldots,\pm n\}_{\mathrm{lin}}.$$

Basic features of the representations Ω can now be summarized as follows.

(i) Each Ω is a homomorphism of $\mathrm{osp}(1,2n)$ on $\mathscr{A}$, the order of the differential operator $\Omega(z)$ being at most 1 if z is an odd element and at most 2 if z is even.

(ii) Let $z\mapsto z^*$ be the involution on $B(0,n)$[7] defined as the antilinear extension of

$$x_{jk}^* := -x_{jk},\quad y_l^* := -iy_l\quad i\equiv\sqrt{-1}. \tag{1.2}$$

Then one has[5]

$$\Omega(z^*) = \Omega(z)^{\#}, \tag{1.3}$$

i.e., Ω is a *-homomorphism. In particular, for elements of the even subalgebra, $\mathrm{sp}(2n,\mathbb{R})$ holds $x^* = -x$, so that $\Omega(x)$ is skew-symmetric:

$$\Omega(x)^{\#} = -\Omega(x),\quad x\in\mathrm{sp}(2n,\mathbb{R}).$$

(iii) All independent Casimir elements of $\mathrm{osp}(1,2n)$ (there is just one for $n = 1$ and two for $n = 2$) are represented by multiples of unity in Λ_N (Schur-irreducibility), the parameters that label Ω being in one-to-one correspondence with these numbers.

The main purpose of this paper is getting algebraically irreducible[8] Hilbert-space *-representations from the linear representations $\Omega^{(n)}$. This is achieved by restricting suitably the representation space of $\Omega^{(n)}$; the procedure is described in detail in Secs. II and III, its essential feature being the requirement that the representations we construct have finite-degenerated vacuum [see Eqs. (2.5)–(2.6)].

In Sec. IV the procedure is applied to the family $\{\Omega^{(1)}\} = \{\Omega_\kappa\colon\kappa\in\mathbb{R}\}$ of linear representations of $B(0,1)$ on the vector space $C_2^\infty(\mathbb{R}^+)$ giving the following results.

(a) A one-parameter family Π of nonequivalent irreducible *-representations of $B(0,1)$ in terms of unbounded operators on $L^2(\mathbb{R}^+)\otimes\mathbb{C}^2$ was obtained. Each representation $\pi = \pi_\kappa\in\Pi$ equals $\Omega_\kappa\restriction\mathscr{D}_\kappa$ for some $\kappa\in(-1/2,\infty)\setminus\{0\}$, $\mathscr{D}_\kappa$ being an Ω_κ-invariant subspace of $C_2^\infty(\mathbb{R}^+)$ such that $\overline{\mathscr{D}}_\kappa = L^2(\mathbb{R}^+)\otimes\mathbb{C}^2$; in addition, π_κ has a nondegenerated vacuum.

(b) The family Π is complete in the following sense: if

 0022-2488/86/122823-09$02.50

$\kappa' \in \mathbb{R} \setminus \{0\}$, $\mathscr{D}'$ is a subspace of $C_2^\infty(\mathbb{R}^+)$ whose intersection with the vacuum subspace is nontrivial, and if $\pi' = \Omega_{\kappa'} \upharpoonright \mathscr{D}'$ is an irreducible *-representation of $B(0,1)$, then π' is equivalent to some $\pi \in \Pi$.

(c) For each $\pi \in \Pi$ and all elements $z \in B(0,1)$ satisfying $z = z^*$ the operators $\pi(z)$ are essentially self-adjoint on $\mathscr{D}$. Particularly, this holds for $z = ix_{jk}$ and $z = \exp(-i\pi/4) y_k$, where $\{x_{jk}, y_k : j,k = \pm 1\}$ is the Racah basis of $B(0,1)$.

(d) If π is restricted to $\mathrm{sp}(2,\mathbb{R}) \sim \mathrm{sl}(2,\mathbb{R})$, i.e., to the even subalgebra of the unique real form $\mathrm{osp}(1,2)$ of $B(0,1)$, a skew-symmetric representation of $\mathrm{sl}(2,\mathbb{R})$ is obtained that equals direct sum of two irreducible skew-symmetric representations of $\mathrm{sl}(2,\mathbb{R})$ on $L^2(\mathbb{R}^2)$. Each of these representations is integrable to a unitary irreducible representation of the universal covering group of $\mathrm{SL}(2,\mathbb{R})$.

II. FORMULATION OF THE PROBLEM

Let $\mathscr{H}$ be a Hilbert space and $\mathscr{D}$ a dense subspace of $\mathscr{H}$. A set $\{a_j, \hat{a}_j^\dagger : j = 1,2,\ldots,n\} \subset \mathrm{End}_{\mathscr{H}} \mathscr{D}$ (see Appendix for definitions) is called pB_n-set with domain $\mathscr{D}$ if

$$[\{\hat{a}_j, \hat{a}_k\}, \hat{a}_l] = 0, \tag{2.1a}$$

$$[\{\hat{a}_j, \hat{a}_k^\dagger\}, \hat{a}_l] = -2\delta_{k-l} \hat{a}_j. \tag{2.1b}$$

The $\hat{a}_j$ ($\hat{a}_j^\dagger$) is interpreted as the jth mode annihilation (creation) operator and

$$\hat{n}_j := \tfrac{1}{2}\{\hat{a}_j, \hat{a}_j^\dagger\} \tag{2.2}$$

as the jth mode particle-number operator of a para-Bose system with n degrees of freedom.[9]

A simple example of a pB_n-set is provided by annihilation and creation operators of usual bosons that satisfy the canonical commutation relation (CCR). It is known that for each $n = 1,2,\ldots,$ such a set for which the operator $\hat{a}_1^\dagger \hat{a}_1 + \cdots + \hat{a}_n^\dagger \hat{a}_n$ is essentially self-adjoint (e.s.a.) is just one up to equivalences.[10] That is why only "non-trivial" solutions of Eqs. (2.1) are of interest, viz. those for which the CCR do not hold. The first example was given for $n = 1$ by Wigner[11]. Later on, Green discovered for arbitrary n, including $n = \infty$, an infinite set of solutions labeled by one integer $p = 1,2,\ldots,$ called "order," and Greenberg with Messiah[12] selected from among them the irreducible ones acting on a Fock space with a unique vacuum.

Our aim is obtaining irreducible pB_n-sets for $n = 1,2$ from the representations $\Omega^{(n)}$. This is possible, at least in principle, because pB_n-sets and representations of $B(0,n)$ are closely related. In order to get a formulation suitable for the purposes of this study, consider another basis $\{b_{jk}, a_l : j,k,l = \pm 1,\ldots, \pm n\}$ in $B(0,n)$ defined via the following map ζ:

$$a_1 \equiv \zeta(y_l) := 2^{-1/2}(y_1 - iy_{-l}), \tag{2.3a}$$

$$b_{jk} \equiv \zeta(x_{jk}) := \tfrac{1}{2}\langle a_j, a_k \rangle = \tfrac{1}{2}(x_{jk} - x_{-j-k}) - \tfrac{1}{2}(x_{-jk} + x_{k-j}). \tag{2.3b}$$

It can easily be verified that all the structure constants are identical with those of the Racah basis [see Eqs. (1.1)]; thus, ζ is an automorphism of $B(0,n)$. Moreover, one gets from (1.2)

$$a_l^* = a_{-l}, \quad b_{jk}^* = b_{-j-k}. \tag{2.3c}$$

By using these facts, one readily gets the sought interrelation of pB_n-sets and *-representations of $B(0,n)$.

Proposition 2.1: If π is an irreducible *-representation of $B(0,n)$ in terms of operators in $\mathrm{End}_{\mathscr{H}} \mathscr{D}$, then $\{\pi(a_j), \pi(a_j)^{\neq} : j = 1,2,\ldots,n\}$ is an irreducible pB_n set.[13]

In view of this assertion, the problem of constructing pB_n-sets from representations $\Omega \equiv \Omega^{(n)}$ can be formulated purely in the language of the representation theory as follows: Given a linear representation Ω on C_N^∞, find an Ω-invariant subspace $\mathscr{D} \subset C_N^\infty$ and introduce a scalar product on $\mathscr{D}$ such that the operators

$$\pi(z) := \Omega(z) \upharpoonright \mathscr{D}, \quad z \in B(0,n), \tag{2.4a}$$

form an irreducible representation of $B(0,n)$ on $\mathscr{H} := \overline{\mathscr{D}}$ and fulfill

$$\pi(z)^\dagger = \Omega(z)^\# \upharpoonright \mathscr{D}. \tag{2.4b}$$

This condition implies that

$$\hat{N} := \sum_{j=1}^{n} \hat{n}_j = \frac{1}{2} \sum_{j=1}^{n} \{\pi(a_j), \pi(a_j)^\dagger\} = \sum_{j=1}^{n} \pi(b_{j-j})$$

has to be a positive operator

$$\hat{N} \geqslant 0, \tag{2.4c}$$

and, due to its interpretation as particle-number operator, must have a nonempty point spectrum. Using the standard considerations based upon the relations $[\hat{N}, \hat{a}_j] = -\hat{a}_j$, $j = 1,\ldots,n$ [cf. Eq. (2.1b)], one concludes that the sought domain $\mathscr{D}$ must have a nontrivial intersection with the vacuum subspace

$$V_\Omega := \{\phi \in C_N^\infty : \tilde{a}_j \phi = 0, \quad j = 1,\ldots,n\}, \quad \tilde{a}_j \equiv \Omega(a_j). \tag{2.5}$$

The usual requirement of uniqueness of the vacuum will be replaced by a weaker condition

$$1 \leqslant \dim \mathscr{D} \cap V_\Omega < \infty, \tag{2.6}$$

since uniqueness, which is essential in the quantum field theory, is a too restrictive condition when systems with a finite number of degrees of freedom are concerned. Representations $\pi \equiv \Omega \upharpoonright \mathscr{D}$, which fulfill Eq. (2.6), are said to have finite-degenerated vacuum (FDV-representations). Let us recall in this context that for representations with a unique vacuum ϕ_0, one has[14]

$$\hat{a}_k \hat{a}_l^\dagger \Phi_0 = p \delta_{k-l} \phi_0, \quad p > 0,$$

where p is independent of k and is called the order (of parastatistics).[12] This relation is in general not valid if the uniqueness of vacuum is violated; in particular, the notion of order does not make sense for representations with degenerated vacuum.

III. GENERAL FEATURES OF THE CONSTRUCTION

We have seen in Sec. II that the problem involves finding an Ω-invariant subspace $\mathscr{D} \subset C_N^\infty$ such that conditions (2.4) and (2.6) are fulfilled. To this purpose we developed a method that can be divided in two steps. In the first one we derive general properties having the form of necessary conditions that follow from the assumption that for a given Ω, there exists a domain $\mathscr{D}$ with all the required properties. In this way, the original family of representations Ω is reduced by excluding all those Ω that do not fulfill the necessary conditions. The conditions themselves are obtained by examining the structure of the subspace

$$\mathscr{D}^{(\text{vac})} := \mathscr{D} \cap V_\Omega .$$

The second step is inductive; it deals with the construction of $\mathscr{D}$ starting with a fixed vector Ψ in the vacuum subspace V_Ω, this vector being fully specified by the above necessary conditions.

The starting point for analyzing the subspace $\mathscr{D}^{(\text{vac})}$ is provided by the following simple assertions.

Lemma 3.1: (i) The linear envelope of $\{b_{j-k}: j,k=1,...,n\}$ is a subalgebra of $\mathrm{sp}(2n,\mathbb{C})$ that is isomorphic to $\mathrm{gl}(n,\mathbb{C})$.

(ii) For each $u \in \mathrm{gl}(n,\mathbb{C})$, one has

$$\Omega(u) v_\Omega \subset V_\Omega . \tag{3.1}$$

Proof: The first statement is due to the fact that b_{j-k} satisfies the same commutation relations as the elements of the standard basis $\{\epsilon_{jk}\}$ of $\mathrm{gl}(n,\mathbb{C})$, if $\epsilon_{jk} \leftrightarrow b_{k-j}$. The relation (3.1) follows from $\langle b_{j-k}, a_l \rangle = -\delta_{k-l} a_j$ and Eq. (2.5). ■

Remark 3.2: Notice that the real linear envelope of $i(b_{j-k}+b_{k-j})$, $(b_{j-k}-b_{-jk})$, $j,k=1,...,n$, equals $u(n)$, which is isomorphic to $\mathrm{sp}(2n,\mathbb{R}) \cap \mathrm{sp}(2n)$, i.e., to the maximal compact subalgebra of $\mathrm{sp}(2n,\mathbb{R})$.

Corollary 3.3: Suppose that $\Omega \restriction \mathscr{D}$ is a FDV-representation; then $\mathscr{D}^{(\text{vac})}$ is a finite-dimensional subspace invariant under $\Omega(u)$, $u \in \mathrm{gl}(n,\mathbb{C})$.

In view of these properties, one can learn much about $\mathscr{D}^{(\text{vac})}$ by applying the theory of finite-dimensional representations of semisimple Lie algebras.[15] Consider the map ω:

$$\mathrm{sl}(n,\mathbb{C}) \ni u \mapsto \omega(u) := \Omega(u) \restriction V_\Omega ;$$

because of (3.1), ω is a representation of $\mathrm{sl}(n,\mathbb{C})$ on V_Ω. Now suppose that for a given Ω there is an Ω-invariant domain $\mathscr{D} \subset C_N^\infty$ such that $\Omega \restriction \mathscr{D}$ fulfills the conditions (2.4b) and (2.6). By the corollary,

$$\omega_{\text{vac}} := \omega \restriction \mathscr{D}^{(\text{vac})}$$

is a finite-dimensional representation of $\mathrm{sl}(n,\mathbb{C})$, and hence

$$\mathscr{D}^{(\text{vac})} = \sum_J{}^{\oplus} V_{\lambda(J)} . \tag{3.2}$$

Each $V_{\lambda(J)}$ is the representation space of an irreducible representation of $\mathrm{sl}(n,\mathbb{C})$ with the highest weight (HW) $\lambda(J)$. Consider the particle-number operator

$$\widetilde{N} := \sum_{j=1}^{n} \Omega(b_{j-j});$$

by Eq. (2.4c), $\widetilde{N}_{\text{vac}} := \widetilde{N} \restriction \mathscr{D}^{(\text{vac})}$ is a positive matrix. Let $\nu \geqslant 0$ be its eigenvalue and W_ν the subspace of eigenvectors corresponding to ν. As $[\omega_{\text{vac}}(u), \widetilde{N}_{\text{vac}}]=0$ for each $u \in \mathrm{sl}(n,\mathbb{C})$ (Lemma 3.1), the subspace W_ν is invariant under ω_{vac}, and thus W_ν equals direct sum of a subsystem of subspaces $V_{\lambda(J)}$ in (3.2). Consequently, to each subspace $V_{\lambda(J)}$ there is an eigenvalue ν of $\widetilde{N}_{\text{vac}}$ such that $V_{\lambda(J)} \subset W_\nu$. In particular, the corresponding HW-vector $\Psi_{\lambda(J)} \in V_{\lambda(J)}$ fulfills

$$\widetilde{N}\Psi_{\lambda(J)} = \nu \Psi_{\lambda(J)} . \tag{3.3}$$

By summarizing, we arrive at the following necessary conditions.

Proposition 3.4: Suppose that for a given Ω there is an Ω-invariant domain $\mathscr{D} \subset C_N^\infty$, such that $\Omega \restriction \mathscr{D}$ is an irreducible FDV-representation having the *-property (2.4b). Then there exist non-negative integers $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{n-1}$, a non-negative ν, and a nonzero $\Psi_\lambda \in \mathscr{D}^{(\text{vac})}$ such that Eq. (3.3) holds,[16]

$$\mathscr{D} = \mathscr{U}(\tilde{a}_1, \tilde{a}_1^\#, ..., \tilde{a}_n, \tilde{a}_n^\#)\Psi_\lambda , \tag{3.4a}$$

and relations

$$\omega(b_{j-j} - b_{j+1,-j-1})\Psi_\lambda = \lambda_j \Psi_\lambda , \tag{3.4b}$$

$$\omega(b_{j+1,-j})\Psi_\lambda = 0 \tag{3.4c}$$

are fulfilled for $j=1,2,...,n-1$. Moreover,

$$\omega_\lambda := \omega \restriction \mathscr{U}(\omega(b_{-21}), \omega(b_{-32}), ..., \omega(b_{-n,n-1}))\Psi_\lambda$$

is an irreducible representation of $\mathrm{sl}(n,\mathbb{C})$.

Since $\Omega(u)$ are linear differential operators, the conditions (3.3), (3.4b), and (3.4c), together with $\Psi_\lambda \in V_\Omega$ represent a system of partial differential equations (PDE). By demanding existence of a nontrivial solution, one gets conditions that relate ν and the integers λ_j to parameters labeling Ω (e.g., for $n=2$ there are two: N and κ—cf. Sec. I). The first step of our approach consists in finding for given $\nu \geqslant 0$ and non-negative integers $\lambda_1 \geqslant \cdots \geqslant \lambda_{n-1}$, the admissible values of these parameters, solving the PDE systems with these values and verifying irreducibility of ω_λ.

Let us pass to the second step. Suppose we found the function Ψ_λ that solves the PDE system for some $\lambda_1 \geqslant \cdots \geqslant \lambda_{n-1}$, $\nu \geqslant 0$, and some fixed representation Ω with admissible parameters. For the domain $\mathscr{D}_\lambda$ given by Eq. (3.4a), we have to check that $\pi_\lambda := \Omega \restriction \mathscr{D}_\lambda$ is irreducible and the FDV condition $1 \leqslant \dim \mathscr{D}_\lambda \cap V_\Omega < \infty$ is fulfilled. In addition, a scalar product must be introduced on $\mathscr{D}_\lambda$ such that (2.4b) will hold.

To this purpose we first try to make the structure of $\mathscr{D}_\lambda$ lucider by finding a basis $\mathscr{E} \subset \mathscr{D}_\lambda$ with some specific properties. Naturally, we demand

$$\Psi_\lambda \in \mathscr{E}. \tag{3.5}$$

As a direct consequence of Eqs. (3.3) and (3.4), one finds that Ψ_λ is a common eigenvector of $\tilde{n}_j := \Omega(b_{j-j})$, $j=1,...,n$:

$$\tilde{n}_j \Psi_\lambda = \nu_j \Psi_\lambda , \tag{3.6}$$

each eigenvalue ν_j being a simple function of ν and $\lambda_1,...,\lambda_{n-1}$. Further, the commutation relation (2.1b) implies for $p=1,2,...$:

$$\tilde{n}_j \tilde{a}_k^p = \tilde{a}_k^p(\tilde{n}_j - p\delta_{j-k}), \quad \tilde{n}_j(\tilde{a}_k^\#)^p = (\tilde{a}_k^\#)^p(\tilde{n}_j + p\delta_{j-k}). \tag{3.7}$$

A straightforward generalization together with Eq. (3.6) yields, for any monomial $\widetilde{M} \in \mathscr{U} := \mathscr{U}(\tilde{a}_1, \tilde{a}_1^\#, ..., \tilde{a}_n, \tilde{a}_n^\#)$,

$$\tilde{n}_j \widetilde{M}\Psi_\lambda = (\nu_j + r_j)\widetilde{M}\Psi_\lambda ,$$

$$r_j = 0, \pm 1, \pm 2, ..., \quad j = 1,2,...,n.$$

Thus, $\mathscr{D}_\lambda$ is spanned by functions $\Phi \in C_N^\infty$ labeled by integers $r_1,...,r_n$, each $\Phi \equiv \Phi_{r_1,...,r_n}$ fulfilling

$$\tilde{n}_j \Phi = (\nu_j + r_j)\Phi, \quad j=1,...,n. \tag{3.6$'$}$$

This is again a system of PDE's and we shall see that particular solutions can be found analytically for both the cases $n = 1,2$.

Let $\mathscr{E}$ be a linearly independent set of solutions to (3.6′) for some infinite system of n-tuples $(r_1,...,r_n)$ [notice that in view of Eqs. (3.5) and (3.6), $(0,...,0)$ must belong to the system] such that

$$\mathscr{U}\mathscr{E} \subset \mathscr{E}_{\text{lin}}. \tag{3.8a}$$

Then one clearly has

$$\mathscr{D}_\lambda \subset (\mathscr{U}\mathscr{E})_{\text{lin}} = \mathscr{E}_{\text{lin}}. \tag{3.8b}$$

Due to the following simple argument one can expect that Eq. (3.8a) will hold as soon as $\mathscr{E}$ is sufficiently large. For any $\Phi \equiv \Phi_{r_1...r_n} \in \mathscr{E}$ one gets from (3.6′) and (3.7)

$$\tilde{n}_j \tilde{a}_k \Phi = (\nu_j + r_j - \delta_{j-k}) \tilde{a}_k \Phi.$$

Comparing with Eq. (3.6′) suggests that $\tilde{a}_k \Phi$ should be identified with $\Phi_{r_1...r_j-1...r_n}$ and thus $\tilde{a}_k \Phi_{r_1...r_n} \in \mathscr{E}$ if $\Phi_{r_1...r_j-1...r_n} \in \mathscr{E}$, etc. In fact, in cases we will consider in the following, it is possible to choose $\mathscr{E}$ such that $\tilde{a}_k \mathscr{E} \subset \mathscr{E} \cup \{0\}$, $\tilde{a}_k^\# \mathscr{E} \subset \mathscr{E}$, i.e., the action of $\tilde{a}_k$ and $\tilde{a}_k^\#$ on any $\Phi \in \mathscr{E}$ is very simple.

As soon as this action is found, one can verify directly whether $\Omega \restriction \mathscr{E}_{\text{lin}}$ is irreducible. If it is so, then from Eq. (3.8b), it follows in view of invariance of $\mathscr{D}_\lambda$ under Ω

$$\mathscr{D}_\lambda = \mathscr{E}_{\text{lin}},$$

and hence $\mathscr{E}$ is a basis in $\mathscr{D}_\lambda$.

Having such a basis is very helpful in introducing a scalar product on $\mathscr{D}_\lambda$ obeying the *-condition (2.4b). This condition requires the operators $\hat{n}_j := \tilde{n}_j \restriction \mathscr{D}_\lambda$ to be symmetric and hence the scalar product must be chosen in such a way that $\mathscr{E}$ becomes an orthogonal set. This can always be done: suppose $\mathscr{E} = \{\Phi_r\}_{r=1}^\infty$ (for simplicity the elements of $\mathscr{E}$ are labeled by a single index) and let

$$(\Phi_r, \Phi_s) := t_r \delta_{r-s}, \quad r,s = 1,2,...,t_r > 0.$$

The Hilbert space $\mathscr{H} = \mathscr{D}_\lambda = \mathscr{E}_{\text{lin}}$ then consists of all functions

$$\Phi \equiv \sum_{r=1}^\infty c_r \Phi_r \quad \text{such that} \quad \sum_{r=1}^\infty t_r |c_r|^2 < \infty.$$

However, this choice does not guarantee that (2.4b) is fulfilled. Let us discuss this point in more detail.

First of all, one can replace Eq. (2.4b) by a simpler condition

$$(\tilde{a}_j \Phi, \Phi) = (\Phi, \tilde{a}_j^\# \Phi), \quad 1 \leqslant j \leqslant n, \quad \Phi \in \mathscr{D}_\lambda, \tag{3.9}$$

since each operator $\Omega(z)$, $z \in B(0,n)$, equals a linear or quadratic function of $\tilde{a}_j, \tilde{a}_j^\#$. All the operators $\tilde{a}_j$ can be expressed as (see Ref. 5 and Sec. I)

$$\tilde{a}_j \equiv \tilde{a} = f_0 + \sum_{k=1}^m f_k p_k.$$

Then by the definition of the #-operation,[5] one finds for any $\Phi \in C_N^\infty \equiv C_N^\infty(M)$, $M \subset \mathbf{R}^m$

$$(\tilde{a}\Phi) \times \Phi - \Phi \times (\tilde{a}^\# \Phi) = \sum_{k=1}^m p_k [(f_k \Phi) \times \Phi],$$

where for $\Phi, \Psi \in C_N^\infty(M)$, $\Phi \equiv (\varphi,...,\varphi_N)$, $\Psi \equiv (\psi_1,...,\psi_N)$, $\varphi_r, \psi_r \in C^\infty(M)$

$$\Phi \times \Psi := \sum_{r=1}^N \varphi_r \bar{\psi}_r.$$

By the Gauss theorem Eq. (3.9) is fulfilled if $\mathscr{D}_\lambda$ is a subspace in

$$\mathscr{H} \equiv \sum_{r=1}^{N} {}^{\oplus} L^2(M), \tag{3.10a}$$

and if all components of each $\Phi \in \mathscr{D}_\lambda$ vanish on the boundary of M. Of course, this requirement is only sufficient for (3.9) and in case it were not compatible with the previous conditions imposed upon $\mathscr{D}_\lambda$, one could try another choice of $\mathscr{H}$. However, verifying the condition (3.9) would then probably be difficult.

According to the general definition, the complete specification of each of the sought representations $\Omega \restriction \mathscr{D}$ includes also a projection $\hat{E}$ on the Hilbert space $\mathscr{H}$ that determines the decomposition $\text{End}_{\mathscr{H}} \mathscr{D} = (\text{End}_{\mathscr{H}} \mathscr{D})_0 \oplus (\text{End}_{\mathscr{H}} \mathscr{D})_1$—cf. Appendix, Eqs. (A2)–(A4). Assume that

$$\mathscr{H} = \sum_{r=1}^{N} {}^{\oplus} \mathscr{G}, \tag{3.10b}$$

$\mathscr{G}$ being a Hilbert space to which belong all components of each $\Phi \in \mathscr{D}$. Then the projection $\hat{E}$ can be chosen as

$$\hat{E} := I_{\mathscr{G}} \otimes E, \quad E := \begin{pmatrix} I_{N/2} & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.11}$$

The choice is implied by the structure of the operators $\Omega(z)$.[1–3] The point is that these operators are expressed in terms of two finite subsets $\mathfrak{M}_0, \mathfrak{M}_1 \subset \text{End}\, \mathbf{C}^N$, and of scalar linear differential operators ξ_α acting on $C^\infty(M)$:

$$\Omega(z) = \sum_\alpha \xi_\alpha \otimes T_\alpha,$$

in such a way that $T_\alpha \in \mathfrak{M}_0$ if z is an even element of $B(0,n)$ and $T_\alpha \in \mathfrak{M}_1$ if z is odd. In addition, the projection $E \in \text{End}\, \mathbf{C}^N$ satisfies $ET_\alpha = T_\alpha E$ if $T_\alpha \in \mathfrak{M}_0$ and $ET_\alpha = T_\alpha (I - E)$ if $T_\alpha \in \mathfrak{M}_1$. It then follows from (3.11) that $\Omega \restriction \mathscr{D}$ will map even and odd elements of $B(0,n)$ in $(\text{End}_{\mathscr{H}} \mathscr{D})_0$ and $(\text{End}_{\mathscr{H}} \mathscr{D})_1$, respectively.

On the other hand, fixing $\hat{E}$ by Eq. (3.11) imposes an additional condition upon the structure of the domain $\mathscr{D}$:

$$\Phi \equiv (\varphi_1,...,\varphi_N) \in \mathscr{D} \Rightarrow (\varphi_1,...,\varphi_{N/2},0,...,0) \in \mathscr{D}$$

[cf. Eq. (A2)]. This will hold, e.g., if

$$\mathscr{D} = \sum_{r=1}^{N} {}^{\oplus} D_r; \tag{3.12}$$

notice that because of Eq. (3.10b) all D_r must be dense in $\mathscr{G}$.

It appears that for representations $\Omega^{(n)}$, $n = 1,2$, of Refs. 1–3, a domain $\mathscr{D}$ can always be found such that (3.12) holds. This is due to the structure of the operators $\tilde{n}_j = \Omega(b_{j-j})$ that allows for decoupling of the system (3.6′) for vector functions $\Phi \equiv (\varphi_1,...,\varphi_N)$ into N-independent systems of n partial differential equations for individual components φ_r. For getting the basis $\mathscr{E}$, one chooses linearly independent sets $\mathscr{E}_r$ of solutions for the rth component and puts

$$\mathscr{E} := \bigcup_{r=1}^N (0,...,\mathscr{E}_r,...,0).$$

This one has

$$\mathscr{E}_{\text{lin}} = \sum_{r=1}^{N}{}^{\oplus} D_r, \quad D_r := (\mathscr{E}_r)_{\text{lin}}.$$

IV. RESULTS FOR $B(0,1)$

A. Specific features of the case $n=1$

The operators $\Omega^{(1)}(z) \equiv \Omega_\kappa(z), z \in B(0,1), \kappa \in \mathbb{R}$, are ordinary differential operators on $C^\infty(\mathbb{R}^+) \otimes \mathbb{C}^2$. The explicit formulas for

$$\widetilde{X}_{jk} := \Omega_\kappa(x_{jk}), \quad \widetilde{Y}_l := \Omega_\kappa(y_1)$$

read[17]

$$\widetilde{X}_{-1-1} = ir^2, \quad \widetilde{X}_{1-1} = r\frac{d}{dr} + \frac{1}{2},$$

$$\widetilde{X}_{11} = i\left(-\frac{d^2}{dr^2} + \frac{\kappa^2}{r^2} - \frac{\kappa}{r^2} \otimes \sigma_3\right), \tag{4.1}$$

$$\widetilde{Y}_{-1} = \eta r \otimes \sigma_2, \quad \widetilde{Y}_1 = \bar{\eta}\left(\frac{d}{dr} \otimes \sigma_2 - i\frac{\kappa}{r} \otimes \sigma_1\right), \quad i \equiv \sqrt{-1}.$$

Here $\eta := \exp(i\pi/4)$ and $\sigma_1, \sigma_2, \sigma_3$ are Pauli matrices. The single Casimir element of $B(0,1)$,

$$k_2 := 2x_{1-1}^2 - \{x_{11}, x_{-1-1}\} + [y_1, y_{-1}],$$

is represented by

$$\widetilde{K}_2 := \Omega_\kappa(k_2) = (2\kappa^2 - \tfrac{1}{2})I.$$

Let us denote the components of any $\Phi \in C^\infty(\mathbb{R}^+) \otimes \mathbb{C}^2$ by $\varphi_\alpha, \alpha = \pm 1$:

$$\Phi \equiv \{\varphi_+, \varphi_-\}, \quad \varphi_\alpha \in C^\infty(\mathbb{R}^+),$$

and let

$$\sigma^\alpha := \tfrac{1}{2}(\sigma_1 - i\alpha\sigma_2).$$

As has been shown in Sec. II, the relevant operators are

$$\tilde{a} := \Omega_\kappa(a_1) = 2^{-1/2}(\widetilde{Y}_1 - i\widetilde{Y}_{-1}),$$

$$\tilde{n} := \Omega_\kappa(b_{1-1}) = \tfrac{1}{2}\{\tilde{a}, \tilde{a}^\#\}.$$

Substituting from (4.1) yields

$$\tilde{a} = \eta 2^{-1/2} \sum_{\alpha = \pm 1} \alpha\left(\frac{d}{dr} + r - \frac{\alpha\kappa}{r}\right) \otimes \sigma^{(\alpha)}, \tag{4.2}$$

$$\tilde{n} = \frac{1}{2}\left(-\frac{d^2}{dr^2} + r^2 + \frac{\kappa^2}{r^2} - \frac{\kappa}{r^2} \otimes \sigma_3\right). \tag{4.3}$$

The condition $\tilde{a}\Psi = 0$, which determines the vacuum subspace $V_{\Omega_\kappa} \equiv V_\kappa$, then becomes

$$\left(\frac{d}{dr} + r - \frac{\alpha\kappa}{r}\right)\Psi_\alpha = 0, \quad \alpha = \pm 1. \tag{4.4}$$

The solution $\exp(-r^2/2)r^{\alpha\kappa}$ belongs to $C^\infty(\mathbb{R}^+)$ for both $\alpha = \pm 1$ and thus

$$V_\kappa = \{\Psi_\kappa^{(+)}, \Psi_\kappa^{(-)}\}_{\text{lin}},$$

$$(\psi_\kappa^{(\beta)})_\alpha(r) := \delta_{\alpha-\beta} \exp\left(-\frac{r}{2}\right)^2 r^{\alpha\kappa}, \tag{4.5}$$

i.e., $\dim V_\kappa = 2$. Hence, the finite degeneracy of vacuum (FDV condition) is automatically fulfilled.

Of the conditions (3.3), (3.4b) and (3.4c), which determine the vector Ψ_λ, only the first one, i.e., $\tilde{n}\Psi_\lambda = \nu\Psi_\lambda$, makes sense for $n = 1$. By using (4.3), one finds

$$\tilde{n}\Psi_\kappa^{(\beta)} = (\beta\kappa + \tfrac{1}{2})\Psi_\kappa^{(\beta)}, \tag{4.6}$$

and thus a vector $\Psi \in V_\kappa$ fulfills $\tilde{n}\Psi = \nu\Psi$ for some $\nu > 0$ if $\Psi = \Psi_\kappa^{(+)}, \kappa > -\frac{1}{2}$, or $\Psi = \Psi_\kappa^{(-)}, \kappa < \frac{1}{2}$. As no highest weight occurs for $n = 1$, and by Eq. (4.6) the eigenvalue ν depends on κ only, we hereafter write Ψ_κ instead of Ψ_λ. Consequently, the necessary conditions of Sec. III are simplified as follows.

Proposition 4.1: Suppose that for a given $\kappa \in \mathbb{R}$, there is an Ω_κ-invariant subspace $\mathscr{D}_\kappa \subset C^\infty(\mathbb{R}^+) \otimes \mathbb{C}^2$ fulfilling $\mathscr{D}_\kappa \cap V_\kappa \neq \{0\}$ and let $\Omega_\kappa \restriction \mathscr{D}_\kappa$ be an irreducible *-representation of $B(0,1)$. Then

$$\mathscr{D}_\kappa \equiv \mathscr{D}(\Psi_\kappa) = \{(\tilde{a}^\#)^k \Psi_\kappa : k = 0,1,\ldots\}_{\text{lin}}, \tag{4.7a}$$

where

$$\Psi_\kappa := \begin{cases} \Psi_\kappa^{(+)}, & \kappa \in (-\frac{1}{2}, \infty) \setminus \{0\}, \\ \Psi_\kappa^{(-)}, & \kappa \in (-\infty, \frac{1}{2}) \setminus \{0\}. \end{cases} \tag{4.7b}$$

Proof: With the help of $\{\tilde{a}, \tilde{a}^\#\} = 2\tilde{n}$, one easily verifies by induction that for each monomial $\widetilde{M}$ in $\tilde{a}, \tilde{a}^\#$ holds $\widetilde{M}\Psi_\kappa \in \{(a^\#)^k \Psi_\kappa : k = 0,1,\ldots\}_{\text{lin}}$. The reason for excluding the values $\kappa = 0$ and $\kappa = \mp\frac{1}{2}$ for $\Psi_\kappa = \Psi_\kappa^{(\pm)}$ is as follows: for $\kappa = 0$, the operators $\tilde{a}, \tilde{a}^\#$ equal direct sum of two identical Schrödinger representations of the canonical commutation relations, which is the case we are not interested in (see the discussion in Sec. II). If $\kappa = \mp\frac{1}{2}$, then by Eq. (4.6) one has for any scalar product on $\mathscr{D}$ under which $\Omega_\kappa \restriction \mathscr{D}$ becomes a *-representation: $\tilde{a}^\#\Psi_\kappa^{(\pm)} = 0$, and, in view of $\tilde{a}\Psi_\kappa^{(\pm)} = 0$, the representation would be trivial. ■

B. Construction of irreducible *-representations

According to Eq. (4.7a), the sought domain $\mathscr{D}_\kappa$ is spanned by functions

$$\Phi_k := (a^\#)^k \Psi_\kappa. \tag{4.8}$$

For getting the functional dependence $r \mapsto \Phi_k(r)$, notice that Eq. (4.2) yields for any $\Phi \equiv \{\varphi_+, \varphi_-\} \in C^\infty(\mathbb{R}^+) \otimes \mathbb{C}^2$,

$$\begin{aligned} (\tilde{a}^\#\phi)_+ &= -\bar{\eta}2^{-\frac{1}{2}}\left(\frac{d}{dr} - r + \frac{\kappa}{r}\right)\varphi_-, \\ (\tilde{a}^\#\Phi)_- &= \bar{\eta}2^{-1/2}\left(\frac{d}{dr} - r - \frac{\kappa}{r}\right)\varphi_+. \end{aligned} \tag{4.9}$$

On the other hand, by applying functional relations for the Laguerre polynomials (Ref. 18, §8.971) to functions

$$f_k^{(\alpha)}(r) := c_k^{(\alpha)} r^{\alpha+1/2} \exp(-r^2/2) L_k^{(\alpha)}(r^2),$$

$$c_k^{(\alpha)} := \left(\frac{2k!}{\Gamma(\alpha + k + 1)}\right)^{1/2}, \tag{4.10}$$

one finds

$$\left(\frac{d}{dr} - \frac{\alpha + \frac{1}{2}}{r} - r\right) f_k^{(\alpha)} = -2(\alpha + k + 1)^{1/2} f_k^{(\alpha+1)},$$

$$\left(\frac{d}{dr} + \frac{\alpha + \frac{1}{2}}{r} - r\right) f_k^{(\alpha+1)} = 2(k+1)^{1/2} f_{k+1}^{(\alpha)}. \tag{4.9'}$$

Consider the first of alternatives (4.7b): $\Psi_\kappa = \Psi_\kappa^{(+)}$, $\kappa \in (-\frac{1}{2}, \infty) \setminus \{0\}$. Comparing (4.5) to (4.10) gives

$$\Psi_\kappa = \Phi_0 = \left(\frac{\Gamma(\kappa + 1/2)}{2}\right)^{1/2} F_0, \quad F_0 := \{f_0^{(\kappa - 1/2)}, 0\}.$$

Equations (4.9) and (4.9′) then yield by induction

$$\Phi_k = (\bar{\eta}\sqrt{2})^k\Big(\frac{1}{2}\Gamma\Big(\kappa+\frac{1}{2}+\Big[\frac{k+1}{2}\Big]\Big)\Big[\frac{k}{2}\Big]!\Big)^{1/2} F_k,$$
$$k=0,1,... \qquad (4.11)$$

(for any $X\in\mathbf{R}$, $[x]$ denotes the largest integer that does not exceed x), with

$$F_{2k}\equiv F_{2k}^{(\kappa)} := \{f_k^{(\kappa-1/2)},0\},$$
$$F_{2k+1}\equiv F_{2k+1}^{(\kappa)} := \{0,-f_k^{(\kappa+1/2)}\}. \qquad (4.12)$$

Clearly, the functions F_k also span $\mathscr{D}(\Psi_\kappa^{(+)})$ and we shall see when considering the *-condition, that working with F_k instead of Φ_k has some technical advantages.

By $\tilde{a}^{\#}\Phi_k = \Phi_{k+1}$ and (4.11), one has

$$\tilde{a}^{\#}F_k = \bar{\eta} d_{k+1}F_{k+1}, \qquad (4.13a)$$

with

$$d_k\equiv d_k(\kappa) := [k+(1-(-1)^k)\kappa]^{1/2}. \qquad (4.13b)$$

For $\tilde{a}\Phi_k$, we get by induction with the help of $\{a,\tilde{a}^{\#}\}=2\tilde{n}$, $[\tilde{n},(\tilde{a}^{\#})^l] = l(\tilde{a}^{\#})^l$, and Eq. (4.6)

$$\tilde{a}\Phi_{2k} = 2k\Phi_{2k-1}, \quad \tilde{a}\Phi_{2k+1} = (2k+2\kappa+1)\Phi_{2k},$$

whence

$$\tilde{a}F_k = \eta d_k F_{k-1}. \qquad (4.13c)$$

Now that the action of operators $\tilde{a}$, $\tilde{a}^{\#}$ on the vectors spanning the domain $\mathscr{D}(\Psi_\kappa^{(+)})$ is known, algebraic irreducibility of $\Omega_\kappa \restriction \mathscr{D}(\psi_\kappa^{(+)})$ can easily be proven. We have to verify that to each $\phi\equiv\alpha_0F_0+\cdots+\alpha_KF_K$, $\alpha_K\neq 0$, there exist operators $T,S\in\mathscr{U}(\tilde{a},a^{\#})$ such that $\phi = TF_0$, $F_0=S\phi$. Existence of T directly follows by Eqs. (4.8) and (4.11); further, (4.13c) yields $\tilde{a}^K\phi = \alpha_K\eta^K d_K d_{K-1}\cdots d_1F_0$, and thus $S\sim\tilde{a}^K$ (notice that $d_k\neq 0$, for $k=1,2,...$).

Next we have to introduce a scalar product$(\cdot,\cdot)$ on $\mathscr{D}(\Psi_\kappa^{(+)})$ such that the condition (3.9) holds. As has been argued in Sec. III, such a scalar product must fulfill

$$(F_k,F_l) = t_k\delta_{k-l}, \quad t_k>0, \quad k,l=0,1,... \,.$$

Now the condition (3.9) is equivalent to

$$(\tilde{a}F_k,F_l) = (F_k,\tilde{a}^{\#}F_l), \quad k,l=0.1,... \,. \qquad (4.14)$$

By Eqs. (4.13) these conditions become

$$d_kt_{k-1}\delta_{k-1-l} = d_kt_k\delta_{k-1-l},$$
$$\text{i.e., } t_{k-1}=t_k, \quad k=1,2,... \,.$$

We thus see that Eq. (4.14) is satisfied iff there is a positive t such that

$$(F_k,F_l) = (F_k,F_l)_t := t\delta_{k-l}, \quad k,l=0,1,... \,. \qquad (4.15)$$

Let $\mathscr{H}_t$ be the Hilbert space obtained by completing $\mathscr{D}(\Psi_\kappa^{(+)})$ under $(\cdot,\cdot)_t$ and let $\pi_\kappa^{(t)}$ be the representation $\Omega_\kappa\restriction\mathscr{D}(\Psi_\kappa^{(+)})$ regarded as a Hilbert-space representation on $\mathscr{H}_t$; especially we set $\pi_\kappa\equiv\pi_\kappa^{(1)}$. Clearly, one has

$$\pi_\kappa^{(t)} = V_t\pi_\kappa V_t^{-1}, \quad t>0,$$

V_t being the unitary map of $\mathscr{H}_1$ onto $\mathscr{H}_t$ given by $V_tF_k := t^{-1/2}F_k$.

Hence, it is sufficient to consider the case $t=1$ only. The functions $f_k^{(\alpha)}$, $k=0,1,...$, form an orthonormal basis in $L^2(\mathbf{R}^+)$ for each $\alpha>-1$, which implies that $\{F_k\}_{k=0}^\infty$ is an orthonormal basis in $L^2(\mathbf{R}^+)\otimes\mathbf{C}^2$ for each $\kappa>-\frac{1}{2}$. Thus $\mathscr{H}_1$ can be chosen as $L^2(\mathbf{R}^+)\otimes\mathbf{C}^2$, this choice being unique up to unitary maps.

The above considerations concerning the choice $\Psi_\kappa := \Psi_\kappa^{(+)}$ can be concluded as follows: for each $\kappa>-\frac{1}{2}$, $\kappa\neq 0$, the linear representation Ω_κ yields an irreducible representation π_κ on $L^2(\mathbf{R}^+)\otimes\mathbf{C}^2$ with domain $\{F_k^{(\kappa)}: k=0,1,...\}_{\text{lin}}$. The representation π_κ satisfies the *-condition (4.14) and is determined uniquely up to unitary equivalence.

For the other choice $\Psi_\kappa := \Psi_\kappa^{(-)}$, $\kappa>\frac{1}{2}$, $\kappa\neq 0$, everything can be repeated step by step. By defining

$$G_{2k}\equiv G_{2k}^{(\kappa)} := \{0,f_k^{(-\kappa-1/2)}\},$$
$$G_{2k+1}\equiv G_{2k+1}^{(\kappa)} := \{f_k^{(-\kappa+1/2)},0\}, \qquad (4.16)$$
$$d_k^{(-)} := d_k(-\kappa),$$

we find that for each $\kappa<\frac{1}{2}$, the relations

$$\tilde{a}G_k = \eta d_k^{(-)}G_{k-1}, \quad \tilde{a}^{\#}G_k = \bar{\eta}d_{k+1}^{(-)}G_{k+1}, \qquad (4.17)$$

determine an irreducible representation ρ_κ of $B(0,1)$ on $L^2(\mathbf{R}^+)\otimes\mathbf{C}^2$, with domain $\{G_k: k=0,1,...\}_{\text{lin}}$. The *-condition (4.14) is satisfied and any other Hilbert-space representation with these properties obtained from $\Omega_\kappa\restriction\mathscr{D}(\Psi_\kappa^{(-)})$ is unitarily equivalent to ρ_κ.

However, the representations ρ_κ are in fact of no interest as for each $\kappa<\frac{1}{2}$ the representations ρ_κ, $\pi_{-\kappa}$ are unitarily equivalent:

$$\rho_\kappa = U\pi_{-\kappa}U^{-1}, \quad U := -iI\otimes\sigma_2.$$

Proof: By (4.12) and (4.16) we see that $UF_k^{(-\kappa)} = G_k^{(\kappa)}$, $k=0,1,...$, i.e., U maps the domains of ρ_κ and $\pi_{-\kappa}$ onto each other. Further, Eqs. (4.13) and (4.17) yield

$$U\tilde{a}(-\kappa)F_k^{(-\kappa)} = \eta d_k(-\kappa)F_{k-1}^{(-\kappa)} = \eta d_k^{(-)}G_{k-1}^{(\kappa)}$$
$$= \tilde{a}(\kappa)G_k^{(\kappa)} = \tilde{a}(\kappa)UF_k^{(-\kappa)},$$

$k=0,1,...$; similarly

$$Ua^{\#}(-\kappa)F_k^{(-\kappa)} = \tilde{a}^{\#}(\kappa)UF_k^{(-\kappa)}. \qquad \blacksquare$$

The main results of this section can be summarized as follows.

Theorem 4.2: (i) For each $\kappa>-\frac{1}{2}$, $\kappa\neq 0$, the operators $\Omega_\kappa(z)\restriction\mathscr{D}_\kappa$, $z\in B(0,1)$, form an irreducible *-representation π_κ of $B(0,1)$ on $L^2(\mathbf{R}^+)\otimes\mathbf{C}^2$ with domain $\mathscr{D}_\kappa := \{F_k^{(\kappa)}: k=0,1,...\}_{\text{lin}}$ specified by Eq. (4.12) and projection $E := I\otimes(\sigma_0+\sigma_3)/2$. In addition, π_κ has nondegenerated vacuum

$$\Psi_\kappa^{(+)} = \Big(\frac{\Gamma(\kappa+\frac{1}{2})}{2}\Big)^{1/2}F_0^{(\kappa)}.$$

(ii) Any two representations in the family

$$\Pi := \{\pi_\kappa: \kappa\in(-\tfrac{1}{2},\infty)\setminus\{0\}\}$$

are nonequivalent.

(iii) The family contains (up to unitary equivalence) all the Hilbert space irreducible *-representations of $B(0,1)$ that can be obtained from linear representations Ω_κ, $\kappa\in\mathbf{R}\setminus\{0\}$, and whose domain contains a vacuum vector.

Remark 4.3: More explicitly, (iii) states the following: Let $\kappa\in\mathbf{R}\setminus\{0\}$, $\mathscr{D}'$ be a subspace in $C^\infty(\mathbf{R}^+)\otimes\mathbf{C}^2$ having nontrivial intersection with the vacuum subspace V_κ, and $\mathscr{H}$ be a Hilbert space such that $\overline{\mathscr{D}}' = \mathscr{H}$, and the operators $\Omega_\kappa(z) \upharpoonright \mathscr{D}'$, $z \in B(0,1)$, form an irreducible *-representation π'_κ of $B(0,1)$ on $\mathscr{H}$ with projection $\hat{E}'$. Then there is a unitary map $V: L^2(\mathbf{R}^+) \otimes \mathbf{C}^2 \to \mathscr{H}$ for which

$$\pi'_\kappa = \begin{cases} V\pi_\kappa V^{-1}, & \text{if } \kappa > -\frac{1}{2}, \\ V\pi_{-\kappa} V^{-1}, & \text{if } \kappa < -\frac{1}{2}, \end{cases} \tag{4.18}$$

and $\hat{E}' = V\hat{E}V^{-1}$ or $\hat{E}' = V(\hat{I} - \hat{E})V^{-1}$.

Proof of the Theorem: (i) It remains to verify that $\hat{E}$ fulfills $\hat{E}\mathscr{D}_\kappa \subsetneq \mathscr{D}_\kappa$ and that for each $\phi \in \mathscr{D}_\kappa$ holds

$$E\Omega_\kappa(x)\Phi = \Omega_\kappa(x)\hat{E}\Phi, \tag{4.19a}$$

if x is any even element of $B(0,1)$,

$$\hat{E}\Omega_\kappa(y)\Phi = \Omega_\kappa(y)(\hat{I} - \hat{E})\Phi \tag{4.19b}$$

if y is odd (cf. Appendix). All these conditions can readily be verified by using Eqs. (4.12) and (4.1).

(ii) Let $\pi_\kappa, \pi_{\kappa'} \in \Pi, \kappa \neq \kappa'$; in view of (4.6) and $[\tilde{n}, \tilde{a}^{\#k}] = k\tilde{a}^{\#k}$ the minimal eigenvalue of $\pi_\kappa(b_{1-1}) \equiv \hat{n} \upharpoonright \mathscr{D}_\kappa$ equals $\kappa + \frac{1}{2}$ and hence $\pi_\kappa, \pi_{\kappa'}$ cannot be equivalent.

(iii) By Proposition 4.1, there is nonzero Ψ'_κ in $\mathscr{D}' \cap V_\kappa$ and the alternative (4.7b) holds for Ψ'_κ. Then Eq. (4.18) ensues from the considerations in the beginning of this section. Further, Eq. (4.19a) implies that $\hat{n}' := \pi'_\kappa(b_{1-1})$ commutes with $\hat{E}'$; then, by (4.18), $\hat{n}'$ has the same spectrum as $\hat{n}$, i.e., a pure-point spectrum with nondegenerate eigenvalues. Hence $\hat{E}'F'_k = p_k F'_k$, $F'_k := VF_k^{(\kappa)}$ if $\kappa > -\frac{1}{2}$, $F'_k := VF_k^{(-\kappa)}$ if $\kappa < -\frac{1}{2}$; moreover, $p_k = 0$ or 1 since $\hat{E}'$ is a projection. Finally, (4.19b) yields for $\hat{a}' \equiv \pi'_\kappa(a_1)$: $\hat{E}'\hat{a}'F'_k = a'(I - \hat{E}')F'_k$ and, by using (4.18) and (4.13c), we find $p_{k-1} = 1 - p_k$, $k = 1,2,\ldots$. Thus, one has either $\hat{E}'F'_{2k} = F'_{2k}$, $\hat{E}'F'_{2k+1} = 0$ or $\hat{E}'F'_{2k} = 0$, $\hat{E}'F'_{2k+1} = F'_{2k+1}$. Since

$$\hat{E}F_{2k} = F_{2k}, \hat{E}F_{2k+1} = 0, \tag{4.20}$$

the first possibility implies $\hat{E}' = V\hat{E}V^{-1}$ and the second implies $\hat{E}' = V(\hat{I} - \hat{E})V^{-1}$. ∎

C. Essential self-adjointness

The *-property (4.14) means that the operators $\hat{X}_{jk} := \pi_\kappa(x_{jk})$, $\hat{Y}_j := \pi_\kappa(y_j)$ satisfy $\hat{X}^\dagger_{jk} = -\hat{X}_{jk}$, $\hat{Y}^\dagger_j = -i\hat{Y}_j$, i.e., $i\hat{X}_{jk}$ and $\hat{\eta}\hat{Y}_j$ are symmetric. By using Nelson's analytic-vector theorem,[19] we will now prove that these operators are moreover essentially self-adjoint (e.s.a.).

Hereafter only $\kappa \in (-\frac{1}{2}, \infty) \setminus \{0\}$ are considered and the notation $\hat{a} \equiv \hat{a}(\kappa) := \pi_\kappa(a_1)$ is used.

Lemma 4.4: Let $\hat{A}_p$ be a monomial in $\hat{a}, \hat{a}^{\neq}$ of pth degree, $p = 1,2,\ldots$. Then

$$\|\hat{A}_p F_k\|^2 < \frac{\Gamma(k + 2\kappa + p + 2)}{\Gamma(k + 2\kappa + 2)}, \quad k = 0,1,\ldots, \tag{4.21}$$

where $\|\cdot\|$ is the norm on $\mathscr{H} \equiv L^2(\mathbf{R}^+) \otimes \mathbf{C}^2$.

Proof: Since $\kappa + \frac{1}{2} > 0$, Eq. (4.13b) yields $d_k^2 < k + 2\kappa + 1$. For $p = 1$ the assertion now follows by Eqs. (4.13a) and (4.13c), and the proof is finished by induction if one realizes that $\hat{A}_{p+1}$ equals either $\hat{A}_p\hat{a}^\dagger$ or $\hat{A}_p\hat{a}$. ∎

Proposition 4.5: If $\hat{P}_d$ is a homogeneous polynomial in $\hat{a}$, $\hat{a}^\ddagger$ of degree d, $d = 1,2$, then each F_k, $k = 0,1,\ldots$, is an analytic vector of $\hat{P}_d$.

Proof: One has

$$\hat{P}_d = \sum_{r=1}^{2^d} \alpha_r \hat{A}_d^{(r)},$$

where $\hat{A}_d^{(r)}$ are the independent monomials of degree d. Let $M := \max|\alpha_r|$; then the estimate (4.21) yields for any F_k, $t > 0$:

$$\sum_{n=0}^\infty \|P_d^n F_k\| \frac{t^n}{n!} < \sum_{n=0}^\infty \sum_{s=1}^{2^{nd}} \frac{(Mt)^n}{n!} \left(\frac{\Gamma(k + 2\kappa + nd + 2)}{\Gamma(k + 2\kappa + 2)} \right)^{1/2} = \sum_{n=0}^\infty \frac{(2^d Mt)^n}{n!} \left(\frac{\Gamma(k + 2\kappa + nd + 2)}{\Gamma(k + 2\kappa + 2)} \right)^{1/2}.$$

This series is convergent for any $t > 0$, if $d = 1$ and for $0 < t < (8M)^{-1}$ if $d = 2$, whence the assertion. ∎

As $\{F_k : k = 0,1,\ldots\}$ is a total set in $\mathscr{H}$, we get, by the Nelson theorem the following corollary.

Corollary 4.6: If $\hat{P}_d = \hat{P}_d^\ddagger$, $d = 1,2$; then $\hat{P}_d$ is e.s.a.; in particular, this holds true for the operators $i\hat{X}_{jk}$, $\hat{\eta}\hat{Y}_j$, $j,k = \pm 1$.

If $\hat{B}$ is a biquadratic homogeneous polynomial, then the above proposition implies that the series

$$\sum_{n=0}^\infty \frac{\|B^n F_k\| t^n}{(2n)!}$$

is convergent for $t < (64M)^{-1}$. Thus $\hat{B}$ has a total set of semianalytic vectors and by the Nussbaum theorem,[19] $\hat{B}$ is e.s.a. if $\hat{B} > 0$. An important example provides the operator

$$\hat{B} \equiv \hat{N} := -(\hat{X}^2_{11} + \hat{X}^2_{1-1} + \hat{X}^2_{-1-1}).$$

Essential self-adjointness of $\hat{N}$ implies that the representation τ_κ of sl(2,**R**) ~ sp(2,**R**), which is obtained by restricting π_κ to the even subalgebra of the unique real form osp(1,2) of $B(0,1)$, is integrable to a unitary representation of the universal covering group of SL(2,**R**) (see Ref. 20). We shall return to this point in the next section.

Remark 4.7: The conclusions concerning integrability of τ_κ and essential self-adjointness of $i\hat{X}_{jk}$ can alternatively be obtained as follows. Introduce a new basis in sl(2,**R**):

$$q_1 := x_{1-1}, \quad q_2 := \frac{(x_{11} - x_{-1-1})}{2},$$

$$q_3 := \frac{(x_{11} + x_{-1-1})}{2} = ib_{1-1}$$

[see (2.3b)] and set $\hat{Q}_r := \pi_\kappa(q_r)$. For the Casimir element of sl(2,**R**), one has $c_2 = 2x^2_{1-1} - \{x_{11}, x_{-1-1}\} = 2(q_1^2 + q_2^2 - q_3^2)$, and thus $\Delta := -(\hat{Q}_1^2 + \hat{Q}_2^2 + \hat{Q}_3^2)$ commutes with $\hat{Q}_3$. Now $\hat{Q}_3 = i\hat{n}$ and since F_k, $k = 0,1,\ldots$, are nondegenerated eigenvectors of $\hat{n}$, they are also eigenvectors (and hence analytic vectors) of Δ. Consequently, Δ is e.s.a., which further implies that any operator $i(a_1\hat{Q}_1 + a_2\hat{Q}_2 + a_3\hat{Q}_3)$, $a_r \in \mathbf{R}$, is e.s.a.[21]

D. Restriction of π_κ to the even subalgebra sl(2,R)$\subset$osp(1,2)

Let τ_κ be the restriction of π_κ to sl(2,$\mathbb{R}$). According to Eq. (4.19a), τ_κ is reduced by the projection $\widehat{E} = I \otimes (\sigma_0 + \sigma_3)/2$:

$$\tau_\kappa = \tau_\kappa^{(+)} \oplus \tau_\kappa^{(-)},$$

the $\tau_\kappa^{(\alpha)}$, $\alpha = \pm 1$, being skew-symmetric representations of sl(2,$\mathbb{R}$) on $L^2(\mathbb{R}^+)$ with domains $\mathscr{D}_\kappa^{(\alpha)} := \{f_k^{(\kappa - \alpha/2)}: k = 0,1,...\}_{\text{lin}}$ [cf. Eqs. (4.12) and (4.20)].

Proposition 4.8: The representations $\tau_\kappa^{(\alpha)}$ are irreducible.

Proof: We have to show that the set $\mathscr{U}^{(\alpha)} := \mathscr{U}(\widehat{X}_{11}^{(\alpha)}, \widehat{X}_{1-1}^{(\alpha)}, \widehat{X}_{-1-1}^{(\alpha)}), \widehat{X}_{jk}^{(\alpha)} := \widehat{X}_{jk} \upharpoonright \mathscr{D}_\kappa^{(\alpha)}$, has no invariant subspaces. By Eqs. (2.3) we see that $\widehat{X}_{jk}$ are homogeneous quadratic polynomials in $\hat{a}, \hat{a}^\ddagger$. Irreducibility of $\tau_\kappa^{(\alpha)}$ can then be verified with the help of Eqs. (4.13) by repeating the argument we used for proving absence of invariant subspaces of $\mathscr{D}(\Psi_\kappa^{(+)})$. ■

Remark 4.9: Let $\mathscr{H}_k^{(\alpha)} \subset L^2(\mathbb{R}^+)$ be the one-dimensional subspace spanned by $f_k^{(\kappa - \alpha/2)}$. Clearly, each of the domains $\mathscr{D}_\kappa^{(\alpha)}$ can be expressed as the algebraic sum of subspaces $\mathscr{H}_k^{(\alpha)}$

$$\mathscr{D}_\kappa^{(\alpha)} = \sum_{k=0}^{\infty} \mathscr{H}_k^{(\alpha)}.$$

The $f^{(\kappa - \alpha/2)}$ are eigenvectors of $\hat{h}^{(\alpha)} := \tau_\kappa^{(\alpha)}(b_{1-1})$ corresponding to eigenvalues

$$\lambda_{k,\kappa}^{(\alpha)} := 2k + \kappa + (|\alpha| - \alpha + 1)/2.$$

Since the maximal compact subalgebra u(1)$\subset$sl(2,$\mathbb{R}$) is spanned by $ib_{1-1} = (x_{11} + x_{-1-1})/2$ (see Remark 3.2), the restriction $\tau_\kappa^{(\alpha)} \upharpoonright$ u(1) equals direct sum of one-dimensional representations of u(1) on $\mathscr{H}_k^{(\alpha)}$ that are uniquely determined by eigenvalues $\lambda_{k,\kappa}^{(\alpha)}$. This means that the so-called weight diagram of $\tau_\kappa^{(\alpha)}$ is $\{\lambda_{k,\kappa}^{(\alpha)}: k = 0,1,...\}$.

Each of $\tau_\kappa^{(\alpha)}$ is integrable to a representation $\mathscr{T}_\kappa^{(\alpha)}$ of $G \equiv \overline{SL(2,\mathbb{R})}$, as the vectors $f_k^{(\kappa - \alpha/2)}$, $k = 0,1,...$, are analytic vectors of

$$\widehat{N}^{(\alpha)} := -((\widehat{X}_{11}^{(\alpha)})^2 + (\widehat{X}_{1-1}^{(\alpha)})^2 + (\widehat{X}_{-1-1}^{(\alpha)})^2),$$

and form a total set in $L^2(\mathbb{R}^+)$. Moreover, $\mathscr{T}_\kappa^{(\alpha)}$ is a unitary irreducible representation (UIR) of G on $L^2(\mathbb{R}^+)$ with the following property[22]: let K be the simply connected subgroup of G whose Lie algebra is u(1); then $\mathscr{T}_\kappa^{(\alpha)} \upharpoonright K$ equals the direct sum of the UIR's of K on $\mathscr{H}_k^{(\alpha)}$, each of them being uniquely determined by the eigenvalue $\lambda_{k,\kappa}^{(\alpha)}$. In fact, $K \sim \mathbb{R}$ and the UIR of K on $\mathscr{H}_k^{(\alpha)}$ is equivalent to $t \mapsto \exp(it\lambda_{k,\kappa}^{(\alpha)})$, $t \in \mathbb{R}$.

V. CONCLUDING REMARKS

The problem of constructing representations of $B(0,1)$ was recently considered [starting with a family of linear representations of $B(0,1)$ equivalent to our $\{\Omega^{(1)}\}$] by Mukunda *et al.*[23] These authors stressed the importance of specifying carefully domains of unbounded Hilbert-space operators that arise from "formal" differential operators $\Omega(z)$, $z \in B(0,1)$. They also corrected some erroneous conclusions of an earlier study.[24]

The "Schrödinger description" of Ref. 24 is in fact identical with our family Π, and so are the representations of osp(1,2) used by D'Hoker and Vinet in their study of dynamical symmetries of Dirac monopole.[25] On the other hand, the results (b)–(d) from the list in Sec. I are new, as well as the approach we used. Its advantages become apparent especially when considering the cases $n \geqslant 2$ for which the subalgebra of sp($2n$,$\mathbb{C}$) that leaves invariant the vacuum subspace is nontrivial (cf. Lemma 3.1 and Proposition 3.4). Construction of irreducible *-representations of $B(0,2)$ based on this approach is in progress.

ACKNOWLEDGMENT

The authors are truly indebted to Dr. M. Bednář for enlightening discussion.

APPENDIX: HILBERT-SPACE *-REPRESENTATIONS OF LIE SUBALGEBRAS BY UNBOUNDED OPERATORS

It is well-known that Hilbert-space representations of a real LSA whose even subalgebra is noncompact have the following property: if even generators are represented by skew-symmetric operators, then at least one of them must be unbounded.[26] That is how one mostly arrives at *-representations of LSA by unbounded operators. Corresponding definitions are obtained by generalizing, on the one hand, the definition of finite-dimensional *-representations of LSA as given, e.g., in Ref. 6, and, on the other hand, that of ∞-dimensional *-representations of Lie algebras.[21]

Let $\mathscr{H}$ be an ∞-dimensional separable Hilbert-space and D its subspace such that

$$\overline{D} = \mathscr{H}. \tag{A1}$$

Consider the set $\text{End}_{\mathscr{H}} D \equiv \text{End}\, D$ of linear operators X on $\mathscr{H}$ satisfying

(i) $D(X) = D$, Ran $X \subset D$, so that D is a common invariant domain for all $X \in \text{End}\, D$,

(ii) $D(X^\dagger) \supset D$, Ran $X^\ddagger \subset D$,

where $X^\dagger$ is the usual Hilbert-space adjoint of X and

$$X^\ddagger := X^\dagger \upharpoonright D.$$

Then End D becomes an associative *-algebra with involution $X \mapsto X^\ddagger$.

For a given projection E on $\mathscr{H}$ such that

$$ED \subset D, \quad ED \neq D, \tag{A2}$$

consider the following subsets of End D:

$$(\text{End}\, D)_0 := \{X \in \text{End}\, D: XE\psi = EX\psi, \ \psi \in D\},$$
$$(\text{End}\, D)_1 := \{X \in \text{End}\, D: XE\psi = (I - E)X\psi, \ \psi \in D\}. \tag{A3}$$

Then one has

$$\text{End}\, D = (\text{End}\, D)_0 \oplus (\text{End}\, D)_1, \tag{A4}$$

and End D becomes a LSA if one defines multiplication $X, Y \mapsto \langle X, Y \rangle$ as the bilinear extension of

$$\langle X, Y \rangle := XY - (-1)^{\alpha\beta} YX, \quad X \in (\text{End}\, D)_\alpha,$$
$$Y \in (\text{End}\, D)_\beta, \quad \alpha, \beta = 0,1. \tag{A5}$$

This LSA, which is completely determined by the associative *-algebra End D and projection E, will be denoted (End D,E).[27] The mapping $X \mapsto X^\ddagger$ preserves the grading[28]:

$$X\in(\mathrm{End}\,D)_\alpha \Rightarrow X^\ddagger\in(\mathrm{End}\,D)_\alpha, \quad \alpha=0,1.$$

Further this mapping is an involution on End D and thus by (A5) one sees that (End D,E) is a *-LSA.

Definition: Let a *-LSA, $\mathscr{A}\equiv\mathscr{A}_0\oplus\mathscr{A}_1$ with multiplication $x, y\mapsto x\cdot y$ and involution $x\mapsto x^*$ be given. Further let D be a subspace in a separable Hilbert space $\mathscr{H}$ and E a projection on $\mathscr{H}$ such that the conditions (A1) and (A2) are fulfilled. Linear mapping π:

$$\mathscr{A}\ni x\mapsto\pi(x)\in\mathrm{End}\,D$$

is a *-representation of $\mathscr{A}$ on $\mathscr{H}$ with domain D and projection E if

(i) $\pi(\mathscr{A}_\alpha)\subset(\mathrm{End}\,D)_\alpha, \quad \alpha=0,1,$

(ii) $\pi(x\cdot y)=\langle\pi(x),\pi(y)\rangle,$

(iii) $\pi(x^*)=(\pi(x))^\ddagger.$

[1]J. Blank, P. Exner, M. Havlíček, and W. Lassner, J. Math. Phys. **23**, 350 (1982).

[2]M. Bednář, J. Blank, P. Exner, and M. Havlíček, JINR E2-82-771, Dubna, 1982.

[3]M. Bednář, J. Blank, P. Exner, and M. Havlíček, JINR E2-83-150, Dubna, 1983.

[4]G. Racah, "Group theory and spectroscopy," CERN Report 61-8, Geneva, 1961. Notice that also a slightly different basis of osp$(1,2n)$ is called Racah—see A. Bincer, J. Math. Phys. **24**, 2546 (1983).

[5]Linear differential operators on $C_N^\infty(M)$, $M\subset\mathbf{R}^m$ are linear maps D

$$\phi\mapsto D\phi:=\sum_{j_1,\dots,j_m} f_{j_1\dots j_m}(p_1^{j_1}\cdots p_m^{j_m})[\phi],$$

where $f_{j_1\dots j_m}$ are C^∞-functions on M with values in the space End $\mathbf{C}^N$ of linear operators on $\mathbf{C}^N$, $p_k^j[\phi]:=\partial^j\phi/\partial x_k^j$ and the summation extends over some finite subset of m-tuples of non-negative integers. Clearly these operators form a vector space that will be denoted Λ_N. The adjoint of D is defined by

$$D^\#\Phi:=\sum_{j_1,\dots,j_m}(-1)^{j_1+\cdots+j_m}(p_1^{j_1}\cdots p_m^{j_m})[f_{j_1\dots j_m}^+\Phi]$$

[see e.g. R. Courant, *Partial Differential Equations* (Interscience, New York, 1962)]. By applying the Liebnitz formula, one finds that $D^\#\in\Lambda_N$ and similarly can be verified that the set Λ_N is an associative algebra and $D\mapsto D^\#$ is an involution on it.

[6]M. Scheunert, "The theory of Lie superalgebras," in *Lecture Notes in Mathematics* (Springer, Berlin, 1979), Vol. 716.

[7]The involution $z\mapsto\vartheta(z)\equiv z^*$ is unique up to equivalence transformations $\vartheta\mapsto\varphi\vartheta\varphi^{-1}$ generated by automorphisms φ of $B(0,n)$—see M. Parker, J. Math. Phys. **21**, 689 (1980).

[8]Representation π (possibly ∞-dimensional) with domain $\mathscr{D}$ is algebraically irreducible if $\{0\}$ and $\mathscr{D}$ are the only π-invariant subspaces of $\mathscr{D}$. Throughout this paper irreducibility always means algebraic irreducibility.

[9]Notice that due to Eq. (2.1b) the operators $\hat{n}_j$ have the basic property of usual particle-number operators, viz. $[\hat{n}_j,\hat{a}_j]=-\hat{a}_j$, $[\hat{n}_j,\hat{a}_j^\ddagger]=\hat{a}_j^\#$. One further has $[\hat{n}_j,\hat{n}_k]=0$, although different modes need not commute: $[\hat{a}_j,\hat{a}_k]\neq0$, $[\hat{a}_j^\ddagger,\hat{a}_k^\ddagger]\neq0$, for $j\neq k$.

[10]J. Dixmier, Comp. Math. **13**, 263 (1956).

[11]E. Wigner, Phys. Rev. **77**, 711 (1950).

[12]H. S. Green, Phys. Rev. **90**, 270 (1953); O. W. Greenberg, A. M. L. Messiah, Phys. Rev. **138**, B 1155 (1965).

[13]Conversely, to each irreducible pB_n-set $\{\hat{a}_j,\hat{a}_j^\ddagger: j=1,2,\dots,n\}$ there is an irreducible *-representation π of $B(0,n)$: $\pi(y_j):=2^{-1/2}(\hat{a}_j+i\hat{a}_j^\ddagger)$, $\pi(y_{-j}):=2^{-1/2}(\hat{a}_j^\ddagger+i\hat{a}_j)$, $j=1,\dots,n$, $\pi(x_{jk}):=\frac{1}{2}\{\pi(y_j),\pi(y_k)\}$, $j,k,=\pm1,\dots,\pm n$, where $i=\sqrt{-1}$, and by Eq. (2.3a) one has $\pi(a_j)=\hat{a}_j$. This interrelation of $B(0,n)$ and pB_n-sets was formulated algebraically by A. Ganchev and T. Palev, J. Math. Phys. **21**, p. 797 (1980).

[14]If the number of degrees of freedom is finite, then p may assume any positive value [this is the case, e.g., for $n=1$—see (4.6)], whereas for infinite systems p is always integer.

[15]M. A. Naimark, *Linear Representations of Groups* (Nauka, Moscow 1976) (in Russian).

[16]If $T_1,\dots,T_n$ are linear operators on a Hilbert space $\mathscr{H}$ with a common invariant domain D, then $\mathscr{U}(T_1,\dots,T_n)$ denotes the subalgebra of End$_{\mathscr{H}}$ D generated by $T_1,\dots,T_n$.

[17]See, Ref. 1. The basis $\{Q_\pm, Q_3, V_\pm\}$ of osp(1,2) and the real parameter c used there are related to the Racah basis [Eqs. (1.1a)–(1.1d)] and parameter κ by $V_+=y_{-1}/2$, $V_-=-y_1/4$, $Q_+=x_{-1-1}$, $Q_-=-x_{11}/4$, $Q_3=x_{1-1}/2$, and $\kappa=2c$, respectively. Accordingly, for the Casimir element K_2: $=Q_3^2+\{Q_+,Q_-\}/2+[V_+,V_-]$ one has $K_2=k_2/8$.

[18]I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Sums and Products* (Fizmatgiz, Moscow 1963) (in Russian).

[19]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. II, §X.6.

[20]G. Warner, *Harmonic Analysis on Semi-simple Lie Groups* (Springer, Berlin, 1972), Vol. I, §4.4.

[21]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN, Warsaw, 1977), §11.5.

[22]Harish-Chandra, Trans. Am. Math. Soc. **75**, 185 (1953).

[23]N. Mukunda, E. C. G. Sudarshan, J. K. Sharma and C. L. Mehta, J. Math. Phys. **21**, 2386 (1980); **22**, 78 (1981).

[24]Y. Ohnuki and S. Kamefuchi, J. Math. Phys. **19**, 67 (1978).

[25]E. D'Hoker and L. Vinet, Phys. Lett. B **137**, 72 (1984); Lett. Math. Phys. **8**, 439 (1984).

[26]H. D. Doebner and O. Melsheimer, Nuovo Cimento **49**, 73 (1967).

[27]Notice that (End D,E) = (End $D,I-E$) since the conditions (A2) and (A3) hold for E iff they hold for $I-E$.

[28]Consider, e.g., the case $X\in(\mathrm{End}\,D)_1$; for any $\varphi,\psi\in D$ one has

$$(EX^\ddagger\varphi,\psi)=(\varphi,XE\psi)=(E'\varphi,X\psi), \quad E':=I-E.$$

Now $E'\varphi\in D\subset D(X^+)$ which implies $(E'\varphi,X\psi)=(X^+E'\varphi,\psi)=(X^\ddagger E'\varphi,\psi)$ and as $\overline{D}=\mathscr{H}$, one has $EX^\ddagger\varphi=X^\ddagger E'\varphi$ for each $\varphi\in D$, i.e., $X^\ddagger\in(\mathrm{End}\,D)_1$.

The generalized atypical supertableaux of the orthosymplectic groups OSP(2|2p)

Michel Gourdin
Universite Pierre et Marie Curie, Laboratoire de Physique Théorique et Hautes Energies, 4 Place Jussieu, 75230 Paris Cedex 05, France

(Received 25 July 1985; accepted for publication 20 August 1986)

The classification and the interpretation of the Young supertableaux of the orthosymplectic group OSP$(2|2p)$ are given. A particular emphasis is made on the generalized atypical supertableaux associated to nonfully reducible atypical representations.

I. INTRODUCTION

The interest for the mathematical aspects of supersymmetry starts with the pioneering paper of Corwin, Ne'eman, and Sternberg[1] closely followed by that of Pais and Rittenberg.[2] After a while a precise mathematical formulation of the Z_2-graded Lie algebra has been given by Kac[3] in a series of papers that contains, in particular, the description of typical and atypical finite-dimensional representations of simple classical superalgebras. A particular set of orthosymplectic superalgebras is considered in this paper.

We introduce a graded vector space $V = V_S \oplus V_A$ and let G be a nondegenerate even bilinear form on V such that the restriction of G to V_S is a symmetrical form and to V_A a skew symmetrical form. Let us introduce the dimensions of V_S and V_A,

$$\dim V_S = m, \quad \dim V_A = 2p, \quad m \geqslant 1, \quad p \geqslant 1.$$

The set of Z_2-graded matrices $A(m,2p)$ leaving invariant the even bilinear form G is, by definition, the orthosymplectic group OSP $(m|2p)$.[4,5]

The subgroup associated to the Bose sector is $\mathrm{SO}(m) \otimes \mathrm{Sp}(2p)$ and the classification of the orthosymplectic superalgebras is made according to the value of m. The special case $m = 2$ plays a particular role, the two Lie groups SO(2) and U(1) being locally isomorphic. This paper is devoted to a study of the Young supertableaux of the orthosymplectic groups OSP$(2|2p)$ whose Lie superalgebras are noted $C(p+1)$ by Kac.[3]

By study we mean a classification of the supertableaux according to their size and a knowledge of the $\mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ components of the supertableaux. The spirit of our investigation is analogous to that used for the supertableaux of the superunitary groups.[6-9]

Young supertableaux for supergroups have been introduced by Dondi and Jarvis,[10] Balentekin and Bars,[11] King,[12] and the particular case of orthosymplectic groups has also been considered by Farmer and Jarvis[13] and Hurni.[14]

We use here the tensor product method briefly described in Sec. III and, as a by-product, we are able to relate the highest and lowest weights of an atypical irreducible representation of $C(p+1)$. The results are given in Appendix A.

For the orthosymplectic groups OSP$(2|2p)$ we encounter two types of supertableaux.

(1) The first type is the supertableaux associated with an irreducible or to a fully reducible representation of OSP$(2|2p)$. They have, by themselves, a well-defined meaning and they will be called irreducible supertableaux. Their properties have already been studied by Farmer and Jarvis.[13] For completeness we briefly recall in Sec. IV the results using our framework of classification.

(2) The second type is the supertableaux associated with nonfully reducible atypical representations of OSP$(2|2p)$. They do not have a meaning by themselves and only a pair of such atypical supertableaux can make sense in terms of representation of OSP$(2|2p)$. The definition of the constituents of the generalized atypical supertableaux and the description of their atypical components are made in Sec. V, which contains the main original results of this paper. The particular case OSP$(2|4)$ associated with the $N = 2$ supersymmetry is explicitly discussed in Appendix B.

The situation turns out to be extremely similar to that found with the supertableaux of the superunitary groups SU$(n|1)$ or SU$(1|n)$.[6] This is due to the fact that in both superalgebras $C(p+1)$ and $A(n-1,0)$ the spectra of the atypical eigenvalues of the U(1) generator are nondegenerate.

A brief discussion is added in Sec. VI concerning the topology of the set of Young supertableaux of OSP$(2|2p)$ in relation to their total number of boxes.

The particular case of the orthosymplectic group OSP$(2|2)$ has been considered in a separate publication[15] and it will not be discussed here.

In order to make this paper self-consistent, the basic facts concerning the superalgebra $C(p+1)$ of the orthosymplectic group OSP$(2|2p)$ have been added in Sec. II. Detailed results can be found in Refs. 3 and 13 and we have retained here only those useful for a good understanding of the supertableau approach.

II. BASIC RESULTS ON THE SUPERALGEBRA $C(p+1)$ (REF. 3)

(1) The superalgebra $C(p+1)$ of the orthosymplectic group OSP$(2|2p)$ belongs to the class I superalgebra and it can be decomposed as

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

The sets Δ_0 and Δ_1 of even and odd roots are given by

$$\begin{aligned} \Delta_0 &= \{\pm e_i \pm e_j; \pm 2e_i\}, \\ \Delta_1 &= \{\pm d \pm e_i\}, \end{aligned} \qquad i \neq j = 1,2,\ldots,p.$$

 0022-2488/86/122832-10$02.50

We call the Cartan subalgebra H, and N^+ and N^- are the set of positive and negative generators,

$$L = H + N^+ + N^-.$$

For $C(p+1)$ the classification of the generators is given in Table I, where $i,j = 1,2,...,p$. The rank of the superalgebra $C(p+1)$ is $p+1$ and the dimensions of the various components are

$$\dim L_0 = p(2p+1) + 1,$$
$$\dim L_{\pm 1} = 2p,$$
$$\dim L = 2p^2 + 5p + 1.$$

The Bose subalgebra L_0 is reductive with the U(1) factor K. The Fermi subalgebra L_1 is reductible, $L_1 = L_{+1} \oplus L_{-1}$, and the generator K can be normalized so that

$$[K,L_{\pm 1}] = \pm L_{\pm 1}.$$

The basis in the Cartan subalgebra is usually chosen as

$$h_1 = B_{11} + K,$$
$$h_{j+1} = B_{ii} - B_{i+1\, i+1}, \quad i = 1,2,...,p-1.$$
$$h_{1+p} = B_{pp},$$

The hidden SO(2) generator K is related to the Cartan generators h_j by

$$K = h_1 - \sum_1^p h_{1+i}.$$

(2) An irreducible finite-dimensional representation of $C(p+1)$ is defined by its highest weight Λ. Of course Λ is annihilated by every positive generator and the Kac–Dynkin parameters are the eigenvalues of the Cartan generators h_j for the highest weight Λ:

$$h_j|\Lambda\rangle = a_j|\Lambda\rangle.$$

Similarly,

$$K|\Lambda\rangle = k_\Lambda|\Lambda\rangle$$

and k_Λ is related to the Kac–Dynkin parameters by

$$k_\Lambda = a_1 - \sum_1^p a_{1+i}.$$

We shall use the following notation for an irreducible representation $R(\Lambda)$:

$$R(\Lambda) \Rightarrow \{a_1|a_2,...,a_{1+p}\},$$

where $a_1(k_\Lambda)$ is any complex number and $a_2,...,a_{1+p}$ are non-negative integers. These last parameters describe an irreducible representation of the symplectic group Sp$(2p)$ and it is convenient to introduce the Young tableau $Y(\Lambda)$ of this representation shown in Fig. 1.

TABLE I. Infinitesimal generators of $C(p+1)$.

H	$\{B_{ii}\}$	K		
N^+	$\{B_{ij}\|i<j\}$ $\{C_{ij} = C_{ji}\}$		$\{F_{+j}\}$ $\{G_{+j}\}$	
N^-	$\{B_{ij}\|i>j\}$ $\{D_{ij} = D_{ji}\}$			$\{F_{-j}\}$ $\{G_{-j}\}$
	Sp$(2p)$	SO(2)	L_{+1}	L_{-1}

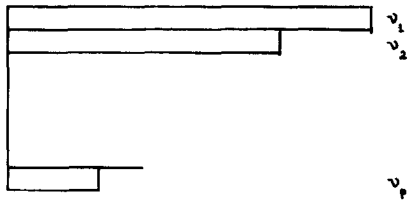


FIG. 1. Young tableau $Y(\Lambda)$.

Here,

$$\nu_j = \sum_j^p a_{1+i}, \quad j = 1,2,...,p. \tag{1}$$

We then have an equivalent description of the highest weight Λ with the Young tableau $Y(\Lambda) = Y(\nu_1,...,\nu_p)$ and the largest eigenvalue k_Λ.

(3) For an irreducible representation $R(\Lambda)$ the construction of the different components associated with the even subalgebra of SO(2) $\otimes$ Sp$(2p)$ is made by applying the odd negative generators F_{-j} and G_{-j} on the highest weight Λ. We then obtain 2^{2p} possible states of the form

$$\prod_{j_1=1}^{j_1=p}\prod_{j_2=1}^{j_2=p} [F_{-j_1}]^{n_{j_1}}[G_{-j_2}]^{n_{j_2}}|\Lambda\rangle, \tag{2}$$

where $n_{j_1}, n_{j_2} = 0$ or 1. The eigenvalue of the generator K for the state (2) is simply given by

$$k_\rho = k_\Lambda - \rho,$$

where

$$\rho = \sum_{j_1} n_{j_1} + \sum_{j_2} n_{j_2}, \quad 0 \leqslant \rho \leqslant 2p.$$

When all the states (2) are coupled to Λ the representation $R(\Lambda)$ is typical. It has 2^{2p} SO(2) $\otimes$ Sp$(2p)$ components distributed in $2p+1$ levels in k,

$$k = k_\Lambda, k_\Lambda - 1,...,k_\Lambda - 2p,$$

and the dimension of the typical $R(\Lambda)$ turns out to be independent of k_Λ and simply related to the dimension of the Sp$(2p)$ Young tableau $Y(\Lambda)$ by

$$\dim R(\Lambda) = 2^{2p} \dim Y(\Lambda). \tag{3}$$

(4) For particular values of k_Λ some states of the set (2) are decoupled and the representation $R(\Lambda)$ is atypical with a dimension less than that given in formula (3).

For the Kac–Dynkin parameter a_1 we have $2p$ possible atypical values

$$a_1 = A_j = \sum_1^j (1 + a_{1+i}),$$

$$a_1 = B_j = \sum_1^p (1 + a_{1+i}) \qquad j = 0,1,...,p-1.$$
$$+ \sum_{j+1}^p (1 + a_{1+i}),$$

The atypical values are non-negative integers in the order

$$A_0 = 0 < A_1 < \cdots < A_{p-1} < B_{p-1} < \cdots < B_1 < B_0,$$

$$B_0 = 2 \sum_1^p (1 + a_{1+i}).$$

Notice that the half-sum

$$\frac{1}{2}(A_j + B_j) = \sum_1^p (1 + a_{1+i})$$

is independent of j and it corresponds to a typical value for a_1.

For the eigenvalue k_Λ and the Young tableau parameters ν_j the atypical situations are the following:

$$\begin{aligned} a_1 = A_j &\Leftrightarrow k_\Lambda + \nu_{1+j} = j, \\ a_1 = B_j &\Leftrightarrow k_\Lambda - \nu_{1+j} = 2p - j, \end{aligned} \quad j = 0,1,...,p-1,$$

and the value $k_\Lambda = p$ is typical.

(5) The construction (2) is equivalent to performing tensor products of the Young tableau $Y(\Lambda)$ by the one column with l boxes $(0<l<p)$ Young tableaux F_l of the fundamental representations of $\mathrm{Sp}(2p)$.[14]

The reduction of an irreducible representation $R(\Lambda)$ of the orthosymplectic group $\mathrm{OSP}(2|2p)$ with respect to the subgroup $\mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ is then written as

$$R(\Lambda) = \bigoplus_{\rho=0}^{\rho=2p} [Y(\Lambda)]_{k_\Lambda} \otimes [\oplus_l F_l]_{-\rho}, \tag{4}$$

where

$$l = L, L-2, ..., L-2n,$$

$$L = \min(\rho, 2-\rho).$$

The sum over l ends at $l=0$ if ρ is even and at $l=1$ if ρ is odd.

The dimension of the fundamental Young tableau F_l is

$$\dim F_l = C_{2p}^l - C_{2p}^{l-2},$$

where C_{2p}^l is as usual the binomial factor

$$C_{2p}^l = (2p)!/l!(2p-l)!.$$

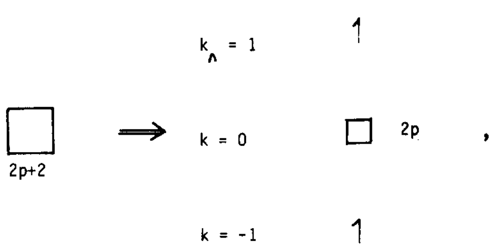
As a consequence the dimension of the level $k = k_\rho$ of a typical $R(\Lambda)$ is simply

$$C_{2p}^\rho \dim Y(\Lambda)$$

and we immediately recover formula (3) for a typical representation.

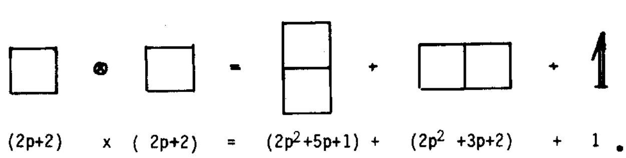
III. METHOD

(1) The method used for the description of the $\mathrm{OSP}(2|2p)$ supertableaux is the tensor product method introduced in previous publications.[6,9] The starting point is the one box supertableau F associated to the fundamental representation $\{1|0\cdots0\}$ of $\mathrm{OSP}(2|2p)$ whose $\mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ components are



where 1 is the singlet representation of $\mathrm{Sp}(2p)$.

(2) The simplest tensor product is

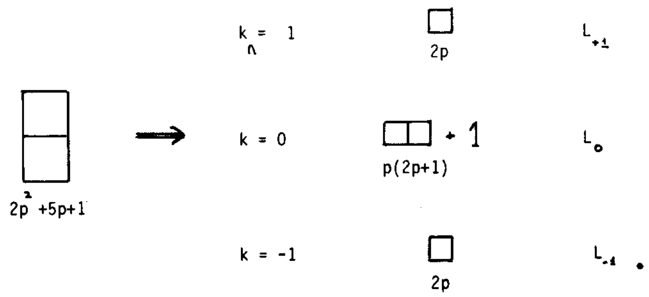


(5)

The dimensions indicated for the supertableaux here and in what follows are computed with the determinant method of Balentekin and Bars.[11]

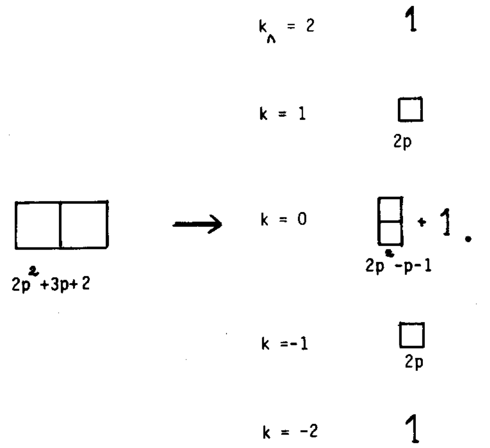
The zero box supertableaux 1 is associated with the singlet atypical representation $\{0|0,...,0\}$ of $\mathrm{OSP}(2|2p)$.

The supertableau with two superantisymmetric boxes is irreducible and it describes the adjoint representation $\{2|1,0,...,0\}$ of $\mathrm{OSP}(2|2)$ with the $\mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ components of the Lie superalgebra $C(p+1)$:



This representation is typical when $p = 1$ and atypical when $p \geqslant 2$.

The supertableau with two supersymmetric boxes is irreducible only when $p \geqslant 2$. The case $p = 1$ has been studied in Ref. 15. For $p \geqslant 2$ it describes the irreducible representation $\{2|0,...,0\}$ of $\mathrm{OSP}(2|2p)$ whose $\mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ components are



This representation is typical when $p = 2$ and atypical when $p \geqslant 3$.

In summary, when $p \geqslant 2$ the tensor product (5) is simply written in terms of irreducible representations

$$\{1|0,...,0\} \otimes \{1|0,...,0\}$$
$$= \{2|1,0,...,0\} \oplus \{2|0,...,0\} \oplus \{0|0,...,0\}.$$

(3) More generally we make the tensor product of an

irreducible supertableau T describing an irreducible typical or atypical representation of OSP$(2|2p)$ for which the SO(2) $\otimes$ Sp$(2p)$ content is known by the one-box supertableau F. We reduce the tensor product, taking into account the invariance of the bilinear even form G. When the subtraction of the invariant subspaces due to the invariance of G cannot be performed, we are in a case of nonfull reductibility for the tensor product and we get nonfully reducible representations to which correspond pairs of atypical supertableaux forming generalized atypical supertableaux. This situation is analogous to that found with the superunitary groups SU$(n|m)$.[6–9]

IV. IRREDUCIBLE SUPERTABLEAUX OF OSP(2|2p) (Refs. 11 and 13)

(1) The class I Young supertableaux of the orthosymplectic group OSP$(2|2p)$ have one row and p columns of arbitrary length and they are conveniently parametrized as indicated in Fig. 2.

(2) The highest weight of the supertableau κ, ν_j is given by a Young tableau of Sp$(2p)$ with p rows of length $\nu_1,\nu_2,\ldots,\nu_p$ and the eigenvalue $k=\kappa$ of the hidden SO(2) generator. The Kac–Dynkin parameters of the supertableau are

$$a_1=\kappa+\nu_1,$$
$$a_{1+i}=\nu_i-\nu_{i+1},\quad i=1,2,\ldots,p-1.$$
$$a_{1+p}=\nu_p,$$

The Kac–Dynkin parameter a_1 can be expressed in terms of the other Kac–Dynkin parameters and the result is

$$a_1=\kappa+\sum_1^p a_{1+i}.$$

We now discuss the three cases $\kappa<p$, $\kappa=p$, and $\kappa>p$.

(3) When $\kappa<p$ we have the supertableau constraints

$$\nu_{1+\kappa}=\cdots=\nu_p=0,$$

which imply the vanishing of $p-\kappa$ Kac–Dynkin parameters

$$a_{1+i}=0,\quad \text{for}\quad i\geqslant\kappa+1.$$

As a consequence the value of a_1 is atypical:

$$a_1=\kappa+\sum_1^\kappa a_{1+i}=\sum_1^\kappa(1+a_{1+i})=A_\kappa.$$

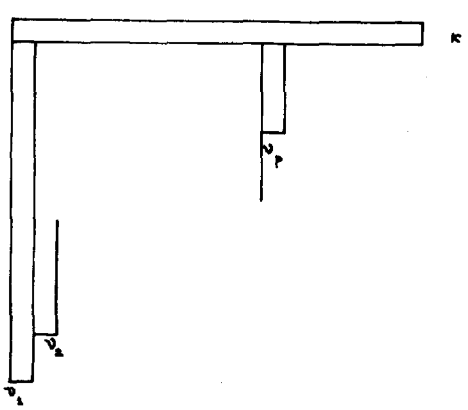


FIG. 2. Supertableau of OSP$(2|2p)$.

The supertableaux $\kappa<p$ are associated to irreducible atypical representations of $C(p+1)$,

$$\{\nu_1+\kappa|\nu_1-\nu_2,\ldots,\nu_{\kappa-1}-\nu_\kappa,\nu_\kappa,0,\ldots,0\}.$$

These representations are self-contragradient and have $2\kappa+1$ levels in the eigenvalues of K.

(4) When $\kappa=p$ the value of a_1 is typical

$$a_1=p+\sum_1^p a_{1+i}=\sum_1^p(1+a_{1+i})=\frac{1}{2}(A_{p-1}+B_{p-1}).$$

The supertableaux $\kappa=p$ are associated to irreducible typical representations of $C(p+1)$

$$\{\nu_1+p|\nu_1-\nu_2,\ldots,\nu_p\}.$$

These representations are self-contragradient with $2p+1$ levels in the eigenvalue of K.

The dimension of the supertableau $\kappa=p$ is

$$2^{2p}N(\nu_1,\ldots,\nu_p),$$

where $N(\nu_1,\ldots,\nu_p)$ is the dimension of the Sp$(2p)$ Young tableau defining the highest weight.

(5) When $\kappa>p$ the Kac–Dynkin parameter a_1 is either typical or atypical of type B_j. If $a_1=B_j$ then $\kappa=\kappa_j$ with

$$\kappa_j=2p-j+\nu_{1+j},\quad j=0,1,\ldots,p-1.$$

At fixed values of $\nu_1,\nu_2,\ldots,\nu_p$, we have p possible atypical values for $\kappa>p$.

(6) Consider the case $\kappa>p$ typical. As a consequence of the reductibility O(2) $\Rightarrow$ SO(2), the typical supertableaux of OSP$(2|2p)$ are associated with a direct sum of two contragradient irreducible typical representations of $C(p+1)$ given by[13]

$$\{\nu_1+\kappa|\nu_1-\nu_2,\ldots,\nu_p\}\oplus\{\nu_1+2p-\kappa|\nu_1-\nu_2,\ldots,\nu_p\}.$$

The two contragradient representations have the same dimension and the dimension of the typical supertableaux $\kappa>p$ is

$$2^{2p+1}N(\nu_1,\ldots,\nu_p).$$

V. GENERALIZED SUPERTABLEAUX OF OSP(2|2p)

(1) We now suppose that the supertableaux T_1 is atypical of type B_j,

$$\kappa^{(1)}=\kappa_j=2p-j+\nu_{1+j}.$$

As a result of the tensor product method T_1 is always simultaneously produced with a second atypical supertableau T_2 and the pair (T_1,T_2), called a generalized atypical supertableau, describes a nonfully reducible representation of $C(p+1)$.

Choosing the notation T_1,T_2 such that $\kappa^{(1)}>\kappa^{(2)}$ we study the relation between T_1 and T_2. For that purpose it is convenient to distinguish the two cases $\nu_{1+j}\geqslant 1$ and $\nu_{1+j}=0$.

(a) Case $\nu_{1+j}\geqslant 1$*:* Let α be the largest non-negative integer such that for the supertableau T_1

$$\nu_{1+j}=\cdots=\nu_{1+j+\alpha},\quad 0\leqslant\alpha\leqslant p-1-j.$$

The supertableau T_2 is obtained from T_1 by suppressing $1+\alpha$ boxes in the first row,

$\kappa^{(2)} = \kappa^{(1)} - (1+\alpha)$,

and $1+\alpha$ boxes in the $(1+\nu_{1+j})$th row,

$\nu_i \Rightarrow \nu_i - 1, \quad \text{for} \quad 1+j \leqslant i \leqslant 1+j+\alpha,$

the other parameters ν_i remaining as for T_1.

The value $\kappa^{(2)}$ is also written

$$\kappa^{(2)} = 2p - j + \nu_{1+j} - (1+\alpha)$$
$$= 2p - (j+\alpha) + \nu_{1+j+\alpha} - 1$$

and the supertableau T_2 is atypical of type $B_{j+\alpha}$. In this case we have

$$\kappa^{(1)} > \kappa^{(2)} > p.$$

An example with $\alpha = 0$ is shown in Fig. 3.

(b) Case $\nu_{1+j} = 0$: The supertableau T_2 is obtained from T_1 by suppressing $2(p-j)$ boxes in the first row,

$$\kappa^{(1)} = 2p - j, \quad \kappa^{(2)} = j,$$

and the parameters $\nu_1, \nu_2, ..., \nu_j$ are identical for T_1 and T_2.

The supertableau T_2 is atypical of type A_j. In this case we have

$$\kappa^{(1)} > p > \kappa^{(2)}.$$

A generalized atypical supertableau of this category is represented in Fig. 4.

(2) The generalized atypical supertableaux $\kappa^{(1)} > \kappa^{(2)} > p$ are associated with a direct sum of two contragradient non-fully reducible atypical representations of $C(p+1)$ with four atypical components each,

$$[C_1 + 2C_2 + C_3] \oplus [\overline{C}_3 + 2\overline{C}_2 + \overline{C}_1].$$

The Kac–Dynkin parameters of the components C_1 and C_2 are, respectively, those of the supertableaux T_1 and T_2. For the component C_3 these parameters are those of the atypical supertableau T_3 obtained from T_2 as T_2 itself has been obtained from T_1. Let us discuss separately the cases $\nu_{1+j} \geqslant 2$ and $\nu_{1+j} = 1$, the parameter ν_{1+j} referring to the supertableau T_1.

(a) Case $\nu_{1+j} \geqslant 2$: We define as α and β the smallest nonnegative integers such that

$$\begin{matrix} \nu_{1+j} - \nu_{2+j+\alpha} \geqslant 1, \\ \nu_{1+j} - \nu_{2+j+\beta} \geqslant 2, \end{matrix} \quad 0 \leqslant \alpha \leqslant \beta \leqslant p - 1 - j.$$

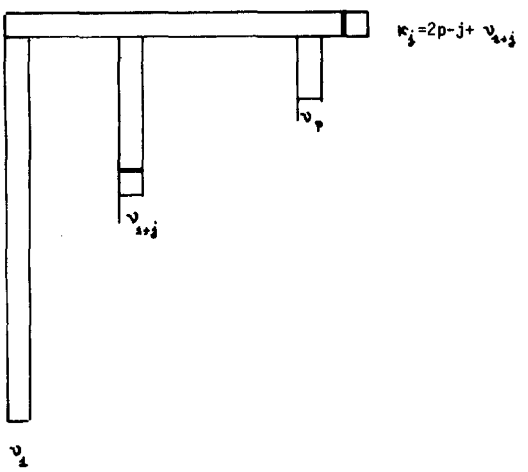


FIG. 3. Generalized atypical supertableau $\kappa^{(1)} > \kappa^{(2)} > p$.

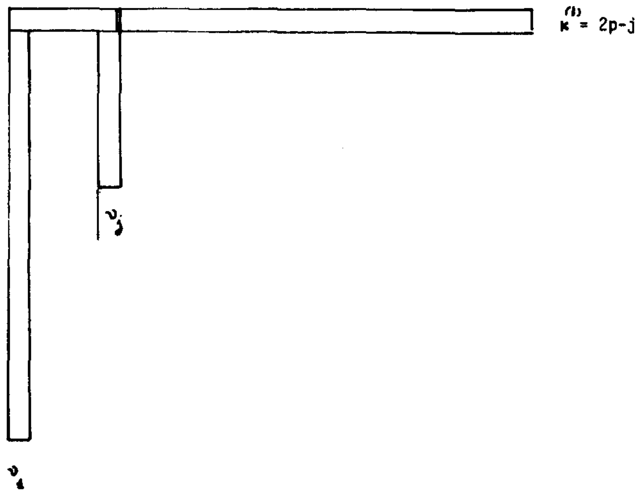


FIG. 4. Generalized atypical supertableau $\kappa^{(1)} > p > \kappa^{(2)}$.

(a-1) $\alpha < \beta$: We have $a_{2+j+\alpha} = 1, a_{2+j+\beta} \geqslant 1$. For $j > 0$ the components C_1, C_2, and C_3 are the following:

$C_1\{a_1 | a_2,...,a_{1+j},0,...,0,a_{2+j+\alpha} = 1,0,...,0,$
$a_{2+j+\beta},...,a_{1+p}\}$,

$C_2\{a_1 - 1 - \alpha | a_2,...,a_{1+j} + 1,0,...,0,0,0,...,0,$
$a_{2+j+\beta},...,a_{1+p}\}$,

$C_3\{a_1 - 2 - \beta | a_2,...,a_{1+j} + 2,0,...,0,0,0,...,0,$
$a_{2+j+\beta} - 1,...,a_{1+p}\}$,

where the Kac–Dynkin parameter a_1 of C_1 is given by

$$a_1 = 2p - j + \sum_1^j a_{1+i} + 2 + 2 \sum_{1+j+\beta}^{p} a_{1+i}.$$

The atypicities of the components C_1, C_2, C_3, are, respectively, $B_j, B_{j+\alpha}, B_{j+\beta}$.

(a-2) $\alpha = \beta$: We have $a_{2+j+\alpha} \geqslant 2$. For $j > 0$ the components C_1, C_2, and C_3 are the following:

$C_1\{a_1 | a_2,...,a_{1+j},0,...,0,a_{2+j+\alpha},...,a_{1+p}\}$,

$C_2\{a_1 - 1 - \alpha | a_2,...,a_{1+j} + 1,0,...,0,$
$a_{2+j+\alpha} - 1,...,a_{1+p}\}$,

$C_3\{a_1 - 2 - \alpha | a_2,...,a_{1+j} + 2,0,...,0,$
$a_{2+j+\alpha} - 2,...,a_{1+p}\}$,

where the Kac–Dynkin parameter a_1 of C_1 is given by

$$a_1 = 2p - j + \sum_1^j a_{1+i} + 2 \sum_{1+j+\alpha}^{p} a_{1+i}.$$

The atypicities of the components C_1, C_2, C_3 are, respectively, $B_j, B_{j+\alpha}, B_{j+\alpha}$.

(b) Case $\nu_{1+j} = 1$: Using the same parameter α as previously, we have $a_{2+j+\alpha} = 1$ and

$$a_{1+i} = 0, \quad \text{for} \quad i \geqslant 2 + j + \alpha.$$

For $j > 0$ the components C_1, C_2, and C_3 are the following:

$C_1\{2p - j + 1 + \nu_1 | a_2,...,a_{1+j},0,...,0,a_{2+j+\alpha} = 1,0,...,0\}$,

$C_2\{2\pi - j - \alpha + \nu_1 | a_2,...,a_{1+j} + 1,0,...,0,0,0,...,0\}$

$C_3\{j + \alpha + \nu_1 | a_2,...,a_{1+j} + 2,0,...,0,0,0,...,0\}$, where the parameter ν_1 of the first column of T_1 is given by

$$\nu_1 = 1 + \sum_{1}^{j} a_{1+i}.$$

The atypicities of the components C_1, C_2, C_3 are, respectively, $B_j, B_{j+\alpha}, A_{j+\alpha}$. Let us notice that, in this case, the component C_3 is self-contragradient, $\overline{C}_3 = C_3$.

(3) The contragradient components $\overline{C}_3, \overline{C}_2, \overline{C}_1$ are obtained by the general relations given in the Appendix A. We define as α and $\overline{\alpha}$ the two largest non-negative integers such that

$$\nu_{1+j-\overline{\alpha}} = \cdots = \nu_{1+j} = \cdots = \nu_{1+j+\alpha},$$
$$0 \leqslant \overline{\alpha} \leqslant j, \quad 0 \leqslant \alpha \leqslant p-1-j,$$

and we study separately the two cases $\overline{\alpha} < j$ and $\overline{\alpha} = j$ always for $j > 0$.

(a) Case $\overline{\alpha} < j$: We have $a_{1+j-\overline{\alpha}} \geqslant 1$ and $a_{2+j+\alpha} \geqslant 1$. For $j > 0$ the components $\overline{C}_3$, $\overline{C}_2$, and $\overline{C}_1$ are the following:

$$\overline{C}_3\{a_1 + 1 + \alpha | a_2, \ldots, a_{1+j-\overline{\alpha}}, 0, \ldots, 0,$$
$$a_{1+j} = 1, 0, \ldots, 0, a_{2+j+\alpha} - 1, \ldots, a_{1+p}\},$$
$$\overline{C}_2\{a_1 | a_2, \ldots, a_{1+j-\overline{\alpha}}, 0, \ldots, 0, 0, 0, \ldots, 0,$$
$$a_{2+j+\alpha}, \ldots, a_{1+p}\},$$
$$\overline{C}_1\{a_1 - 1 - \overline{\alpha} | a_2, \ldots, a_{1+j-\overline{\alpha}} - 1, 0, \ldots, 0, 0,$$
$$a_{2+j} = 1, 0, \ldots, 0, a_{2+j+\alpha}, \ldots, a_{1+p}\},$$

where the Kac–Dynkin parameter a_1 of $\overline{C}_2$ is given by

$$a_1 = j + \sum_{1}^{j-\overline{\alpha}} a_{1+i}.$$

The atypicities of the components $\overline{C}_3, \overline{C}_2, \overline{C}_1$ are, respectively, $A_{j+\alpha}, A_j, A_{j-\overline{\alpha}}$.

(b) Case $\overline{\alpha} = j$: We have $a_{2+j+\alpha} \geqslant 1$ and $a_{1+i} = 0$ for $1 \leqslant i \leqslant j + \alpha$.

For $j > 0$ the components $\overline{C}_3, \overline{C}_2, \overline{C}_1$ are the following:

$$\overline{C}_3\{j + 1 + \alpha | 0, \ldots, 0, a_{1+j} = 1, 0, \ldots, 0,$$
$$a_{2+j+\alpha} - 1, \ldots, a_{1+p}\},$$
$$\overline{C}_2\{j | 0, \ldots, 0, 0, 0, \ldots, 0, a_{2+j+\alpha}, \ldots, a_{1+p}\},$$
$$\overline{C}_1\{0 | 0, \ldots, 0, 0, a_{2+j} = 1, 0, \ldots, 0, a_{2+j+\alpha}, \ldots, a_{1+p}\}.$$

The atypicities of the components $\overline{C}_3, \overline{C}_2, \overline{C}_1$, are, respectively, $A_{j+\alpha}, A_j, A_0$.

We easily check that when $\nu_{1+j} = 1$ for the supertableau T_1 the component $\overline{C}_3$ is self-contragradient and it reduces to the component C_3 previously determined in the case (b) $\nu_{1+j} = 1$.

(4) Now we discuss the case $j = 0$. The general formulas given in the subsections (2) and (3) have to be slightly modified. Of course here we have $\overline{\alpha} = 0$ and we distinguish three cases.

(a) Case $\nu_i \geqslant 2$, $\alpha < \beta$:

$C_1\{a_1 | 0, \ldots, 0, a_{2+\alpha} = 1, 0, \ldots, 0, a_{2+\beta}, \ldots, a_{1+p}\}$,
atypicity B_0,
$C_2\{a_1 - 2 - \alpha | 0, \ldots, 0, 0, 0, \ldots, 0, a_{2+\beta}, \ldots, a_{1+p}\}$,
atypicity B_α,
$C_3\{a_1 - 4 - \beta | 0, \ldots, 0, 0, 0, \ldots, 0, a_{2+\beta} - 1, \ldots, a_{1+p}\}$,
atypicity B_β,
$\overline{C}_3\{\alpha | 0, \ldots, 0, 0, 0, \ldots, 0, a_{2+\beta}, \ldots, a_{1+p}\}$,
atypicity A_α,
$\overline{C}_2\{0 | 0, \ldots, 0, a_{2+\alpha} = 1, 0, \ldots, 0, a_{2+\beta}, \ldots, a_{1+p}\}$,
atypicity A_0,
$\overline{C}_1\{0 | 1, \ldots, 0, a_{2+\alpha} = 1, 0, \ldots, 0, a_{2+\beta}, \ldots, a_{1+p}\}$,
atypicity A_0,

where the Kac–Dynkin parameters a_1 of the component C_1 is given by

$$a_1 = 2p + 2 + 2 \sum_{1+\beta}^{p} a_{1+i}.$$

(b) Case $\nu_i \geqslant 2$, $\alpha = \beta$:

$C_1\{a_1 | 0, \ldots, 0, a_{2+\alpha}, \ldots, a_{1+p}\}$,	atypicity B_0,
$C_2\{a_1 - 2 - \alpha | 0, \ldots, 0, a_{2+\alpha} - 1, \ldots, a_{1+p}\}$,	atypicity B_α,
$C_3\{a_1 - 4 - \alpha | 0, \ldots, 0, a_{2+\alpha} - 2, \ldots, a_{1+p}\}$,	atypicity B_α,
$\overline{C}_3\{\alpha | 0, \ldots, 0, a_{2+\alpha} - 1, \ldots, a_{1+p}\}$,	atypicity A_α,
$\overline{C}_2\{0 | 0, \ldots, 0, a_{2+\alpha}, \ldots, a_{1+p}\}$,	atypicity A_0,
$\overline{C}_1\{0 | 1, 0, \ldots, 0, a_{2+\alpha}, \ldots, a_{1+p}\}$,	atypicity A_0,

where the Kac–Dynkin parameters a_1 of the component C_1 is given by

$$a_1 = 2p + 2 \sum_{1+\alpha}^{p} (a_{1+i}).$$

(c) Case $\nu_i = 1$:

$C_1\{2p + 2 | 0, \ldots, 0, a_{2+\alpha} = 1, 0, \ldots, 0\}$,	atypicity B_0,
$C_2\{2p - \alpha | 0, \ldots, 0, 0, 0, \ldots, 0\}$,	atypicity B_α,
$C_3 = \overline{C}_3\{\alpha | 0, \ldots, 0, 0, 0, \ldots, 0\}$,	atypicity A_α,
$\overline{C}_2\{0 | 0, \ldots, 0, a_{2+\alpha} = 1, 0, \ldots, 0\}$,	atypicity A_0,
$\overline{C}_1\{0 | 0, \ldots, 0, a_{2+\alpha} = 1, 0, \ldots, 0\}$,	atypicity A_0.

(5) The generalized atypical supertableaux $\kappa^1 > p > \kappa^2$ are associated to self-contragradient nonfully reducible atypical representations of $C(p+1)$ with four atypical components

$$[C_1 + 2C_2 + \overline{C}_1].$$

Of course, the atypical component C_2 is self-contragradient, $\overline{C}_2 = C_2$. The Kac–Dynkin parameters of the components C_1 and C_2 are those of the supertableaux T_1 and T_2. In order to determine the component $\overline{C}_1$ we use the same parameter $\overline{\alpha}$ as previously, which is here the largest non-negative integer such that

$$\nu_{1+j-\overline{\alpha}} = 0, \quad 0 \leqslant \overline{\alpha} \leqslant j.$$

It is convenient to study separately the two cases $\overline{\alpha} < j$ and $\overline{\alpha} = j$.

(a) Case $\overline{\alpha} < j$: we have $a_{1+j-\overline{\alpha}} \geqslant 1$, $a_{1+i} = 0$ for $i \geqslant 1 + j - \overline{\alpha}$. The component $C_1, C_2, \overline{C}_1$ are the following:

$$C_1\{2p - j + \nu_1 | a_2, \ldots, a_{1+j-\overline{\alpha}}, 0, \ldots, 0\},$$
$$C_2\{j + \nu_1 | a_2, \ldots, a_{1+j-\overline{\alpha}}, 0, \ldots, 0\},$$
$$\overline{C}_1\{j - 1 - \overline{\alpha} + \nu_1 | a_2, \ldots, a_{1+j-\overline{\alpha}} - 1, 0, \ldots, a_{2+j} = 1, 0, 0\},$$

where the parameter ν_1 of the supertableau T_1 is given by

$$\nu_1 = \sum_1^{j-\bar{\alpha}} a_{1+i}.$$

The atypicity of the components C_1, C_2, and $\overline{C}_1$ are, respectively, B_j, A_j, and $A_{j-\bar{\alpha}}$. The dimension of those generalized atypical supertableaux is

$$2^{2p+1}N(\nu_1,...,\nu_{j-\bar{\alpha}},0,...,0).$$

(b) Case $\bar{\alpha}=j$: The supertableaux T_1 and T_2 have only one row and the atypical components $C_1, C_2, \overline{C}_1$ are the following:

$$C_1\{2p-j|0,...,0\},$$
$$C_2\{j|0,...,0\},$$
$$\overline{C}_1\{0|0,...,0,a_{2+j}=1,0,0\}.$$

The atypicity of the components C_1, C_2, and $\overline{C}_1$ are, respectively, B_j, A_j, and A_0. The dimension of these generalized atypical supertableaux is simply 2^{2p+1}.

VI. TOPOLOGY OF THE SET T OF SUPERTABLEAUX

(1) Let us define as S the set of representations of $C(p+1)$, where the Kac–Dynkin parameter a_1 is an algebraic integer. In the tensor product

$$R_1 \otimes R_2 = \bigoplus_j R_j,$$

if R_1 and $R_2 \in S$ then $R_j \in S$ for all j's. Of course the set S contains irreducible typical, irreducible atypical, and nonfully reducible atypical representations.

(2) The set T of supertableaux of OSP$(2|2p)$ describe self-contragradient irreducible, fully reducible, and nonfully reducible representations of the set S.

(3) The $(2p+2)\times(2p+2)$ matrices commuting with all orthosymplectic matrices of OSP$(2|2p)$ are, by the Schur lemma proportional to the unit matrix I_{2p+2}. The only possibilities on the field of real numbers are I_{2p+2} and $-I_{2p+2}$. Therefore the center of the orthosymplectic group OSP$(2|2p)$ is isomorphic to Z_2 and the two groups OSP$(2|2p)$ and OSP$(2|2p)/Z_2$ are locally isomorphic.

(4) As a first consequence the set S of representations of OSP$(2|2p)$ can be divided into two classes S_A and S_B by using a linear combination C of the Kac–Dynkin parameters defined by

$$C = \sum_1^{p+1} j a_j.$$

For the class S_A,C is even and for the class S_B,C is odd.

(5) As a second consequence the set T of supertableaux of OSP$(2|2p)$ can also be divided into two classes T_A and T_B by using now a parameter N counting the total number of boxes of a supertableau of T.

$$N = \kappa + \sum_1^p \nu_j.$$

For the class T_A,N is even and for the class T_B,N is odd. Of course the class $T_A\,(T_B)$ supertableaux are associated to self-contragradient representations of the class $S_A\,(S_B)$.

(6) The supertableaux of the set T are generated, by tensor product, from the one box supertableau of the fundamental representation:

$$\square \in T_B, \quad \{1|0,...,0\} \in S_B.$$

The supertableaux of the class T_A are the supertableaux of the factor group OSP$(2|2p)/Z_2$ and they are generated, by tensor product, from the two-box supertableau of the adjoint representation.

$$\begin{array}{|c|}\hline \\ \hline \\ \hline\end{array} \in T_A, \quad \{2|1,0,...,0\} \in S_A.$$

(7) Let us remark that, by construction, the atypical supertableaux T_1 and T_2 forming a generalized atypical supertableau differ by an even number of boxes and therefore T_1 and T_2 belong to the same class. As a consequence the notion of class applies to generalized atypical supertableaux.

VII. CONCLUDING REMARKS

We have studied the Young supertableaux of the orthosymplectic groups OSP$(2|2p)$, $p \geqslant 2$, and we have determined the relation between these supertableaux and some self-contragradient representations of $C(p+1)$: (i) irreducible typical and atypical; (ii) fully reducible typical; and (iii) nonfully reducible atypical.

As previously in the case of superunitary groups only a subset of representations of $C(p+1)$ is described by supertableaux. Of course, this subset is closed under the tensor product and it possesses a structure in two classes.

The supertableaux of the orthosymplectic groups OSP$(m|2p)m \geqslant 3$ can be analyzed with the same techniques. The simple case of irreducible supertableaux has been treated by Farmer and Jarvis.[13] That of generalized atypical supertableaux is nontrivial especially with the appearance of degeneracies in the spectrum of the atypical values.[7,9] This work is now in progress.

ACKNOWLEDGMENTS

The author takes this opportunity to thank Guido Altarelli, Nicolá Cabibbo, and Luciano Maiani for the hospitality extended to him at the University of Rome where part of this work has been done.

APPENDIX A: CONTRAGRADIENT IRREDUCIBLE REPRESENTATIONS

As a result of the tensor product method we are able to determine the lowest weight of an irreducible representation R of $C(p+1)$ or, equivalently, the highest weight of the contragradient representation $\overline{R}$ of R.

It is convenient to describe the highest weight of $R(\overline{R})$ by a Young tableau $Y(\overline{Y})$ of the symplectic group Sp$(2p)$ and the largest eigenvalue $k(\bar{k})$ of the SO(2) generator.

1. R and $\overline{R}$ are typical

This case is well known and we simply have[13]

$$\overline{Y} = Y, \quad \bar{k} + k = 2p.$$

In the reduction OSP$(2|2p) \Rightarrow$ SO(2) $\otimes$ Sp$(2p)$ the typical representations R and $\overline{R}$ have $2p+1$ levels in k. In the description with Kac–Dynkin parameters we get

$$R \Rightarrow \{a_1|a_2,...,a_{1+p}\},$$
$$\bar{R} \Rightarrow \{\bar{a}_1|a_2,...,a_{1+p}\},$$

with the relation

$$a_1 + \bar{a}_1 = 2p + 2\sum_1^p a_{1+i}.$$

The particular case of a self-contragradient typical representation is obtained when $k = \bar{k} = p$.

2. R is atypical of type B_j

We have

$$k_\Lambda = 2p - j + \nu_{1+j}.$$

Let us define for the Young tableau Y the largest non-negative integer $\bar{\alpha}$ such that

$$\nu_{1+j-\bar{\alpha}} = \cdots = \nu_{1+j}, \quad 0 \leqslant \bar{\alpha} \leqslant j.$$

Then the contragradient representation $\bar{R}$ is atypical of type $A_{j-\bar{\alpha}}$ and the highest weight of $\bar{R}$ is defined by

$$\bar{Y}\begin{cases}\bar{\nu}_i = \nu_i + 1, & \text{for } 1 + j - \bar{\alpha} \leqslant i \leqslant 1 + j,\\ \bar{\nu}_i = \nu_1, & \text{for the other } i\text{'s},\end{cases}$$
$$\bar{k}_{\bar{\Lambda}} = 2p - k_\Lambda - (1 + \bar{\alpha}).$$

In the reduction $\mathrm{OSP}(2|2p) \Rightarrow \mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ the two contragradient representations R and $\bar{R}$ have $2p - \bar{\alpha}$ levels in k.

For the description of R and $\bar{R}$ with Kac–Dynkin parameters it is convenient to distinguish the two cases $\bar{\alpha} < j$ and $\bar{\alpha} = j$.

(a) Case $\bar{\alpha} < j$:

$$R \ \{a_1|\cdots a_{1+j-\bar{\alpha}},0,...,0,a_{2+j},...,a_{1+p}\},$$
$$\bar{R} \ \{\bar{a}_1|\cdots a_{1+j-\bar{\alpha}} - 1,0,...,a_{2+j} + 1,...,a_{1+p}\},$$

where

$$a_1 = 2p - j + \sum_1^{j-\bar{\alpha}} a_{1+i} + 2\sum_{1+j}^p a_{1+i},$$
$$\bar{a}_1 = j - \bar{\alpha} - 1 + \sum_1^{j-\bar{\alpha}} a_{1+i}.$$

(b) Case $\bar{\alpha} = j$:

$$R \ \{a_1|0,...,0,a_{2+j},...,a_{1+p}\},$$
$$R \ \{0|0,...,0,a_{2+j} + 1,...,a_{1+p}\},$$

where

$$a_1 = 2p - j + 2\sum_{1+j}^p a_{1+i}.$$

3. R is atypical of type A_j

We have

$$k_\Lambda = j - \nu_{1+j}.$$

When $\nu_{1+j} = 0$ the representation R is self-contragradient and in the reduction $\mathrm{OSP}(2|2p) \Rightarrow \mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ it contains $2j + 1$ levels in k.

When $\nu_{1+j} \geqslant 1$ we define, for the Young tableau Y, the largest non-negative integer α such that

$$\nu_{1+j} = \cdots = \nu_{1+j+\alpha}, \quad 0 \leqslant \alpha \leqslant p - 1 - j.$$

Then the contragradient representation $\bar{R}$ is atypical of type $B_{j+\alpha}$ and the highest weight of $\bar{R}$ is defined by

$$\bar{Y}\begin{cases}\bar{\nu}_i = \nu_i - 1, & \text{for } 1 + j \leqslant i \leqslant 1 + j + \alpha,\\ \bar{\nu}_i = \nu_i, & \text{for the other } i\text{'s},\end{cases}$$
$$\bar{k}_{\bar{\Lambda}} = 2p - k_\Lambda - (1 + \alpha).$$

In the reduction $\mathrm{OSP}(2|2p) \Rightarrow \mathrm{SO}(2) \otimes \mathrm{Sp}(2p)$ the two contragradient representations R and $\bar{R}$ have $2p - \alpha$ levels in k.

For the description of R and $\bar{R}$ with Kac–Dynkin parameters it is convenient to distinguish the two cases $j > 0$ and $j = 0$.

(a) Case $j > 0$:

$$R \ \{a_1|a_2,...,a_{1+j},0,...,0,a_{2+j+\alpha},...,a_{1+p}\},$$
$$R \ \{\bar{a}_1|a_2,...,a_{1+j} + 1,0,...,0,a_{2+j+\alpha} - 1,...,a_{1+p}\},$$

where

$$a_1 = j + \sum_1^j a_{1+i},$$
$$\bar{a}_1 = 2p - j - \alpha - 1 + \sum_1^j a_{j+i} + 2\sum_{1+j+\alpha}^p a_{1+i}.$$

(b) Case $j=0$:

$$R \ \{0|0,...,0,a_{2+\alpha},...,a_{1+p}\},$$
$$\bar{R} \ \{\bar{a}_1|0,...,0,a_{2+\alpha} - 1,...,a_{1+p}\},$$

where

$$\bar{a}_1 = 2p - \alpha + 2\sum_{1+\alpha}^p a_{1+i}.$$

Of course the formulas of subsections 2 and 3 are complementary.

APPENDIX B: GENERALIZED ATYPICAL SUPERTABLEAUX OF OSP(2|4)

As an illustration of the general expressions given in Sec. V we give the complete list of the generalized atypical supertableaux of OSP(2|4) with their atypical components. We have two possible atypicities for the supertableau T_1.

(i) Atypicity B_0, $\kappa^{(1)} = 4 + \nu_1$.

The parameter $\bar{\alpha}$ is zero and for the parameters α and β we have $0 \leqslant \alpha \leqslant \beta \leqslant 1$.

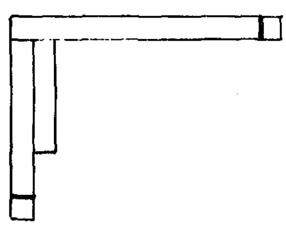
(ii) Atypicity B_1, $\kappa^{(1)} = 3 + \nu_2$.

The parameters α and β are zero and the parameter $\bar{\alpha}$ can take two values $0 \leqslant \bar{\alpha} \leqslant 1$.

1. The supertableau T_1 has the atypicity B_0

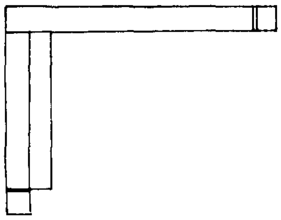
We have six possible cases.

(a) $\nu_1 \geqslant 2$, $\nu_1 - \nu_2 \geqslant 2$, $\alpha = \beta = 0$:



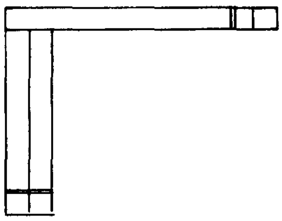
$$[\{2\nu_1+4|\nu_1-\nu_2,\nu_2\}+2\{2\nu_1+2|\nu_1-\nu_2-1,\nu_2\}+\{2\nu_1|\nu_1-\nu_2-2,\nu_2\}]$$
$$\oplus[\{0|\nu_1-\nu_2-1,\nu_2\}+2\{0|\nu_1-\nu_2,\nu_2\}+\{0|\nu_1-\nu_2+1,\nu_2\}].$$

(b) $\nu_1 \geqslant 2,\ \nu_1-\nu_2=1,\ \alpha=0,\ \beta=1$:



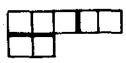
$$[\{2\nu_1+4|1,\nu_1-1\}+2\{2\nu_1+2|0,\nu_1-1\}+\{2\nu_1-1|0,\nu_1-2\}]\oplus[\{0|0,\nu_1-1\}+2\{0|1,\nu_1-1\}+\{0|2,\nu_1-1\}.$$

(c) $\nu_1 \geqslant 2,\ \nu_1=\nu_2,\ \alpha=\beta=1$:



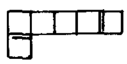
$$[\{2\nu_1+4|0,\nu_1\}+2\{2\nu_1+1|0,\nu_1-1\}+\{2\nu_1-2|0,\nu_1-2\}]\oplus[\{1|0,\nu_1-1\}+2\{0|0,\nu_1\}+\{0|1,\nu_1\}].$$

(d) $\nu_1=1=\nu_2,\ \alpha=1$:



$$[\{6|0,1\}_{70}+2\{3|0,0\}_{10}+\{1|0,0\}_6]_{96}\oplus[\{1|0,0\}_6+2\{0|0,1\}_{10}+\{0|1,1\}_{70}]_{96}.$$

(e) $\nu_1=1,\ \nu_2=0,\ \alpha=0$:



$$[\{6|1,0\}_{49}+2\{4|0,0\}_{15}+\{0|0,0\}_1]_{80}\oplus[\{0|0,0\}_1+2\{0|1,0\}_{15}+\{0|2,0\}_{49}]_{80}.$$

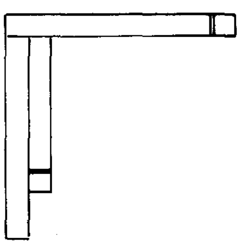
(f) $\nu_1=0$:

$$[\{4|0,0\}_{15}+2\{0(|0,0\}_1+\{0|1,0\}_{15}]_{32}.$$

2. The suptertableau T has the atypicity B_1

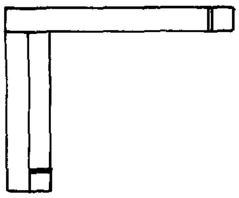
We have again six possible cases.

(a) $\nu_2 \geqslant 2,\ \nu_1-\nu_2 \geqslant 1,\ \bar{\alpha}=0$:



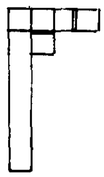
$$[\{\nu_1+\nu_2+3|\nu_1-\nu_2,\nu_2\}+2\{\nu_1+\nu_2+2|\nu_1-\nu_2+1,\nu_2-1\}+\{\nu_1+\nu_2+1|\nu_1-\nu_2+2,\nu_2-2\}]$$
$$\oplus[\{\nu_1-\nu_2+2|\nu_1-\nu_2+1,\nu_2+1\}+2\{\nu_1-\nu_2+1|\nu_1-\nu_2,\nu_2\}+\{\nu_1-\nu_2|\nu_1-\nu_2+1,\nu_2-1\}].$$

(b) $\nu_2 > 2,\ \nu_1 = \nu_2,\ \bar{\alpha} = 1$:



$$[\{2\nu_2+3|0,\nu_2\}+2\{2\nu_2+2|1,\nu_2-1\}+\{2\nu_2+1|2,\nu_2-2\}]\oplus[\{2|1,\nu_2-1\}+2\{1|0,\nu_2\}+\{0|0,\nu_2+1\}].$$

(c) $\nu_2 = 1,\ \nu_1 - \nu_2 > 1,\ \bar{\alpha} = 0$:

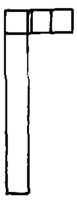


$$[\{\nu_1+4|\nu_1-1,1\}+2\{\nu_1+3|\nu_1,0\}+\{\nu_1+1|\nu_1,0\}]\oplus[\{\nu_1+1|\nu_1,0\}+2\{\nu_1|\nu_1-1,1\}+\{\nu_1-1|\nu_1-2,2\}].$$

(d) $\nu_2 = 1 = \nu_1,\ \bar{\alpha} = 1$:

$$[\{5|0,1\}_{35}+2\{4|1,0\}_{45}+\{2|1,0\}_{19}]_{144}\oplus[\{2|1,0\}_{19}+2\{1|0,1\}_{45}+\{0|0,2\}_{35}]_{144}.$$

(e) $\nu_2 = 0,\ \nu_1 = 1,\ \bar{\alpha} = 0$:



$$[\{3+\nu_1|\nu_1,0\}+2\{1+\nu_1|\nu_1,0\}+\{\nu_1|\nu_1-1,0\}].$$

(f) $\nu_1 = \nu_2 = 0,\ \bar{\alpha} = 1$:

$$[\{3|0,0\}_{10}+2\{1|0,0\}_{6}+\{0|0,1\}_{10}]_{32}.$$

[1] L. Corwin, Y. Ne'eman, and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[2] A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975).
[3] V. C. Kac, Adv. Math. **26**, 8 (1977); Commun. Math. Phys. **53**, 31 (1977); *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), p. 597.
[4] P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).
[5] W. Nahm and M. Scheunert, J. Math. Phys. **17**, 868 (1976).
[6] F. Delduc and M. Gourdin, J. Math. Phys. **25**, 1681 (1984).
[7] F. Delduc and M. Gourdin, J. Math. Phys. **26**, 1865 (1985).
[8] M. Gourdin, "Naturally atypical supertableaux of the superunitary group SU(3/2)," PAR LPTHE 84-26, 1984.
[9] M. Gourdin, "Naturally typical supertableaux of the superunitary group SU(3/2)," PAR LPTHE 84-31, 1984.
[10] P. H. Dondi and P. D. Jarvis, Z. Phys. C **4**, 201 (1980); J. Phys. A Math. Gen. **14**, 547 (1981).
[11] A. B. Balentekin and I. Bars, J. Math. Phys. **22**, 1149, 1810 (1981).
[12] R. C. King, J. Phys. A Math. Gen. **8**, 429 (1975); *Lecture Notes in Physics*, Vol. 180 (Springer, Berlin, 1983), p. 41.
[13] R. J. Farmer and P. D. Jarvis, J. Phys. A Math. Gen. **17**, 2365 (1984).
[14] J. P. Hurni, "Young supertableaux of Lie superalgebras," Preprint UGVA-DPT 1984/04. 420 (Genéve).
[15] M. Gourdin, J. Math. Phys. **27**, 61 (1986).

Principal five-dimensional subalgebras of Lie superalgebras

Joris Van der Jeugt[a)]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B9000 Gent, Belgium

(Received 27 March 1986; accepted for publication 6 August 1986)

The analog of sl(2) for Lie superalgebras is osp(1,2), a five-dimensional superalgebra. All basic classical Lie superalgebras L that contain a principal five-dimensional osp(1,2) subalgebra are classified. Moreover, the decomposition of the standard representation and of the adjoint representation of L into irreducible components of the principal osp(1,2) subalgebra is given.

I. INTRODUCTION

Principal three-dimensional subalgebras of Lie algebras have turned out to be important in many physical models.[1] For instance, let $b_{j,m}^{(+)}$ $(m = -j, -j+1,..., +j)$ be a set of boson creation and annihilation operators with angular momentum number j and projection m. Then, the algebra acting in the boson space and consisting of quadratic products leaving the total number of bosons invariant is $u(2j+1)$. If j is an integer, this algebra contains $so(2j+1)$. The so(3) subalgebra describing the physical angular momentum of the system in the chain $so(3) \subset so(2j+1) \subset u(2j+1)$ $(j \in \mathbb{N})$ is the principal three-dimensional subalgebra of $so(2j+1)$ and $u(2j+1)$.

From the mathematical point of view, principal three-dimensional subalgebras have been discussed by Dynkin[2] and Kostant.[3] Important mathematical applications were the combinatorial results obtained by Hughes,[4] and later generalized by Stanley.[5]

In the present paper, we investigate principal subalgebras of the basic classical Lie superalgebras. The superalgebra corresponding with the three-dimensional Lie algebra sl(2) is the five-dimensional Lie superalgebra osp(1,2), sometimes denoted by $B(0,1)$. Basic classical Lie superalgebras were classified by Kac[6] and are of type $A(m,n)$, $B(m,n)$ $C(n)$, $D(m,n)$, $G(3)$, $F(4)$, or $D(2,1;\alpha)$. Throughout this paper, we shall use the notation of Ref. 6. It is no longer true that every basic classical Lie superalgebra contains a principal five-dimensional subalgebra. This fact was already observed by Stanley,[7] but although he finds only one general class of superalgebras containing a principal osp(1,2), the series $A(n+1,n)$, we prove in this paper that several classes of orthosymplectic Lie superalgebras also have a principal five-dimensional subalgebra.

The structure of the paper is as follows. In Sec. II, irreducible representations of osp(1,2) are analyzed, and they give rise to certain inclusion relations of osp(1,2) into special linear and orthosymplectic Lie superalgebras. With the definition of a principal five-dimensional subalgebra in Sec. III, it is easy to see that the inclusion relations of Sec. II are principal. Then, a necessary condition for the existence of a principal osp(1,2) subalgebra is given and analyzed for all the basic classical Lie superalgebras. In Sec. IV, we give a realization of the principal five-dimensional subalgebra in the standard representation for all the classes of Lie superalgebras satisfying the condition of Sec. III. Finally, in Sec. V and in Table I, the main results are summarized.

II. IRREDUCIBLE REPRESENTATIONS OF osp(1,2)

The orthosymplectic Lie superalgebra osp(1,2) can be defined as the set of 3×3 complex matrices of the form

$$\left[\begin{array}{c|cc} 0 & d & e \\ \hline e & a & b \\ -d & c & -a \end{array}\right], \tag{2.1}$$

together with the multiplication rule

$$[x,y] = x\cdot y - (-1)^{\xi\eta} y\cdot x, \quad x\in L_\xi, \quad y \in L_\eta, \tag{2.2}$$

where ξ, $\eta \in \{\bar 0, \bar 1\} = Z_2$, and $L = L_{\bar 0} \oplus L_{\bar 1}$ is the splitting of $L = \mathrm{osp}(1,2)$ into the even and odd subspace. Here, $L_{\bar 0}$ is spanned by matrices (2.1) with $d = e = 0$, and $L_{\bar 1}$ by matrices with $a = b = c = 0$. The even subalgebra equals sl(2). We choose the following basis for osp(1,2):

$$j_0 = \frac{1}{2}\left[\begin{array}{c|cc} 0 & & \\ \hline & 1 & 0 \\ & 0 & -1 \end{array}\right], \quad j_+ = \left[\begin{array}{c|cc} 0 & & \\ \hline & 0 & 1 \\ & 0 & 0 \end{array}\right],$$
$$j_- = \left[\begin{array}{c|cc} 0 & & \\ \hline & 0 & 0 \\ & 1 & 0 \end{array}\right], \tag{2.3}$$
$$q_{1/2} = \left[\begin{array}{c|cc} & 0 & 1 \\ \hline 1 & & \\ 0 & & \end{array}\right], \quad q_{-1/2} = \left[\begin{array}{c|cc} & -1 & 0 \\ \hline 0 & & \\ 1 & & \end{array}\right].$$

Then, the standard (anti-) commutation relations are given by

$$\begin{aligned} &[j_0, j_\pm] = \pm j_\pm, \quad [j_+, j_-] = 2j_0, \\ &[j_0, q_{\pm 1/2}] = \pm \tfrac{1}{2} q_{\pm 1/2}, \quad [j_\pm, q_{\mp 1/2}] = q_{\pm 1/2}, \\ &[q_{\pm 1/2}, q_{\pm 1/2}] = \pm 2 j_\pm, \quad [q_{1/2}, q_{-1/2}] = -2j_0. \end{aligned} \tag{2.4}$$

Irreducible representations (irreps) of osp(1,2) have been studied by several authors.[8,9] In particular, the irreducible spaces V on which osp(1,2) acts are graded: $V = V_{\bar 0} \oplus V_{\bar 1}$. Both $V_{\bar 0}$ and $V_{\bar 1}$ are irreducible sl(2) representations. This is why osp(1,2) irreps are called "dispin." An sl(2) module $V(j)$ is characterized by an integer or a half-integer: (j), $j \in \mathbb{N} \cup \frac{1}{2}\mathbb{N}$. Then j is the maximal j_0 eigenvalue, and $\dim V(j) = 2j+1$. All finite-dimensional irreducible representations of osp(1,2) are determined by[8,9]

[a)] Senior Research Assistant of the National Fund for Scientific Research, Belgium.

$$[a] = (a) \oplus (a - \tfrac{1}{2}) \quad (a = \tfrac{1}{2}, 1, \tfrac{3}{2}, \ldots). \tag{2.5}$$

For $a = 0$ $[a] = (a)$ is the trivial representation. Note that

$$\dim V[a] = 4a + 1. \tag{2.6}$$

The basis vectors of $V(j)$ are denoted by $|j,m\rangle$ $(m = -j, -j+1, \ldots, +j)$. From other studies[8] we copy the explicit form of the actions of the osp(1,2) elements on the basis vectors of the module $V[a]$,

$$\begin{aligned} &j_0 |j,m\rangle = m |j,m\rangle \\ &(j = a \text{ or } j = a - \tfrac{1}{2}), \\ &j_\pm |j,m\rangle = [(j \mp m)(j \pm m + 1)]^{1/2} |j,m \pm 1\rangle; \end{aligned} \tag{2.7}$$

$$\begin{aligned} &q_{1/2} |a,m\rangle = (a-m)^{1/2} |a - \tfrac{1}{2}, m + \tfrac{1}{2}\rangle, \\ &q_{-1/2} |a,m\rangle = -(a+m)^{1/2} |a - \tfrac{1}{2}, m - \tfrac{1}{2}\rangle, \\ &q_{1/2} |a - \tfrac{1}{2}, m - \tfrac{1}{2}\rangle = (a+m)^{1/2} |a,m\rangle, \\ &q_{-1/2} |a - \tfrac{1}{2}, m - \tfrac{1}{2}\rangle = (a - m + 1)^{1/2} |a, m-1\rangle. \end{aligned} \tag{2.8}$$

In fact, (2.7) and (2.8) determine the form of the matrix representation of osp(1,2) on the $(4a+1)$-dimensional space $V[a]$. Taking the following order for the basis of $V[a]$:

$$\begin{aligned} &|a,a\rangle, |a,a-1\rangle, \ldots, |a,-a\rangle, |a - \tfrac{1}{2}, a - \tfrac{1}{2}\rangle, \\ &|a - \tfrac{1}{2}, a - \tfrac{3}{2}\rangle, \ldots, |a - \tfrac{1}{2}, -a + \tfrac{1}{2}\rangle, \end{aligned} \tag{2.9}$$

we find as matrix representatives $\rho(x)$, for $x \in \mathrm{osp}(1,2)$,

$$\rho(j_0) = \mathrm{diag}(a, a-1, \ldots, -a; a - \tfrac{1}{2}, a - \tfrac{3}{2}, \ldots, -a + \tfrac{1}{2}), \tag{2.10}$$

$$\rho(j_+) = \left[\begin{array}{cccccc|cccccc}
0 & [1\cdot(2a)]^{1/2} & & & & & & & & & & \\
 & 0 & [2\cdot(2a-1)]^{1/2} & & & & & & & & & \\
 & & 0 & \ddots & & & & & & & & \\
 & & & \ddots & & & & & & & & \\
 & & & & 0 & [(2a)\cdot 1]^{1/2} & & & & & & \\
 & & & & & 0 & & & & & & \\
\hline
 & & & & & & 0 & [1\cdot(2a-1)]^{1/2} & & & & \\
 & & & & & & & 0 & [2\cdot(2a-2)]^{1/2} & & & \\
 & & & & & & & & 0 & \ddots & & \\
 & & & & & & & & & \ddots & & \\
 & & & & & & & & & & 0 & [(2a-1)\cdot 1]^{1/2} \\
 & & & & & & & & & & & 0
\end{array}\right] \tag{2.11}$$

$$\rho(j_-) = \rho(j_+)', \tag{2.12}$$

$$\rho(q_{1/2}) = \left[\begin{array}{c|c} & \begin{matrix}[2a]^{1/2} & & & & \\ 0 & [2a-1]^{1/2} & & & \\ & 0 & \ddots & & \\ & & \ddots & \sqrt{2} & \\ & & & 0 & \sqrt{1} \\ & & & & 0\end{matrix} \\ \hline \begin{matrix} 0 & \sqrt{1} & & \\ & 0 & \sqrt{2} & \\ & \ddots & \ddots & \\ & & 0 & [2a]^{1/2}\end{matrix} & \end{array}\right], \tag{2.13}$$

$$\rho(q_{-1/2}) = \left[\begin{array}{c|c} & \begin{matrix} 0 & & & \\ \sqrt{1} & 0 & & \\ & \sqrt{2} & 0 & \\ & & \ddots & \ddots & 0 \\ & & & & [2a]^{1/2}\end{matrix} \\ \hline \begin{matrix} -[2a]^{1/2} & 0 & & \\ & -[2a-1]^{1/2} & \ddots & \\ & & \ddots & 0 \\ & & -\sqrt{1} & 0\end{matrix} & \end{array}\right]. \tag{2.14}$$

Hence, it is obvious that ρ is a homomorphism from $\mathrm{osp}(1,2)$ into the Lie superalgebra $\mathrm{spl}(2a+1,2a)$ [sometimes denoted as $\mathrm{sl}(2a+1/2a)$ or as $A(2a,2a-1)$]. This was already observed by Stanley.[7] As a consequence we have $\mathrm{osp}(1,2) \subset \mathrm{spl}(2a+1,2a)$.

Now we want to investigate whether there are in general any other Lie superalgebras appearing in the last inclusion relation. First, we consider the situation with $a \in \mathbb{N}$. Define the following matrices:

$$[\beta_1] = \begin{bmatrix} & & & 1 \\ & & -1 & \\ & ⋰ & & \\ -1 & & & \\ 1 & & & \end{bmatrix}$$
$$\times((2a+1)\times(2a+1)\ \text{matrix}), \tag{2.15}$$

$$[\beta_2] = \begin{bmatrix} & & & 1 \\ & & -1 & \\ & 1 & & \\ \cdot^{\cdot^{\cdot}} & & & \\ -1 & & & \end{bmatrix} \times (2a \times 2a \text{ matrix}), \tag{2.16}$$

and let

$$[\beta] = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}. \tag{2.17}$$

Then $[\beta]$ defines a bilinear form β on the representation space $V[a]$ of the irrep $[a]$ with the basis given by (2.9). With $V_{\bar{0}} = V(a)$ and $V_{\bar{1}} = V(a-\frac{1}{2})$, one can check that (i) β is homogeneous, $\beta(V_{\bar{0}}, V_{\bar{1}}) = 0$; (ii) β is nondegenerate; and (iii) β is supersymmetric,

$$\beta(f,g) = (-1)^{\varphi\gamma}\beta(g,f), \quad \forall f \in V_\varphi, \quad \forall g \in V_\gamma.$$

Making use of the explicit matrices given in (2.10)–(2.14), it is easy to verify that for any homogeneous element x of osp(1,2), one has

$$[\rho(x)]^T[\beta] + (-1)^\xi[\beta][\rho(x)] = 0,$$
$$\forall x \in \operatorname{osp}(1,2)_\xi \quad (\xi = \bar{0}, \bar{1}). \tag{2.18}$$

Herein, $[\rho(x)]^T$ is the supertranspose[10] of a graded matrix, defined by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^T = \begin{bmatrix} a^t & -c^t \\ b^t & d^t \end{bmatrix}. \tag{2.19}$$

Equation (2.18) implies that

$$\beta(\rho(x)\cdot f, g) + (-1)^{\varphi\xi}\beta(f, \rho(x)\cdot g) = 0,$$
$$\forall x \in \operatorname{osp}(1,2)_\xi, \quad \forall f \in V_\varphi, \quad \forall g \in V. \tag{2.20}$$

Consequently, the matrices (2.10)–(2.14) are elements of the orthosymplectic subalgebra $\operatorname{osp}(2a+1,2a)$ contained in $\operatorname{spl}(2a+1,2a)$, and we conclude

$$\operatorname{osp}(1,2) \subset \operatorname{osp}(2a+1,2a) \subset \operatorname{spl}(2a+1,2a), \quad a \in \mathbb{N}. \tag{2.21}$$

When a is a half-integer, a similar analysis leads to

$$\operatorname{osp}(1,2) \subset \operatorname{osp}(2a,2a+1) \subset \operatorname{spl}(2a,2+1), \quad (a-\tfrac{1}{2}) \in \mathbb{N}. \tag{2.22}$$

III. PRINCIPAL FIVE-DIMENSIONAL SUBALGEBRAS

Definition: Let $L = L_{\bar{0}} \oplus L_{\bar{1}}$ be a basic classical Lie superalgebra. Let $L' = \operatorname{span}\{j_0, j_+, j_-, q_{1/2}, q_{-1/2}\}$ be a subalgebra of L with standard relations given by (2.4). Then L' is called a principal five-dimensional subalgebra of L if $\operatorname{span}\{j_0, j_+, j_-\}$ is a principal three-dimensional subalgebra of the Lie algebra $L_{\bar{0}}$.

Principal three-dimensional subalgebras of semisimple Lie algebras are well known. One of the most important properties[3] of the principal sl(2) of a semisimple Lie algebra G is the following: a sl(2) subalgebra of G is principal if and only if the number of irreducible components occurring in the complete reduction of the adjoint representation of sl(2) on G is equal to the rank of G.

With the given definition, it follows immediately from Sec. II that $\operatorname{spl}(n+1,n)$, $\operatorname{osp}(2n+1,2n)$, and $\operatorname{osp}(2n-1,2n)$ ($n \in \mathbb{N}$) contain a principal five-dimensional subalgebra. Now we want to find out which basic classical Lie superalgebras have a principal osp(1,2). There is actually an easy procedure leading to a necessary condition. Indeed, let L be a basic classical Lie superalgebra, and let sl(2) be the principal three-dimensional subalgebra of $L_{\bar{0}}$. Let $(a_1) + (a_2) + \cdots + (a_k)$ be the decomposition of the adjoint representation of sl(2) on $L_{\bar{0}}$, and $(b_1) + (b_2) + \cdots + (b_l)$ be the decomposition on $L_{\bar{1}}$. Then a necessary condition for L in order to contain a principal osp(1,2) subalgebra is that the sequence of numbers $\{a_1,\ldots,a_k,b_1,\ldots,b_l\}$ can be split in couples of the form $\{a_i,b_j\}$ with $b_j = a_i \pm \frac{1}{2}$, or in sets only containing the number 0: $\{0\}$. This follows from the fact that every finite-dimensional representation of osp(1,2) is completely reducible[10] in irreps of the form (2.5) or the trivial representation.

As an example, consider $L = \operatorname{spl}(m,n)$ with $m > n$. Then $L_{\bar{0}} = \operatorname{sl}(m) \oplus \operatorname{sl}(n) \oplus \mathbb{C}$ and $L_{\bar{1}} = \operatorname{sl}_m \otimes \operatorname{sl}_n^* \oplus \operatorname{sl}_m^* \otimes \operatorname{sl}_n$. For the principal sl(2) subalgebra of $L_{\bar{0}}$, $L_{\bar{0}}$ decomposes as

$$(1) + (2) + \cdots + (m-1)$$
$$+ (1) + (2) + \cdots + (n-1) + (0). \tag{3.1}$$

Since $\operatorname{sl}_m^{(*)}$ decomposes as $((m-1)/2)$ and $\operatorname{sl}_n^{(*)}$ as $((n-1)/2)$, we obtain the following decomposition for $L_{\bar{1}}$:

$$\left(\frac{m-n}{2}\right) + \left(\frac{m-n}{2}+1\right) + \cdots + \left(\frac{m+n-2}{2}\right)$$
$$+ \left(\frac{m-n}{2}\right) + \left(\frac{m-n}{2}+1\right) + \cdots + \left(\frac{m+n-2}{2}\right). \tag{3.2}$$

Then it is easy to see that (3.1) and (3.2) can be split in subspaces of the form $(a) + (a-\frac{1}{2})$ only if $m = n+1$. Hence, only $\operatorname{spl}(n+1,n)$ can contain a principal osp(1,2) subalgebra. That it actually does have a principal five-dimensional subalgebra follows from Sec. II. Note that the adjoint representation of $\operatorname{spl}(n+1,n)$ decomposes in osp(1,2) irreps as follows:

$$[\tfrac{1}{2}] + [1] + [\tfrac{3}{2}] + \cdots + [n-\tfrac{1}{2}] + [n], \tag{3.3}$$

and from Sec. II one derives that the standard representation of $\operatorname{spl}(n+1,n)$ decomposes in the irrep $[n/2]$ of the principal osp(1,2).

A similar analysis for Lie superalgebras of type $B(m,n)$ shows that only $\operatorname{osp}(2n+1,2n)$ and $\operatorname{osp}(2n-1,2n)$ satisfy the necessary condition. The Lie superalgebras $C(n)$ never contain a principal osp(1,2) subalgebra if $n > 2$, and when $n = 2$ we have $C(2) = \operatorname{spl}(2,1)$. The only Lie superalgebras of type $D(m,n)$ satisfying the necessary condition are $\operatorname{osp}(2n+2,2n)$ and $\operatorname{osp}(2n,2n)$. Finally, among the exceptional Lie superalgebras there are simply the $D(2,1;\alpha)$ algebras, which can contain a principal osp(1,2). That the previously mentioned Lie superalgebras really do contain a principal five-dimensional subalgebra follows from the fact that the principal osp(1,2) can be realized in the standard representation of the Lie superalgebra. This will be discussed in the following section.

IV. REALIZATION OF THE PRINCIPAL osp(1,2)

For $spl(n+1,n)$, $osp(2n+1,2n)$, and $osp(2n-1,2n)$, the realization of the principal osp(1,2) subalgebra has already been given in Sec. II. From (2.10), one can see that the standard representation of $osp(2n+1,2n)$ decomposes into the irrep $[n]$ when restricted to its principal osp(1,2). When the adjoint representation is restricted to the principal sl(2) subalgebra of $L_{\bar 0}$, one verifies that $L_{\bar 0} = so(2n+1) \oplus sp(2n)$ decomposes into

$$(1)+(3)+\cdots+(2n-1)+(1)+(3)+\cdots+(2n-1),$$

and that $L_{\bar 1} = so_{2n+1} \otimes sp_{2n}$ decomposes into

$$(n)\times(n-\tfrac{1}{2}) = (\tfrac{1}{2})+(\tfrac{3}{2})+\cdots+(2n-\tfrac{1}{2}).$$

Hence, the decomposition of the adjoint representation of $osp(2n+1,2n)$ into irreps of its principal osp(1,2) subalgebra is given by

$$[1]+[3]+\cdots+[2n-1]+[\tfrac{3}{2}] + [\tfrac{7}{2}]+\cdots+[2n-\tfrac{1}{2}]\,. \tag{4.1}$$

Similarly, the standard representation of $osp(2n-1,2n)$ decomposes into $[n-\frac{1}{2}]$, and the adjoint representation decomposes into

$$[1]+[3]+\cdots+[2n-1]+[\tfrac{3}{2}] + [\tfrac{7}{2}]+\cdots+[2n-\tfrac{5}{2}] \tag{4.2}$$

when restricted to its principal osp(1,2) subalgebra.

Now, consider the Lie superalgebra $L = osp(2a+2,2a)$ $(a \in \mathbb{N})$. We shall use the realization given in (2.10)–(2.14) in order to construct a realization of the principal osp(1,2) subalgebra of $osp(2a+2,2a)$. We know that the decomposition of $L_{\bar 0}$ into irreps of its principal sl(2) subalgebra is given by $(0)+(a)+(a-\frac{1}{2})$. Hence, in a similar notation as in Sec. II, we choose the following basis vectors for the representation space V:

$$|0,0\rangle, |a,a\rangle, |a,a-1\rangle,\ldots,|a,-a\rangle, |a-\tfrac{1}{2},a-\tfrac{1}{2}\rangle, \quad |a-\tfrac{1}{2},a-\tfrac{3}{2}\rangle,\ldots,|a-\tfrac{1}{2},-a+\tfrac{1}{2}\rangle. \tag{4.3}$$

Then, acting on this basis the elements of osp(1,2) have the following matrix realization:

$$\rho'(x) = \left[\begin{array}{c|c} 0 & \cdots\ 0 \\ \hline \vdots & \rho(x) \\ 0 & \end{array}\right], \tag{4.4}$$

where $\rho(x)$ is given by (2.10)–(2.14). Hence, (4.4) is a realization of osp(1,2) in block matrices of type $[(2a+2)+(2a)]\times[(2a+2)+(2a)]$. Moreover

$$[\rho'(x)]^T[\beta'] + (-1)^\xi[\beta'][\rho'(x)] = 0, \quad \forall x \in osp(1,2)_\xi, \tag{4.5}$$

where

$$[\beta'] = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \beta_1 & 0 \\ \vdots & & \\ 0 & 0 & \beta_2 \end{array}\right]. \tag{4.6}$$

The bilinear form β' defined by (4.6) on $V = V_{\bar 0} \oplus V_{\bar 1}$, with $V_{\bar 0} = V(0)+V(a)$ and $V_{\bar 1} = V(a-\frac{1}{2})$, is homogeneous, nondegenerate, and supersymmetric. It follows that (4.4) is a realization of the principal osp(1,2) into $osp(2a+2,2a)$.

For $osp(2n,2n)$, one makes use of the realization of the principal osp(1,2) in the superalgebra $osp(2a,2a+1)$, $(a-\frac{1}{2}) \in \mathbb{N}$, and one performs the same construction as in (4.4) in order to obtain the principal osp(1,2) contained in $osp(2a+1,2a+1)$.

Finally, for the exceptional Lie superalgebras $D(2,1;\alpha)$ we prefer to use the notation of Scheunert[10] $\Gamma(\sigma_1,\sigma_2,\sigma_3)$. The connection between $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ and $D(2,1;\alpha)$ has been given in Ref. 11. The even part of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ is $sl(2)\oplus sl(2)\oplus sl(2)$, spanned by $s_{0,\pm}$, $t_{0,\pm}$ and $u_{0,\pm}$. The odd part is equal to $sl_2 \otimes sl_2 \otimes sl_2$ and its basis vectors are denoted by $R_{\alpha,\beta,\gamma}$ $(\alpha,\beta,\gamma = \pm\frac{1}{2})$. The product relations in $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ have been given explicitly in a previous paper.[11] Since $(\sigma_1,\sigma_2,\sigma_3)$ are determined up to an arbitrary factor, we can choose $4\sigma_1\sigma_2\sigma_3 = 1$. Then, the explicit expressions for

TABLE I. The Lie superalgebras L having a principal osp(1,2) subalgebra. Besides the dimension and rank of L, we also list the decomposition of the standard representation ρ_S and of the adjoint representation ρ_A when decomposed into irreps of the principal five-dimensional subalgebra.

L	dim L	rank L	ρ_S	ρ_A
$A(n,n-1) = spl(n+1,n)$	$4n^2+4n$	$2n$	$[n/2]$	$[1]+[2]+\cdots+[n] + [\frac{1}{2}]+[\frac{3}{2}]+\cdots+[n-\frac{1}{2}]$
$B(n,n) = osp(2n+1,2n)$	$4n(2n+1)$	$2n$	$[n]$	$[1]+[3]+\cdots+[2n-1] + [\frac{3}{2}]+[\frac{7}{2}]+\cdots+[2n-\frac{1}{2}]$
$B(n-1,n) = osp(2n-1,2n)$	$8n^2-4n+1$	$2n-1$	$[n-\frac{1}{2}]$	$[1]+[3]+\cdots+[2n-1] + [\frac{3}{2}]+[\frac{7}{2}]+\cdots+[2n-\frac{5}{2}]$
$D(n+1,n) = osp(2n+2,2n)$	$8n^2+8n+1$	$2n+1$	$[0]+[n]$	$[1]+[3]+\cdots+[2n-1] + [\frac{3}{2}]+[\frac{7}{2}]+\cdots+[2n-\frac{1}{2}]+[n]$
$D(n,n) = osp(2n,2n)$	$8n^2$	$2n$	$[0]+[n-\frac{1}{2}]$	$[1]+[3]+\cdots+[2n-1] + [\frac{3}{2}]+[\frac{7}{2}]+\cdots+[2n-\frac{5}{2}]+[n-\frac{1}{2}]$
$D(2,1;\alpha)$	17	3	$\cdots$	$[1]+[1]+[\frac{3}{2}]$

the basis elements of the principal osp(1,2) subalgebra of $\Gamma(\sigma_1,\sigma_2,\sigma_3)$ are given by

$$j_\mu = s_\mu + t_\mu + u_\mu \quad (\mu = 0, \pm),$$
$$q_{\pm 1/2} = \pm \sqrt{2}[\sigma_1 R_{\mp 1/2, \pm 1/2, \pm 1/2} + \sigma_2 R_{\pm 1/2, \mp 1/2, \pm 1/2} + \sigma_3 R_{\pm 1/2, \pm 1/2, \mp 1/2}] \,. \tag{4.7}$$

V. CONCLUSION

It has been known for a long time that all semisimple Lie algebras contain a principal sl(2) subalgebra. For basic classical Lie superalgebras, the analog of a principal sl(2) is a principal five-dimensional subalgebra osp(1,2). Our analysis shows that not all the basic Lie superalgebras contain a principal osp(1,2) subalgebra, but only the ones given in Table I. For all these cases, we have obtained an explicit realization of the principal osp(1,2) in the standard representation of the Lie superalgebra. Table I lists the reduction of the standard representations and of the adjoint representations for the Lie superalgebras L decomposed into irreps of their principal five-dimensional subalgebra. From this table, one can notice that the number of osp(1,2) irreps in which the adjoint representation of a Lie superalgebra L decomposes when restricted to its principal osp(1,2) subalgebra is equal to the rank of L. This is the analog of the defining property[3] of a principal sl(2) subalgebra of a Lie algebra.

In a paper by Stanley[7] some combinatorial properties concerning unimodality were derived by studying the principal osp(1,2) subalgebra of spl$(n+1,n)$ or of pl$(n+1,n)$ [sometimes denoted as gl$(n+1/n)$]. It is clear that the analysis of representations of the Lie superalgebras in Table I may give rise to similar interesting combinatorial properties, but this study falls beyond the scope of the present paper.

[1]M. Hamermesh, *Group Theory and its Applications to Physical Problems* (Addison–Wesley, London, 1962).
[2]E. B. Dynkin, Am. Math. Soc. Transl., Ser. 2 **6**, 111 (1957).
[3]B. Kostant, Am. J. Math. **81**, 973 (1959).
[4]J. W. B. Hughes, in *Number Theory and Algebra*, edited by H. Zassenhaus (Academic, New York, 1977), p. 135.
[5]R. P. Stanley, in "Combinatorics, representation theory and statistical methods in groups," *Lectures Notes in Pure and Applied Mathematics*, Vol. 57, edited by T. V. Narayana, R. M. Mathsen, and J. G. Williams, (Dekker, New York, 1980), p. 127.
[6]V. G. Kac, Adv. Math. **26**, 8 (1977).
[7]R. P. Stanley, Stud. Appl. Math. **72**, 263 (1985).
[8]J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981).
[9]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[10]M. Scheunert, *Springer Lecture Notes in Mathematics*, Vol. 716 (Springer, Berlin, 1979).
[11]J. Van der Jeugt, J. Math. Phys. **26**, 913 (1985).

Are all the equations of the Kadomtsev–Petviashvili hierarchy integrable?

B. Dorizzi and B. Grammaticos
Département de Mathématiques, Centre National d'Etudes de Télécommunications, F-92131, Issy-les-Moulineaux, France
A. Ramani
Centre de Physique Théorique, Ecole Polytechnique, F-91128, Palaiseau, France
P. Winternitz
Centre de Recherches Mathématiques, Université de Montréal, C. P. 6128, Succursale A, Montréal, Québec, Canada H3C 3J7

(Received 14 April 1986; accepted for publication 28 July 1986)

The Kadomtsev–Petviashvili (KP) hierarchy is an infinite set of nonlinear partial differential equations in which the number of independent variables increases indefinitely as one proceeds down the hierarchy. Since these equations were obtained as part of a group theoretical approach to soliton equations it would appear that the KP hierarchy provides integrable scalar equations with an arbitrary number of independent variables. It is shown, by investigating a specific equation in $3+1$ dimensions, that the higher equations in the KP hierarchy are only integrable in a conditional sense. The equation under study, taken in isolation, does not pass certain well-known and reliable integrability tests. Thus, applying Painlevé analysis, we find that solutions exist, allowing movable critical points. Furthermore, solitary wave solutions are shown to exist that do not behave like solitons in multiple collisions. On the other hand, if the dependence of a solution on the first $2+1$ variables is restricted by the fact that it should also satisfy the KP equation itself, then the integrability conditions in the other dimensions are satisfied. "Conditional integrability" thus means that linear techniques will provide only those solutions of equations in the hierarchy that simultaneously satisfy lower equations in the same hierarchy.

I. INTRODUCTION

The purpose of this article is to discuss the integrability of an infinite set of nonlinear partial differential equations (PDE's) proposed recently by Jimbo and Miwa,[1] and called the Kadomtsev–Petviashvilli (KP) hierarchy. We shall use the word "integrable" to mean that a given nonlinear PDE can be integrated by essentially linear techniques, such as the inverse scattering method,[2] the "dressing method,"[3] or various group theoretical approaches.[1,4] Most of the equations integrable in the above sense involve only $1+1$ variables (the Korteweg–de Vries equation, the nonlinear Schrödinger equation, the sine–Gordon equation, or the equations of the nonlinear σ model being prime examples[2,3]). Some well known examples of integrable equations also exist in $2+1$ dimensions. These include, e.g., the Kadomtsev–Petviashvili[5] equation and the Davey–Stewartson equation,[6] which are of considerable physical interest. The only integrable system of PDE's in n dimensions that we are aware of is written for $n\times n$ matrix functions and has very exceptional geometrical properties (it generalizes the sine–Gordon equation).[7,8]

Since most of the nonlinear equations of physics are written in $3+1$ dimensions or more, the question of whether such equations can be integrable by linear techniques is of cardinal importance. In other words, is integrability a low-dimensional accident or does it occur in arbitrary dimensions?

We shall not delve here into the philosophical implications of the existence or nonexistence of integrable equations in higher dimensions. Rather, we intend to show that one current belief in the existence of integrable equations involving arbitrary dimensions in the hierarchies of equations presented by Jimbo and Miwa,[1] in particular the KP hierarchy, is unfounded, at least in the most direct sense.

In a recent publication[1] Jimbo and Miwa have reviewed and further developed an approach to soliton-type equations based on the representation theory of infinite-dimensional Lie algebras and Lie groups. The approach is primarily due to the Kyoto school and goes back to the original work of Sato and Sato.[9] The KP hierarchy of equations is the most basic one in this approach and this hierarchy, as well as its solutions, are obtained from the representation of the algebra gl(∞). Other equations and their hierarchies are associated with infinite-dimensional subalgebras of gl(∞), such as the orthogonal and symplectic algebras B_∞, C_∞, and D_∞, or Kac–Moody Lie algebras. The Hirota bilinear formalism[10] as well as the formalism of Lax pairs[2] are incorporated in this approach in a natural manner.

The KP hierarchy is presented[1] in the Hirota formalism in terms of the D operators, defined by their action on bilinear expressions:

$$\begin{aligned}[D^a_{x_1} D^b_{x_2}\cdots]\,\tau(x_1,x_2,\ldots)\cdot\tau(x_1,x_2,\ldots)\\ =(\partial_{x_1}-\partial_{x_1'})^a(\partial_{x_2}-\partial_{x_2'})^b\\ \cdots\tau(x_1,x_2,\ldots)\tau(x_1',x_2',\ldots)\big|_{x_1'=x_1,x_2'=x_2\cdots}\,. \end{aligned}\tag{1.1}$$

The first four equations of the KP hierarchy are written in this formalism as

$$[D_1^4+3D_2^2-4D_1D_3]\tau\cdot\tau=0\,, \tag{1.2}$$

$$[(D_1^3+2D_3)D_2-3D_1D_4]\tau\cdot\tau=0\,, \tag{1.3}$$

$$[D_1^6-20D_1^3D_3-80D_3^2+144D_1D_5-45D_1^2D_2^2]\tau\cdot\tau=0\,, \tag{1.4}$$

$$[D_1^6+4D_1^3D_3-32D_3^2-9D_1^2D_2^2+36D_2D_4]\tau\cdot\tau=0\,. \tag{1.5}$$

Equation (1.2) involves only three variables, $x_1=x$, $x_2=y$, $x_3=t$, and is actually the KP equation itself. Thus putting

$$w=2\frac{\partial}{\partial x}\log\tau,\quad u=2\frac{\partial^2}{\partial x^2}\log\tau\,,$$

we reduce (1.2) to either the potential KP equation

$$w_{xxxx}+6w_xw_{xx}+3w_{yy}-4w_{xt}=0\,, \tag{1.6}$$

or to the more standard form of the KP equation

$$[4u_t-6u\,u_x-u_{xxx}]_x-3u_{yy}=0\,. \tag{1.7}$$

Equations (1.3)–(1.5) involve four independent variables each and higher members of the hierarchy involve higher dimensions, in principle arbitrarily high ones. This would seem to indicate that the KP hierarchy, as well as other hierarchies,[1] provide integrable equations involving arbitrary numbers of independent variables. Moreover, this impression is given credence by the fact that Jimbo and Miwa[1] give a τ-function solution to the entire KP hierarchy and it has the form of an N-soliton solution.

Our aim is to point out that while the entire KP hierarchy, taken together, is integrable,[1] individual equations in the series, taken out of context, fail the usual integrability tests. More specifically we shall investigate the second equation in the hierarchy, namely (1.3), which we rewrite in more standard notation as

$$w_{xxxy}+3w_{xy}w_x+3w_yw_{xx}+2w_{yt}-3w_{xz}=0 \tag{1.8}$$

[we have put $w=2(\partial/\partial x)\log\tau$ and $x_1=x$, $x_2=y$, $x_3=t$, and $x_4=z$].

In Sec. II we perform a Painlevé analysis[11–13] of this equation. As a PDE in four variables it does not pass the test. Hence the equation does not have the Painlevé property and has solutions that are not single-valued functions in the neighborhood of their singularity surfaces. Such behavior is generally considered to be incompatible with integrability.[2,11,12] On the other hand, we show that if a solution w of (1.8) also satisfies the KP equation (1.6) (as a function of x, y, and t) then it will have the Painlevé property (in all four variables x, y, z, and t).

This leads us to the concept of conditional integrability: Any given equation in the hierarchy must be considered together with the lower equations in the hierarchy: the common solutions of all these equations have the usual integrability properties.

In Sec. III we perform a second test. We consider solitary wave solutions of Eq. (1.8) and ask whether they behave like solitons with respect to mutual interactions. Indeed, solitary waves exist for numerous nonlinear PDE's and two-solitary wave solutions exist for any equation that can be cast into Hirota's bilinear form. The existence of N-soliton solutions for $N\geqslant 3$ is a quite nontrivial phenomenon that can be considered to be an indication of integrability.[14] We show that solitary wave solutions of (1.8) can be composed into three-soliton solutions only in the case when the obtained solution also solves the KP equation (1.6). Thus, the integrability condition is again satisfied in a conditional sense only.

In Sec. IV we calculate the group of Lie symmetries of Eq. (1.8). While this group is infinite dimensional, its Lie algebra does not have the structure of a loop algebra, typical for integrable equations in $2+1$ dimensions. Section V is devoted to conclusions.

II. THE PAINLEVÉ ANALYSIS

According to the "Painlevé conjecture" due to Ablowitz, Ramani, and Segur[11] (ARS), whenever a system of partial differential equations is integrable by the inverse scattering method (or related linear techniques), all ordinary differential equations, obtained from this PDE by symmetry reduction, will have the Painlevé property. This means that the only movable singularities (i.e., singularities depending on the initial conditions) of any solution of such a nonlinear ODE are poles (in the complex plane of the independent variable). This in particular excludes movable singularities of the branch point type (in addition to essential singularities). A subsequent extension of the ARS approach by Weiss *et al.*[12] made it possible to bypass the symmetry reductions and deal directly with the PDE itself. A singularity manifold $\phi(z_1,...,z_n)$ is introduced[12] in the complex space of the independent variables and a necessary condition for the PDE to have the Painlevé property is that an expression of the form

$$w=\frac{1}{\phi^\alpha}\sum_{k=0}^{\infty}a_k\phi^k \tag{2.1}$$

exists for a general class of solutions of the PDE (fitting essentially arbitrary Cauchy data). Here α must be an integer, and the coefficients a_k are analytic functions of the independent variables in the neighborhood of the singularity surface $\phi=0$. For an Nth-order PDE the solution (2.1) must involve N arbitrary functions, namely ϕ and $N-1$ of the coefficients a_k (the corresponding values of k are "resonances").[11,12] An important improvement of this Painlevé test is due to Kruskal[13] and consists of the requirement that the function $\phi(z_1,...,z_n)$ should be linear in one of the variables, say z_n, and that the coefficients a_k should only depend on the remaining variables, i.e.,

$$\phi=z_n+\psi(z_1,...,z_{n-1}),\quad a_k=a_k(z_1,...,z_{n-1})\,, \tag{2.2}$$

in (2.1). This Painlevé test is completely algorithmic, has been applied to a host of PDE's, and has proven to be a most successful integrability criterion.

We now apply the Painlevé test to Eq. (1.8) of the Jimbo–Miwa hierarchy. We write the solution $w(x,y,z,t)$ in the form (2.1) with

$$\phi=x+\psi(y,t,z),\quad a_k=a_k(y,t,z)\,. \tag{2.3}$$

Substituting (2.3) into (1.8) and following the usual procedure,[11,12] we find

$$\alpha = 1, \quad a_0 = 2. \tag{2.4}$$

The "resonances," i.e., the values at which the recursion relations obtained from (1.8) do not determine the functions a_k, are found to be $k = -1$, 1, 4, and 6. The final step is to compute the coefficients a_2, a_3, and a_5 from the recursion relation and to verify that the resonance conditions, i.e., the compatibility conditions for the existence of the free functions a_1, a_4, and a_6, are verified. This is a tedious computation that is best performed by algebraic computations on a computer (we used REDUCE as a language). The result is

$$\begin{aligned} &a_1(y,t,z) = \text{free}, \\ &a_2 = (-2\phi_t\phi_y + 3\phi_z - 6a_{1,y})/12\phi_y, \\ &a_3 = (2a_{1,yy}\phi_y - 2a_{1,y}\phi_{yy} \\ &\qquad + 2\phi_{ty}\phi_{yy} - \phi_{yz}\phi_y + \phi_{yy}\phi_z)/16\phi_y^3, \\ &a_4 = \text{free}. \end{aligned} \tag{2.5}$$

The expression for a_5 is too long to reproduce here (it is available from the authors on request). At the resonance $k = 6$ we obtain a condition which is not satisfied identically. Indeed the resonance condition here is

$$\begin{aligned} R = &-2a_{1,yz}\phi_{yy} + 2a_{1,yy}\phi_{yz} + 2\phi_{ty}\phi_{yz}\phi_y \\ &- 2\phi_{tz}\phi_{yy}\phi_y - \phi_{yz}^2 + \phi_{yy}\phi_{zz} = 0. \end{aligned} \tag{2.6}$$

We see that the condition $R = 0$ is an equation relating the functions $a_1(y,t,z)$ and $\phi(y,t,z)$, rather than an identity. The conclusion is that Eq. (1.8), taken on its own, does not satisfy the Painlevé criterion and is therefore presumably not integrable.

Let us now consider Eq. (1.8) together with the preceding equations in the hierarchy, which in this case is simply the KP equation (1.6) itself. The KP equation does not involve the variable z, so we fix $z = z_0$ and write a singular expansion for the solutions of the KP equation (which is well known to satisfy the Painlevé requirement):

$$\begin{aligned} &w(x,y,z_0,t) = \frac{1}{\tilde{\phi}^\alpha} \sum_{k=0}^{\infty} b_k \tilde{\phi}^k, \\ &\qquad \tilde{\phi} = x + \tilde{\psi}(y,t,z_0), \quad b_k = b_k(y,t,z_0). \end{aligned} \tag{2.7}$$

As for Eq. (1.8), we find

$$\alpha = 1, \quad b_0 = 2,$$

and resonances at $k = -1$, 1, 4, and 6. We have

$$b_1 = \text{free}, \quad b_2 = \tfrac{1}{3}\tilde{\phi}_t - \tfrac{1}{4}\tilde{\phi}_y^2. \tag{2.8}$$

In the spirit of "conditional integrability" of a nonlinear PDE, introduced in the Introduction, we now require that $w(x, y, z, t)$ be a solution of the Jimbo–Miwa (JM) equation (1.8) and simultaneously, for $z = z_0$ fixed, a solution of the KP equation. This means that we must have

$$\tilde{\psi}(y,t,z_0) = \psi(y,t,z_0) \text{ and } b_k(y,t,z_0) = a_k(y,t,z_0) \tag{2.9}$$

for all values of k. In particular $b_2 = a_2$ implies

$$\phi_z = 2a_{1,y} + 2\phi_t\phi_y - \phi_y^3. \tag{2.10}$$

Using condition (2.10) we can show that $a_3 = b_3$; we then choose $a_4 = b_4$ (since both are free) and obtain $a_5 = b_5$. At order $k = 6$ the compatibility condition for the KP equation is satisfied automatically (the KP is integrable). Moreover, the compatibility condition for the JM equation (1.8) at $k = 6$ is now also satisfied, i.e., $R = 0$ in (2.6) is a consequence of (2.10).

We see that "conditional integrability" in the case at hand means that Eq. (1.8) satisfies the necessary conditions for the Painlevé property only for a subclass of solutions. These are solutions for which the evolution of the singularity manifold ϕ in the z direction is determined by Eq. (2.10) for initial data given at some $z = z_0$ by an arbitrary function of y and t. For integrability in the usual sense ϕ should be an arbitrary function of all three variables y, t, and z.

III. THREE-SOLITON SOLUTIONS

We now turn to the integrability criterion related to "soliton" solutions. We are taking the attitude that solitary waves for a partial differential equation only deserve to be called solitions if N-soliton solutions exist for this equation for all values of N. Following Hirota[10] we look for the N-soliton solutions in the following way. We define the quantities

$$\eta_i = k_i x + l_i y + m_i t + n_i z \quad (i = 1,2,...), \tag{3.1}$$

where the sets of numbers (k_i, l_i, m_i, n_i) satisfy the following dispersion relations:

$$P_{\mathrm{KP}}(k_i, l_i, m_i) = k_i^4 + 3l_i^3 - 4l_i m_i = 0 \tag{3.2}$$

for the KP equation (1.2) and

$$P_{\mathrm{JM}}(k_i, l_i, m_i, n_i) \equiv (k_i^3 + 2m_i)l_i - 3k_i n_i = 0 \tag{3.3}$$

for the "Jimbo–Miwa equation" (1.3). Using the variables η_i we define the following solutions.

A. "One-soliton" solutions

The form is

$$\tau = 1 + e^{\eta_1}. \tag{3.4}$$

This provides a solution of the KP equations (1.2) if (3.2) is satisfied (z is then a parameter that can be absorbed into the soliton phase). The function τ is a solution of Eq. (1.3) if (3.3) is satisfied and a common solution of both equations if (3.2) and (3.3) hold simultaneously.

B. "Two-soliton" solutions

The form is

$$\tau = 1 + e^{\eta_1} + e^{\eta_2} + A_{12}e^{\eta_1 + \eta_2}, \tag{3.5}$$

with

$$A_{12} = \frac{P(k_1 - k_2, l_1 - l_2, m_1 - m_2, n_1 - n_2)}{P(k_1 + k_2, l_1 + l_2, m_1 + m_2, n_1 + n_2)}. \tag{3.6}$$

If P in (3.6) is taken to be P_{KP} and (3.2) is satisfied for both sets (k_i, l_i, m_i) $(i = 1,2)$, then τ in (3.5) is a solution of the KP equation (1.2). Similarly, if P is P_{JM} and (3.3) is satisfied, then τ in (3.5) is a solution of the JM equation (1.3) [whether (3.2) is satisfied or not]. If both (3.2) and (3.3) are satisfied then the two expressions for A_{12} coincide and τ is a common solution of the KP equation (1.2) and the JM equation (1.3). The common value of A_{12} in this case is

$$A_{12}=\frac{k_1^2 k_2^2 (k_1-k_2)^2-(k_1l_2-k_2l_1)^2}{k_1^2 k_2^2 (k_1+k_2)^2-(k_1l_2-k_2l_1)^2}. \quad (3.7)$$

C. "Three-soliton" solutions

If a three-soliton solution exists in the Hirota formalism then its form must be

$$\tau=1+e^{\eta_1}+e^{\eta_2}+e^{\eta_3}+A_{12}e^{\eta_1+\eta_2}+A_{13}e^{\eta_1+\eta_3} + A_{23}e^{\eta_2+\eta_3}+A_{12}A_{13}A_{23}e^{\eta_1+\eta_2+\eta_3}. \quad (3.8)$$

Moreover, (3.8) is a solution only if the quantity

$$Q=P(k_1+k_2+k_3,...)A_{12}A_{13}A_{23}+P(k_1+k_2-k_3,...)A_{12} + P(k_1-k_2+k_3,...)A_{23}+P(-k_1+k_2+k_3,...)A_{23} \quad (3.9)$$

vanishes identically.

If P is $P_{\rm KP}$ in (3.9) and (3.2) is satisfied for all three sets (k_i,l_i,m_i) $(i=1,2,3)$, then the quantity Q does indeed vanish. This is of course well known: the KP equation does have three-soliton solutions (and N-soliton solutions for any N).

If, on the other hand, P is $P_{\rm JM}$ in (3.6) and (3.9) and we request only that (3.3) be satisfied for the three sets (k_i,m_i,n_i,l_i) $(i=1,2,3)$, then Q is *not* identically zero. We do not reproduce its value here (obtained by a REDUCE calculation) since it is longer than this entire article. Thus, in general (3.8) is *not* a solution of the JM equations (1.3).

Let us now require that (3.2) be satisfied, in addition to (3.3), and let P be $P_{\rm JM}$ in (3.9), as before. For A_{12} we obtain the expression (3.7) (for both $P_{\rm KP}$ and $P_{\rm JM}$). In this case we find that the quantity Q in (3.9) does vanish and hence τ of (3.8) is a solution of the JM equation (1.3). This was to be expected: we simply reobtain the N-soliton solutions (for $N=3$) obtained more generally by Jimbo and Miwa[1] for the entire KP hierarchy.

The crucial point that we are making is that the variables η_i for these N-soliton solutions must satisfy both (3.2) and (3.3). Solitary waves of the JM equation (1.3) [or equivalently (1.8)] do not, in general, interact as solitons in collisions of three or more at a time. Thus, the criterion of the existence of multisoliton solutions again leads to the conditional integrability of Eq. (1.3). Integrability in the usual (unconditional) sense would imply that all solitary wave solutions of the equation should interact like solitons in interactions of arbitrary multiplicity.

IV. THE SYMMETRY GROUP

Standard algorithms exist for calculating the group of Lie symmetries of a differential equation, or system of equations. Here we consider the symmetry group in the most direct sense of the word, i.e., the group of point transformations

$$\mathbf{x}'=\Lambda_g(\mathbf{x},w), \quad w'=\Omega_g(\mathbf{x},w), \quad (4.1)$$

such that $w'(\mathbf{x}')$ is a solution, whenever $w(\mathbf{x})$ is one.

The symmetry groups of the Kadomtsev–Petviashvili equation,[15] the Davey–Stewartson equation,[16] and other integrable equations in $2+1$ dimensions have recently been calculated. They are all infinite-dimensional and have a specific loop group structure. More specifically, the corresponding infinite-dimensional Lie algebras involve arbitrary functions of time t. When these are expanded into formal Laurent series we obtain a structure that can be identified with a subalgebra of an affine-type Kac–Moody algebra (with no central extension).[15]

Let us now turn to the Jimbo–Miwa equation (1.8). As usual[15–17] we look for a general element of the Lie algebra of the symmetry group in the form

$$V=\xi_1\,\partial_x+\xi_2\,\partial_y+\xi_3\,\partial_z+\xi_4\,\partial_t+\psi\,\partial_w, \quad (4.2)$$

where ξ_i and ψ are functions of x, y, z, t, and w. The considered equation (1.8) is a fourth-order one; hence we need to construct the fourth prolongation of the vector field (4.2) and request that it annihilate the equation on the solution space:

$$pr^{(4)}V\cdot\Delta(\mathbf{x},w)|_{\Delta(\mathbf{x},w)=0}=0, \quad (4.3)$$

where $\Delta(\mathbf{x},w)$ is the right-hand side of Eq. (1.8). Applying a previously written MACSYMA package[16] to the case of the Jimbo–Miwa equation (1.8), we obtain a system of 25 simple first-order linear partial differential equations for the functions ξ_i and ψ. Solving these determining equations we find that the symmetry algebra is infinite dimensional. It depends on five arbitrary constants, three arbitrary functions of one variable $f_1(z), f_2(z)$, and $g(t)$, and one function of two variables $H(z,t)$.

A basis for this Lie algebra is given by the operators

$$\begin{aligned}
&P_1=\partial_x, \quad P_2=\partial_y, \quad P_3=\partial_z, \quad P_4=\partial_t,\\
&D_1=x\,\partial_x+2z\,\partial_z+3t\,\partial_t,\\
&D_2=x\,\partial_x-2y\,\partial_y+3t\,\partial_t,\\
&X(f_1)=f_1(z)\partial_y+\tfrac{3}{4}f_1'(z)t\,\partial_x\\
&\qquad -\left[\tfrac{1}{2}f_1'(z)x+\tfrac{3}{4}f_1''(z)ty\right]\partial_w, \qquad (4.4)\\
&Y(f_2)=f_2(z)\partial_x-f_2'(z)y\,\partial_w, \quad f_2'\neq 0,\\
&Z(g)=g(t)\partial_x+\tfrac{2}{3}g'(t)x\,\partial_w, \quad g'\neq 0,\\
&W(H)=H(z,t)\partial_w
\end{aligned}$$

(the prime indicates differentiation with respect to the argument).

The algebra (4.4) can be integrated in a simple manner to provide the invariance group of Eq. (1.8). This can in turn be used to generate solutions, to perform symmetry reduction, etc. We see that all expected symmetries are present: the four translations P_i $(i=1,...,4)$, two independent dilations D_1, D_2, Galilei transformations in the x direction for $g(t)=t$, "quasirotations" in the z-y plane for $f_1(z)=z$, or in the z-x plane for $f_2(z)=z$, etc. The generator $W(H)$ simply expresses the fact that an arbitrary function of z and t can be added to any solution.

In any case, our purpose is not to solve Eq. (1.8), nor to analyze its symmetry group in detail. We wish simply to point out that the loop group structure that occurs for the KP equation and other integrable equations in more than $1+1$ dimensions is absent here. The reason is that the loop group structure requires the presence of terms of the type $h(t)\partial_t+...$, or $k(z)\partial_z+...$ that are absent in the case under consideration.

V. CONCLUSIONS AND COMMENTS

In summary let us reemphasize that the Painlevé criterion, the "multisoliton" criterion, and symmetry group considerations all agree to suggest strongly that Eq. (1.8) is "conditionally" integrable, rather than integrable on its own. The integrability properties only manifest themselves in solutions $w(x, y, z, t)$ that for $z = z_0$ fixed are also solutions of the KP equation (1.6).

To show that this is not an isolated occurrence, we recall that a similar situation, conditional integrability, has been encountered earlier,[18] but for two equations involving the same number of variables. In fact these equations can be obtained from the same KP hierarchy by reductions. More specifically these are what Jimbo and Miwa call two-reductions: one considers a situation in which there is no dependence on any of the even indexed variables. Starting from the KP equation (1.2) we obtain the Korteweg–de Vries (KdV) equation in Hirota form:

$$[D_1^4 - 4D_1D_3]\tau\cdot\tau = 0\,, \tag{5.1}$$

while Eq. (1.5) reduces to

$$[D_1^6 - 4D_1^3 D_3 - 32D_3^2]\tau\cdot\tau = 0\,. \tag{5.2}$$

Conditional integrability in this case simply means that all the solutions of the integrable KdV equation (5.1) also provide solutions of (5.2). Thus, let us assume that some $\tau\equiv\tau_0$ is a solution of the KdV equation (5.1). Substituting into (5.2) we obtain

$$[D_1^6 + 4D_1^3 D_3 - 32D_3^2]\tau_0\cdot\tau_0 = g(x_3)\tau_0^2\,, \tag{5.3}$$

where $g(x_3)$ is some function of x_3 only. Moreover, $G(x_3)\tau_0$ will be a solution of (5.1) for any function $G(x_3)$. Substituting $G(x_3)\tau_0$ into (5.2) we obtain

$$-64(\ln G)'' + g = 0\,. \tag{5.4}$$

Hence, choosing $G(x_3)$ so as to satisfy (5.4) we obtain a function $\tau = G(x_3)\tau_0$ satisfying both (5.1) and (5.2).

As was argued earlier,[18] Eq. (5.2) is presumably not integrable as it stands since (i) it does not have the Painlevé property; and (ii) in general its solitary waves cannot be combined to provide three-soliton solutions.

Indeed, the only equations of the form

$$(\lambda D_1^6 + \mu D_1^3 D_3 + \nu D_3^2)(\tau\cdot\tau) = 0 \tag{5.5}$$

that satisfy both of the above criteria have either $\lambda = 0$, or $\lambda\nu = -\mu^2/5$, while in (5.2) we have $\lambda\nu = -2\mu^2$.

The situation with three-soliton solutions of (5.2) was analyzed in detail in the Appendix of Ref. 18. In a nutshell, this equation, being second order in time, has two types of solitary wave solutions. The first type also satisfies the KdV. These solitary waves behave like solitons and provide N-soliton solutions for any N. The second type of solitary wave has a different dispersion relation and these waves do not behave like solitons in collisions of three or more waves.

As a final comment we mention that for integrable equations in the usual sense, linear techniques provide large classes of solutions, e.g., all solutions decaying sufficiently rapidly at infinity, or all solutions satisfying certain quasiperiodicity solutions. For conditionally integrable equations, considered on their own, only a small subclass of solutions is obtained. Indeed the condition (2.10) for the singularity surface, or (3.2) together with (3.3) for the solitary wave parameters for the JM equation (1.8), are very serious restrictions and are certainly not satisfied by a generic solution. We conclude that the problem of the existence of genuinely integrable scalar equations in more than $2 + 1$ dimensions remains open.

ACKNOWLEDGMENTS

The research of one of the authors (P.W.) was partially supported by the Natural Sciences and Engineering Research Council of Canada and the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR) du Gouvernement du Québec.

[1]M. Jimbo and T. Miwa, Publ. Res. Inst. Math. Sci., Kyoto Univ. **19**, 943 (1983).

[2]M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).

[3]S. P. Novikov, S. V. Manakov, L. P. Pitaevskiĭ, and V. E. Zakharov, *Theory of Solitons—The Inverse Method* (Plenum, New York, 1984).

[4]V. Kac, *Infinite Dimensional Lie Algebras* (Birkhauser, Boston, 1983).

[5]B. B. Kadomtsev and V. I. Petviashvili, Sov. Phys. Dokl. **15**, 539 (1970).

[6]A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).

[7]K. Tenenblat and C. L. Terng, Ann. Math. **111**, 477 (1980).

[8]K. Tenenblat, Bol. Soc. Brasil. Mat. **16** 67 (1985).

[9]M. Sato and Y. Sato (Mori), Res. Inst. Math. Sci. Kôkyûroku Kyoto Univ. **388**, 183 (1980); **414**, 181 (1981).

[10]R. Hirota, in *Solitons*, edited by R. K. Bullough and P. J. Chaudrey (Springer, Berlin, 1980).

[11]M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715, 1006 (1980).

[12]J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).

[13]M. Kruskal (private communication); M. Jimbo, M. D. Kruskal, and T. Miwa, Phys. Lett. A **92**, 59 (1982).

[14]M. Ito, J. Phys. Soc. Jpn. **49**, 771 (1980).

[15]D. David, N. Kamran, D. Levi, and P. Winternitz, Phys. Rev. Lett. **55**, 2111 (1985); J. Math. Phys. **27**, 1225 (1986).

[16]B. Champagne and P. Winternitz, Preprint CRM-1278, Montreal, 1985.

[17]P. Olver, J. Differ. Geom. **14**, 497 (1979).

[18]A. Ramani, in *Fourth International Conference on Collective Phenomena*, edited by J. L. Lebowitz (New York Academy of Sciences, New York, 1981), p. 54.

Prolongation structures of nonlinear equations and infinite-dimensional algebras

Minoru Omote
Institute of Physics, University of Tsukuba, Ibaraki 305, Japan

(Received 30 December 1985; accepted for publication 14 August 1986)

Prolongation structures of the sine–Gordon equation, the Ernst equation, and the chiral model are systematically discussed. It is shown that the prolongation structures generate the Kac–Moody algebra for the sine–Gordon equation and another type of infinite-dimensional algebra for the Ernst equation. This algebra includes the Kac–Moody algebra and the Virasoro algebra as its subalgebra.

I. INTRODUCTION

As is well known, completely integrable models in two-dimensional space-time have common features. They have the Bäcklund transformations, an infinite number of conserved quantities, and the Lax pairs. They can be solved by the inverse scattering method or the Riemann–Hilbert transformations. There are the Kac–Moody algebras associated with these integrable models.

In this paper we will discuss another common feature of integrable models called the prolongation structure.[1–4] It will be shown that the structure plays important roles in obtaining the Bäcklund transformations and the Lax pairs, and that it reveals infinite-dimensional algebras that these models have implicitly.

As an introduction to the prolongation structure of the nonlinear equation we will consider the Bäcklund transformation and the Lax pair of the sine–Gordon equation. We know that the Bäcklund transformation of the sine–Gordon equation is given by

$$\begin{aligned} \partial_\xi \phi' &= -\partial_\xi \phi - 2\lambda \sin \tfrac{1}{2}(\phi - \phi'), \\ \partial_\eta \phi' &= \partial_\eta \phi + (2/\lambda) \sin \tfrac{1}{2}(\phi + \phi'), \end{aligned} \tag{1.1}$$

where λ is an arbitrary constant. Now we will define a pseudopotential q by

$$q = \tan[(\phi + \phi')/4], \tag{1.2}$$

then it can be shown that q satisfies the Riccati-type differential equations

$$\begin{aligned} \partial_\xi q &= \lambda \cos\phi\, q + (\lambda/2) \sin\phi (q^2 - 1), \\ \partial_\eta q &= \tfrac{1}{2}\partial_\eta \phi (q^2 + 1) + (1/\lambda) q, \end{aligned} \tag{1.3}$$

and that (1.1) can be rewritten in terms of q as

$$\begin{aligned} \partial_\xi \phi' = &-\partial_\xi \phi + [2\lambda/(q^2+1)] \\ &\times \{2q \cos\phi + (q^2 - 1)\sin\phi\}, \\ \partial_\eta \phi' = &\ \partial_\eta \phi + (4/\lambda) q/(q^2 + 1). \end{aligned} \tag{1.4}$$

Next we will express q in terms of two functions ψ_1 and ψ_2

$$q = \psi_1/\psi_2, \tag{1.5}$$

then we have a linear auxiliary equation (the Lax pair)

$$\begin{aligned} \partial_\xi \psi &= \frac{\lambda}{2}\begin{bmatrix} \cos\phi & -\sin\phi \\ -\sin\phi & -\cos\phi \end{bmatrix}\psi, \\ \partial_\eta \psi &= \frac{1}{2}\begin{bmatrix} 1/\lambda & \partial_\eta\phi \\ -\partial_\eta \phi & -1/\lambda \end{bmatrix}\psi, \end{aligned} \tag{1.6}$$

where $\psi = \binom{\psi_1}{\psi_2}$.

Thus we can see that the pseudopotential plays an important role in the discussions of the Bäcklund transformations and the Lax pair. In this point of view it is valuable to find a pseudopotential for a nonlinear equation. This problem to obtain a differential equation for a pseudopotential can be formulated in the method of the prolongation.

In the next section we will discuss the prolongation structure of the sine–Gordon equation and show that it gives an infinite-dimensional algebra associated with the sine–Gordon equation as well as the Riccati-type equation (1.3).

In Secs. III and IV prolongation structures of the Ernst equation and the chiral model are discussed.

II. SINE–GORDON EQUATION

The field equation of the sine–Gordon equation is given in terms of two-forms α_i $(i = 1,2)$:

$$\begin{aligned} \alpha_1 &= d\phi \wedge d\xi - \pi\, d\eta \wedge d\xi, \\ \alpha_2 &= d\pi \wedge d\eta + \sin\phi\, d\eta \wedge d\xi. \end{aligned} \tag{2.1}$$

For the systems of two-forms (2.1) we will assume that prolongation forms can be given by one-forms Ω^i $(i = 1,2,\ldots,p)$,

$$\Omega^i = -dq^i + F^i(\pi,\phi,q)d\eta + G^i(\pi,\phi,q)d\xi, \tag{2.2}$$

where p can be determined later, and F^i and G^i are functions of field variables ϕ,π and the newly introduced pseudopotentials q^i.

From the integrability condition of q^i, which can be expressed as

$$d\Omega^i \in I(\Omega,\alpha), \tag{2.3}$$

where $I(\Omega,\alpha)$ is an ideal generated by the set $\{\alpha^i\}$ and $\{\Omega^i\}$, we have differential equations for F^i and G^i

$$\begin{aligned} &\partial\phi\, F^i = 0, \quad \partial\pi\, G^i = 0, \\ &G^j \partial_j F^i - F^j \partial_j G^i + \sin\phi\, \partial\pi\, F^i - \pi\, \partial\phi\, G^i = 0. \end{aligned} \tag{2.4}$$

 0022-2488/86/122853-08$02.50

In (2.4) $\partial_i = \partial/\partial q_j$. By solving these equations we find[4]

$$F^i = X_0^i + X_1^i \pi, \quad G^i = Y_0^i \sin\phi + Y_1^i \cos\phi, \tag{2.5}$$

where X_a^i and Y_a^i $(a = 0,1)$ are functions of q^i only, and they are assumed to satisfy the following equations:

$$\begin{aligned} &X_0^j \partial_j Y_0^i - Y_0^j \partial_j X_0^i = X_1^i, \quad X_0^j \partial_j Y_1^i - Y_0^j \partial_j X_0^i = 0, \\ &X_1^j \partial_j Y_0^i - Y_0^j \partial_j X_1^i = Y_1^i, \\ &X_1^j \partial_j Y_1^i - Y_1^j \partial_j X_0^i = -Y_0. \end{aligned} \tag{2.6}$$

From (2.6) we can see that vector fields in a q-space (a prolongation space) X_a and Y_a defined by

$$X_a = X_a^i(q)\frac{\partial}{\partial q^i}, \quad Y_a = Y_a^i(q)\frac{\partial}{\partial q^i}, \tag{2.7}$$

satisfy the following commutator products (see Appendix A):

$$\begin{aligned} &[X_0, Y_0] = X_1, \quad [X_0, Y_1] = 0, \\ &[X_1, Y_0] = Y_1, \quad [X_1, Y_1] = -Y_0. \end{aligned} \tag{2.8}$$

Thus we find that the set of vectors $\{X_a, Y_a\}$ generates an incomplete set of commutator products, because $[X_0, X_1]$ and $[Y_0, Y_1]$ are not given yet.

In order to constitute a Lie algebra from (2.8) it can be shown that there are two ways. In the first course we will show that the set of commutator products (2.8) can be closed with finite number of vectors X_a and Y_a $(a = 0,1)$, and that they generate a finite-dimensional Lie algebra. In the second, on the other hand, we find that there appears an infinite-dimensional algebra and that (2.8) can be included in the set of commutator products of the elements. In the following we will consider both cases and will give explicit representations of the algebras in the prolongation space.

A. Finite-dimensional algebra

We will define vector fields X_2 and Y_2 by

$$X_2 = [X_0, X_1], \quad Y_2 = [Y_0, Y_1], \tag{2.9}$$

then from the Jacobi identity it can be shown that X_2 and Y_2 satisfy commutator products

$$\begin{aligned} &[X_2, Y_1] = -X_1, \quad [X_2, Y_0] = 0, \quad [Y_2, X_0] = Y_0, \\ &[Y_2, X_1] = 0, \quad [X_2, Y_2] = Y_1. \end{aligned} \tag{2.10}$$

Now we will assume that X_2 and Y_2 are given by linear combinations of X_0, X_1, Y_0, and Y_1. The coefficients of the linear combinations can be determined so that X_2 and Y_2 satisfy the above commutator products (2.9) and (2.10). Thus we find

$$X_2 = (1/\lambda^2)Y_0, \quad Y_2 = -\lambda^2 X_1, \tag{2.11}$$

where λ is an arbitrary parameter. With these results we have the Lie algebra of X_a, Y_a $(a = 1,2)$;

$$\begin{aligned} &[X_0, Y_0] = X_1, \quad [X_0, Y_1] = 0, \quad [X_0, X_1] = (1/\lambda^2)Y_0, \\ &[X_1, Y_0] = Y_1, \quad [X_1, Y_1] = -Y_0, \quad [Y_0, Y_1] = -\lambda^2 X_1. \end{aligned} \tag{2.12}$$

In this algebra the vector C defined by $C = \lambda^2 X_0 - Y_1$ commutes with all elements X_a, Y_a.

This finite-dimensional algebra can be shown to have a nonlinear representation in a one-dimensional prolongation space

$$\begin{aligned} &X_0 = \frac{1}{\lambda} q \frac{\partial}{\partial q}, \quad X_1 = \frac{1}{2}(q^2+1)\frac{\partial}{\partial q}, \\ &Y_0 = \frac{\lambda}{2}(q^2-1)\frac{\partial}{\partial q}, \quad Y_1 = \lambda q \frac{\partial}{\partial q}, \end{aligned} \tag{2.13}$$

and $C = 0$. In this case F and G have only one component, respectively, and are written as

$$\begin{aligned} &F = (1/\lambda)q + \tfrac{1}{2}(q^2+1)\pi, \\ &G = (\lambda/2)(q^2-1)\sin\phi + \lambda q \cos\phi. \end{aligned} \tag{2.14}$$

Then we find that the pseudopotential q satisfy the Riccati-type equations

$$\begin{aligned} &\partial_\eta q = (1/\lambda)q + \tfrac{1}{2}\pi(q^2+1), \\ &\partial_\xi q = \lambda q \cos\phi + (\lambda/2)(q^2-1)\sin\phi, \end{aligned} \tag{2.15}$$

which have the same forms as (1.3). As shown already we can give the Bäcklund transformations in terms of q (1.4).

B. Infinite-dimensional algebra

Now we will discuss the second case and will assume that X_2 and Y_2 are linearly independent from X_0, X_1, Y_0, and Y_1. Then we have an extended set of vectors that has six elements X_a, Y_a $(a = 0,1,2)$. The commutator products of this set, however, are still incomplete since $[X_0, X_2]$, $[X_1, X_2]$, $[Y_0, Y_2]$, and $[Y_1, Y_2]$ are not determined. Here we will again assume that these commutator products give new independent vectors and we find another incomplete set of commutator products. This procedure makes an infinite number of series and presents an infinite number of vectors as well as their commutator products.

It can be shown that these vectors generate the Kac–Moody algebra ($A_1^{(1)}$ type), and that the algebra can be represented in terms of infinite number of vectors $\overset{(i)}{T}_a$ $(a = 1,2,3,\ i = 0, \pm 1, \pm 2,..., \pm\infty)$ in the infinite-dimensional prolongation space (see Appendix A). With this representation the vector fields X_0, X_1, Y_0, and Y_1 are found to be given by some elements of the algebra:

$$X_0 = \overset{(1)}{T}_1, \quad X_1 = -i\overset{(0)}{T}_3, \quad Y_0 = -\overset{(-1)}{T}_2, \quad Y_1 = \overset{(-1)}{T}_1. \tag{2.16}$$

Then we can see that components of F and G have the following forms:

$$\begin{aligned} &F_1^i = \tfrac{1}{2} q_1^{(i+1)} + (\pi/2) q_2^{(i)}, \\ &F_2^i = -\tfrac{1}{2} q_2^{(i+1)} - (\pi/2) q_2^{(i)}, \\ &G_1^i = -\tfrac{1}{2} q_2^{(i-1)} \sin\phi + \tfrac{1}{2} q_1^{(i-1)} \cos\phi, \\ &G_2^i = -\tfrac{1}{2} q_1^{(i-1)} \sin\phi - \tfrac{1}{2} q_2^{(i-1)} \cos\phi, \end{aligned} \tag{2.17}$$

and that pseudopotentials satisfy equations given by

$$
\partial_\eta \begin{bmatrix} q_1^{(i)} \\ q_2^{(i)} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} q_1^{(i+1)} \\ q_2^{(i+1)} \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} q_1^{(i)} \\ q_2^{(i)} \end{bmatrix}, \tag{2.18}
$$

$$
\partial_\xi \begin{bmatrix} q_1^{(i)} \\ q_2^{(i)} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \cos\phi & -\sin\phi \\ -\sin\phi & -\cos\phi \end{bmatrix} \begin{bmatrix} q_1^{(i-1)} \\ q_2^{(i-1)} \end{bmatrix}.
$$

These results show that there is the infinite-dimensional algebra associated with the prolongation structure of the sine–Gordon equation, which guarantees the existence of the infinite number of pseudopotentials.

Next we will introduce a parameter-dependent potential $\psi(\lambda)$ by

$$
\begin{bmatrix} \psi_1(\lambda) \\ \psi_2(\lambda) \end{bmatrix} = \sum_{n=-\infty}^{\infty} (\lambda)^n \begin{bmatrix} q_1^{(n)} \\ q_2^{(n)} \end{bmatrix}. \tag{2.19}
$$

Then it can be shown that $\psi(\lambda)$ satisfies the linear auxiliary equation (1.6) of the sine–Gordon equation.

In the above discussions we showed that the prolongation structure of the sine–Gordon equation provides the finite-dimensional algebra as well as the infinite-dimensional Lie algebra. We also showed that the former algebra gave the pseudopotential of the Bäcklund transformations and that the latter algebra indicated the existence of the linear auxiliary equations.

III. ERNST'S EQUATION

The field equation is given by

$$
(\partial_\rho^2 + \partial_z^2 + (1/\rho)\partial_\rho)E = (1/T)(\partial_\rho E\,\partial_\rho E + \partial_z E\,\partial_z E), \tag{3.1}
$$

where E is the Ernst potential and $2T = E + E^*$. In terms of complex variables $x^1 = \rho - iz$, $x^2 = \rho + iz$, this equation can be rewritten as

$$
\partial_1 \partial_2 E = -(1/4\rho)(\partial_1 E + \partial_2 E) + (1/T)\partial_1 E\,\partial_2 E. \tag{3.2}
$$

We will define new fields a_i, b_i $(i = 1,2)$ by

$$
a_i = (1/2T)\partial_i E, \quad b_i = (1/2T)\partial_i E^*, \tag{3.3}
$$

then we find that these fields satisfy first-order differential equations

$$
\begin{aligned}
\partial_1 a_2 &= (a_1 - b_1)a_2 - (1/4\rho)(a_1 + a_2), \\
\partial_2 a_1 &= (a_2 - b_2)a_1 - (1/4\rho)(a_1 + a_2), \\
\partial_1 b_2 &= (b_1 - a_1)b_2 - (1/4\rho)(b_1 + b_2), \\
\partial_2 b_1 &= (b_2 - a_2)b_1 - (1/4\rho)(b_1 + b_2).
\end{aligned} \tag{3.4}
$$

Conversely it can be shown that if a_i, b_i satisfy the above equations we can obtain E, which satisfies Ernst's equation.

Equations for a_i and b_i are written in terms of two-forms as

$$
\begin{aligned}
\alpha_1 &= da_1 \wedge dx^1 + \{a_1(a_2 - b_2) \\
&\quad - (1/4\rho)(a_1 + a_2)\}dx^1 \wedge dx^2, \\
\alpha_2 &= da_2 \wedge dx^2 + \{a_2(a_1 - b_1) \\
&\quad - (1/4\rho)(a_1 + a_2)\}dx^2 \wedge dx^1, \\
\beta_1 &= db_1 \wedge dx^1 + \{b_1(a_2 - b_2) \\
&\quad + (1/4\rho)(b_1 + b_2)\}dx^2 \wedge dx^1, \\
\beta_2 &= db_2 \wedge dx^2 + \{b_2(a_1 - b_1) \\
&\quad + (1/4\rho)(b_1 + b_2)\}dx^1 \wedge dx^2.
\end{aligned} \tag{3.5}
$$

As before we will define the prolongation form Ω^i,

$$
\Omega^i = -dq^i + F^i dx^1 + G^i dx^2, \tag{3.6}
$$

then we can obtain prolongation structures for Ernst's equation[5]

$$
\begin{aligned}
F^i &= X_0^i a_1 + X_1^i b_1 + X_2^i, \\
G^i &= Y_0^i a_2 + Y_1^i b_2 + Y_2^i,
\end{aligned} \tag{3.7}
$$

where X_a^i, Y_a^i $(a = 0,1,2)$ are, in general, functions of independent variables x^1, x^2 as well as pseudopotentials. In this case vector fields X_a, Y_a, components of which are given by X_a^i, Y_a^i, satisfy an incomplete set of commutator products

$$
\begin{aligned}
&[X_0, Y_0] = X_0 - Y_0, \quad [X_1, Y_0] = -X_1 + Y_0, \\
&[X_0, Y_1] = -X_0 + Y_1, \quad [X_1, Y_1] = X_1 - Y_1,
\end{aligned} \tag{3.8}
$$

and

$$
\begin{aligned}
&[X_2, Y_0] = -(1/4\rho)(X_0 - Y_0) - \partial_1 Y_0, \\
&[X_2, Y_1] = -(1/4\rho)(X_1 - Y_1) - \partial_1 Y_1, \\
&[X_0, Y_2] = -(1/4\rho)(X_0 - Y_0) + \partial_2 X_0, \\
&[X_1, Y_2] = -(1/4\rho)(X_1 - Y_1) + \partial_2 X_1, \\
&[X_2, Y_2] = \partial_2 X_2 - \partial_1 Y_2.
\end{aligned} \tag{3.9}
$$

Now we will look for the Lie algebras associated with the sets of commutator products (3.8) and (3.9).

A. Finite-dimensional algebra

Vector fields X_3 and Y_3 defined by

$$
X_3 = [X_0, X_1], \quad Y_3 = [Y_0, Y_1], \tag{3.10}
$$

are assumed to be given in the linear combinations of X_a, Y_a $(a = 0,1)$ and are found to be

$$
X_3 = -\zeta^2(Y_0 - Y_1), \quad Y_3 = -\zeta^{-2}(X_0 - X_1). \tag{3.11}
$$

Here the coefficient ζ remains to be undetermined. From (3.8) and (3.11) we see that vector fields X_a and Y_a $(a = 0,1)$ generate a closed algebra and it can be shown that in the one-dimensional prolongation space they have their representations of the form

$$
\begin{aligned}
X_0 &= -q(1 + \zeta q)\frac{\partial}{\partial q}, \quad X_1 = (\zeta + q)\frac{\partial}{\partial q}, \\
Y_0 &= -q(1 + \zeta^{-1} q)\frac{\partial}{\partial q}, \quad Y_1 = (q + \zeta^{-1})\frac{\partial}{\partial q}.
\end{aligned} \tag{3.12}
$$

From (3.9) and (3.12) we see that $X_2 = 0$ and $Y_2 = 0$ and that ζ is not a constant parameter, but it satisfies

$$\partial_1\zeta = (\zeta/4\rho)(\zeta^2-1),\quad \partial_2\zeta = (1/4\rho\zeta)(\zeta^2-1)\,. \tag{3.13}$$

These representations of X_a and Y_a give the Riccati-type equations of the pseudopotential q

$$\begin{aligned}\partial_1 q &= -q(1+\zeta q)a_1 + (q+\zeta)b_1\,,\\ \partial_2 q &= -q(1+\zeta^{-1}q)a_2 + (q+\zeta^{-1})b_2\,.\end{aligned} \tag{3.14}$$

On the other hand, we know that the Bäcklund transformation of the Ernst equation is given by[6]

$$\begin{aligned}a_1' &= \frac{1}{2\theta}\Big\{(1-\zeta^2)(\theta^2+1)+2\zeta^2\theta\\ &\quad + (1-\zeta^2)(\theta-1)\sqrt{\theta^2+2\frac{1+\zeta^2}{1-\zeta^2}\theta+1}\Big\}a_1\\ &\quad + \frac{(1-\zeta^2)}{8\rho\theta}\Big(\theta-1+\sqrt{\theta^2+2\frac{1+\zeta^2}{1-\zeta^2}\theta+1}\Big),\\ a_2' &= \frac{1}{2\zeta^2\theta}\Big\{(1-\zeta^2)(\theta^2+1)\\ &\quad + 2\theta+(1-\zeta^2)(\theta+1)\sqrt{\theta^2+2\frac{1+\zeta^2}{1-\zeta^2}\theta+1}\Big\}a_2\\ &\quad - \frac{1-\zeta^2}{8\rho\theta\zeta^2}\Big(\theta+1+\sqrt{\theta^2+2\frac{1+\zeta^2}{1-\zeta^2}\theta+1}\Big),\end{aligned} \tag{3.15}$$

where $\theta = (i/\rho)TT'$. We will introduce q by $\theta = (\zeta^2-1)^{-1}(q+\zeta)(1+\zeta q)q^{-1}$, which can be shown to satisfy (3.14), then we can rewrite the Bäcklund transformation in the form[5,7]

$$\begin{aligned}a_1' &= -\frac{q(1+\zeta q)}{q+\zeta}a_1 + \frac{1}{4\rho}\frac{q(1-\zeta^2)}{q+\zeta}\,,\\ a_2' &= -\frac{q(1+\zeta^{-1}q)}{q+\zeta^{-1}}a_2 - \frac{1}{4\rho}\frac{q(1-\zeta^2)}{\zeta(1+\zeta q)}\,.\end{aligned} \tag{3.16}$$

B. Infinite-dimensional algebra

By repeating the same procedure in the previous section we can show that (3.8) and (3.9) belong to the coupled Kac–Moody and Virasoro algebra of $\overset{(m)}{T_a}$ and $\overset{(m)}{D}$ (see Appendix A) and the X_a and Y_a are expressed by

$$\begin{aligned}X_0 &= (T_-^{(-1)}+T_3^{(0)}),\quad X_1=(T_+^{(-1)}-T_3^{(0)})\,,\\ X_2 &= -(1/4\rho)(D^{(-2)}-D^{(0)})\,,\\ Y_0 &= (T_-^{(1)}+T_3^{(0)}),\quad Y_1=(T_+^{(1)}-T_3^{(0)})\,,\\ Y_2 &= -(1/4\rho)(D^{(0)}-D^{(2)})\,.\end{aligned} \tag{3.17}$$

These representations indicate that there are an infinite number of pseudopotentials that satisfy the following differential equations:

$$\begin{aligned}dq_1^{(n)} &= \tfrac12\{(q_1^{(n-1)}+iq_2^{(n)}-iq_2^{(n-1)})a_1+(q_1^{(n-1)}-iq_2^{(n)}+iq_2^{(n-1)})b_1\}dx^1\\ &\quad+\tfrac12\{(q_1^{(n+1)}+iq_2^{(n)}-iq_2^{(n+1)})a_2+(q_1^{(n+1)}-iq_2^{(n)}+iq_2^{(n+1)})b_2\}dx^2\\ &\quad-(1/4\rho)\{(n-2)q_1^{(n-2)}-nq_1^{(n)}\}dx^1-(1/4\rho)\{nq_1^{(n)}-(n+2)q_1^{(n+2)}\}dx^2\,,\\ dq_2^{(n)} &= \tfrac12\{-(q_2^{(n-1)}+iq_1^{(n)}+iq_1^{(n-1)})a_1+(-q_2^{(n-1)}+iq_1^{(n)}+iq_1^{(n-1)})b_1\}dx^1\\ &\quad+\tfrac12\{-(q_2^{(n+1)}+iq_1^{(n)}+iq_1^{(n+1)})a_2+(-q_2^{(n+1)}+iq_1^{(n)}+iq_1^{(n+1)})b_2\}dx^2\\ &\quad-(1/4\rho)\{(n-2)q_2^{(n-2)}-nq_2^{(n)}\}dx^1(1/4\rho)\{nq_2^{(n)}-(n+2)q_2^{(n+2)}\}dx^2\,.\end{aligned} \tag{3.18}$$

In this case it can be seen that (3.18) cannot be summed up in terms of the parameter-dependent potential $\psi(\lambda)$, since in the right-hand side of (3.18) we have terms with coefficients depending on n. However we will consider another kind of potential $\psi(\zeta)$

$$\begin{bmatrix}\psi_1(\zeta)\\ \psi_2(\zeta)\end{bmatrix} = \sum_{n=-\infty}^{\infty}\zeta^n\begin{bmatrix}q_1^{(n)}\\ q_2^{(n)}\end{bmatrix}, \tag{3.19}$$

where ζ is a function of independent variables, then

$$d\psi_i = \sum_{n=-\infty}^{\infty}\zeta^n\,dq_i^{(n)} + \sum_{n=-\infty}^{\infty} n\zeta^{n-1}q_i^{(n)}\,d\zeta\,. \tag{3.20}$$

In (3.20) $d\zeta$ can be determined so that the second term of the right-hand side cancels the n-dependent terms in (3.18). Thus we find (3.13) again. From (3.18) and (3.20) we obtain a linear auxiliary equation for the Ernst equation

$$\begin{aligned}d\psi &= \begin{bmatrix}(\zeta/2)(a_1+b_1) & (i/2)(1-\zeta)(a_1-b_1)\\ -(i/2)(1+\zeta)(a_1-b_1) & -(\zeta/2)(a_1+b_1)\end{bmatrix}\psi\,dx^1\\ &\quad+\begin{bmatrix}(\zeta^{-1}/2)(a_2+b_2) & (i/2)(1-\zeta^{-1})(a_2-b_2)\\ -(i/2)(1+\zeta^{-1})(a_2-b_2) & -(\zeta^{-1}/2)(a_2+b_2)\end{bmatrix}\psi\,dx^2\,,\end{aligned} \tag{3.21}$$

which was found before.[7]

IV. CHIRAL MODEL

A field equation of the chiral model is given by

$$\partial_y(g^{-1}\partial_x g) + \partial_x(g^{-1}\partial_y g) = 0, \tag{4.1}$$

where g is a $r\times r$ matrix field with $\det g = 1$. We will define new matrices A and B as

$$A = g^{-1}\partial_x g, \quad B = g^{-1}\partial_y g, \tag{4.2}$$

where $\mathrm{Tr}\,A = \mathrm{Tr}\,B = 0$, then we find these matrices satisfy first-order field equations

$$\begin{aligned}\partial_y A &= \tfrac{1}{2}(AB - BA), \\ \partial_x B &= -\tfrac{1}{2}(AB - BA).\end{aligned} \tag{4.3}$$

These field equations are rewritten in terms of $r\times r$ matrix valued two-forms and

$$\begin{aligned}\alpha &= dA\wedge dx + \tfrac{1}{2}(AB - BA)dx\wedge dy, \\ \beta &= dB\wedge dy + \tfrac{1}{2}(AB - BA)dx\wedge dy.\end{aligned} \tag{4.4}$$

As in the previous sections we assume that prolongation forms are given by

$$\Omega^i = -dq^i + F^i dx + G^i dy, \tag{4.5}$$

where F^i and G^i are functions of q^i, A, and B. Then we have

$$\begin{aligned}&\nabla(A^a{}_b)G^i = 0, \quad \nabla(B^a{}_b)F^i = 0, \\ &G^j\partial_j F^i - F^j\partial_j G^i \\ &\quad + \tfrac{1}{2}\{\nabla(A^a{}_b)F^i + \nabla(B^a{}_b)G^i\}(AB - BA)^a = 0,\end{aligned} \tag{4.6}$$

where

$$\nabla(A^a{}_b)F^i = \left(\frac{\partial F^i}{\partial A^a{}_b} - \frac{1}{r}\sum_c \frac{\partial F^i}{\partial A^c{}_c}\delta^b{}_a\right).$$

From (4.6) we have the prolongation structure

$$F^i = X^{b,i}_a A^a{}_b, \quad G^i = Y^{b,i}_a B^a{}_b, \tag{4.7}$$

and

$$[X^a{}_b, Y^c{}_d] = \tfrac{1}{2}\delta^a{}_d(X+Y)^c{}_b - \tfrac{1}{2}\delta^c{}_b(X+Y)^a{}_d, \tag{4.8}$$

where $X^a{}_b$ and $Y^a{}_b$ are vector fields in the prolongation space and are defined by

$$X^a{}_b = X^{a,i}_b \frac{\partial}{\partial q^i}, \quad Y^a{}_b = Y^{a,i}_b \frac{\partial}{\partial q^i}. \tag{4.9}$$

In order to have explicit forms of vector fields X and Y we will discuss both possibilities of a finite-dimensional algebra and an infinite-dimensional algebra that include (4.8) in commutator products of their elements.

A. Finite-dimensional algebra

We will assume that commutator products $[X^a{}_b, X^c{}_d]$ and $[Y^a{}_b, Y^c{}_d]$ can be expressed in terms of linear combinations of $X^a{}_b$ and $Y^a{}_b$. Then the Jacobi identities determine coefficients of the linear combinations and we find

$$\begin{aligned}[X^a{}_b, X^c{}_d] &= \tfrac{1}{2}(\delta^a{}_d X^c{}_b - \delta^c{}_b X^a{}_d) \\ &\quad + (\lambda^2/2)(\delta^a{}_d Y^c{}_b - \delta^c{}_b Y^a{}_d), \\ [Y^a{}_b, Y^c{}_d] &= (1/2\lambda^2)(\delta^a{}_d X^c{}_b - \delta^c{}_b X^a{}_d) \\ &\quad + \tfrac{1}{2}(\delta^a{}_d Y^c{}_b - \delta^c{}_b Y^a{}_d),\end{aligned} \tag{4.10}$$

where λ is an arbitrary parameter. Equations (4.8) and (4.10) indicate that vector fields $X^a{}_b$ and $Y^a{}_b$ generate finite-dimensional algebras, which have parameter-dependent structure constants. It can be shown that representations of $X^a{}_b$ and $Y^a{}_b$ are given by

$$\begin{aligned}X^a{}_b &= \frac{1}{2(\beta+1)}\{q^a{}_d q^c{}_b + q^c{}_b\delta^a{}_d \\ &\quad - (2\beta+1)q^a{}_d\delta^c{}_b + \delta^a{}_d\delta^c{}_b\}\frac{\partial}{\partial q^c{}_d}, \\ Y^a{}_b &= \frac{1}{2(\beta-1)}\{q^a{}_d q^c{}_b - q^c{}_b\delta^a{}_d \\ &\quad - (2\beta-1)q^a{}_d\delta^c{}_b + \delta^a{}_c\delta^d{}_b\}\frac{\partial}{\partial q^c{}_d},\end{aligned} \tag{4.11}$$

where $\beta = (1+\lambda^2)/(1-\lambda^2)$. From (4.5), (4.7), (4.9), and (4.11) we have

$$\begin{aligned}dq^a{}_b &= [1/2(\beta+1)]\{q^a{}_c A^c{}_d q^d{}_b + q^a{}_c A^c{}_b \\ &\quad - (2\beta+1)A^a{}_c q^c{}_b + A^a{}_b\}dx \\ &\quad + [1/2(\beta-1)]\{q^a{}_c B^c{}_d q^d{}_b \\ &\quad - q^a{}_c B^c{}_b - (2\beta-1)B^a{}_c q^c{}_b + B^a{}_b\}dy,\end{aligned} \tag{4.12}$$

which was obtained before,[8] and it can be shown that the Bäcklund transformation can be expressed in terms of $q^a{}_b$.

B. Infinite-dimensional algebra

We can show that an infinite-dimensional algebra associated with the prolongation structure of the chiral model is given by

$$\begin{aligned}\left[\overset{(m)}{M}{}^a{}_b, \overset{(n)}{M}{}^c{}_d\right] &= \delta^c{}_b \overset{(m+n+1)}{M}{}^a{}_d - \delta^a{}_d \overset{(m+n+1)}{M}{}^c{}_b, \\ \left[\overset{(m)}{N}{}^a{}_b, \overset{(n)}{N}{}^c{}_d\right] &= \delta^c{}_b \overset{(m+n+1)}{N}{}^a{}_d - \delta^a{}_d \overset{(m+n+1)}{N}{}^c{}_b, \\ \left[\overset{(m)}{M}{}^a{}_b, \overset{(n)}{N}{}^c{}_d\right] &= \sum_{l=0}^{m}\binom{m+n-l}{n}(\delta^c{}_b\overset{(l)}{M}{}^a{}_d - \delta^a{}_d\overset{(l)}{M}{}^c{}_b) \\ &\quad + \sum_{l=0}^{n}\binom{m+n-l}{m}(\delta^c{}_d\overset{(l)}{N}{}^a{}_d - \delta^a{}_d\overset{(l)}{N}{}^c{}_b),\end{aligned} \tag{4.13}$$

where m and n run from 0 to $+\infty$ and $\binom{m}{n}$ is the binomial coefficients. It can be shown that the vector fields $\overset{(m)}{M}{}^a{}_b$ and $\overset{(m)}{N}{}^a{}_b$ are represented in the prolongation space as follows (see Appendix B):

$$\overset{(m)}{M}{}^a{}_b = (-2)^{m+1}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{k-m-1}\binom{k-l-1}{m} q_l{}^a\frac{\partial}{\partial q_k{}^b},$$

$$\overset{(m)}{N}{}^a{}_b = (-2)^{m+1}$$
$$\times\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{k-m-1}\binom{k-l-1}{m}(-1)^{k-l} q_l{}^a\frac{\partial}{\partial q_k{}^b}. \tag{4.14}$$

Since $X^a{}_b$ and $Y^a{}_b$ are given by

$$X^a{}_b = -\tfrac{1}{2}\overset{(0)}{M}{}^a{}_b,\quad Y^a{}_b = -\tfrac{1}{2}\overset{(0)}{N}{}^a{}_b, \tag{4.15}$$

we have

$$X^a{}_b = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{k-1} q_l{}^a\frac{\partial}{\partial q_k^b} = \delta^c{}_b\sum_{l=-\infty}^{\infty}\sum_{l=-\infty}^{k-1} q_l{}^a\frac{\partial}{\partial q_k^c},$$
$$Y^a{}_b = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{k-1}(-1)^{k-l}q_l{}^a\frac{\partial}{\partial q_k{}^b} \tag{4.16}$$
$$= \delta^c{}_b\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{k-1}(-1)^{k-l}q_l{}^a\frac{\partial}{\partial q_k{}^c},$$

and

$$F_k{}^a = (X^c{}_b)_k{}^a A^b{}_c = \sum_{l=-\infty}^{k-1} q_l{}^b A^a{}_b,$$
$$G_k{}^a = (Y^c{}_b)_k{}^a B^b{}_c = \sum_{l=-\infty}^{k-1}(-1)^{k-l}q_l{}^b B^a{}_b. \tag{4.17}$$

With this representation we see that an infinite number of pseudopotentials satisfy

$$dq_k{}^a = A^a{}_b\sum_{l=-\infty}^{k-1} q_l{}^b\,dx + B^a{}_b\sum_{l=-\infty}^{k-1}(-1)^{k-l}q_l{}^b\,dy. \tag{4.18}$$

Now we will introduce a parameter-dependent potential $\psi^a(\lambda)$ by

$$\psi^a(\lambda) = \sum_{\lambda=-\infty}^{\infty}(\lambda)^k q_k{}^a, \tag{4.19}$$

then we obtain a linear auxiliary equation for the chiral model[9]

$$d\psi = [\lambda/(1-\lambda)]A\psi\,dx - [\lambda/(1+\lambda)]B\psi\,dy. \tag{4.20}$$

V. CONCLUSIONS

In the previous sections we have shown that the prolongation structures of the sine–Gordon equation, the Ernst equation, and the chiral model give incomplete sets of comutator products of vector fields in the prolongation space. It was found that there were two ways of constructing the Lie algebras that include these sets of commutator products. In the first way we found finite-dimensional algebras that have bilinear representations of vector fields in the prolongation spaces. From these representations we obtained the Riccati-type equations of pseudopotentials for the Bäcklund transformations.

In the second way there appeared infinite-dimensional algebras associated with the nonlinear equations. It was shown that the linear representations of these algebras gave the linear auxiliary equations.

Then we can summarize the above-mentioned results of the prolongation structure as follows:

nonlinear equation
↓
prolongation structure
↓ finite-dimensional algebra → the Riccati equation → the Bäcklund transformation
↓ infinite-dimensional algebra → the Lax pair → inverse scattering problem.

Here we have to notice that the Ernst equation has the new kind of infinite-dimensional algebra that includes the Kac–Moody algebra and the Virasoro algebra as its subalgebra, and that the x-dependent spectral function $\zeta(x)$ comes from the appearance of the Virasoro algebra.

In this paper we have discussed only nonlinear equations in two-dimensional space-time. However, I think that our method of the prolongation structure can be generalized to be applicable to higher-dimensional nonlinear equations.[10–12]

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor Ge Mo Lin for helpful discussions.

APPENDIX A: THE KAC–MOODY AND THE VIRASORO ALGEBRA

We will consider a set of vector fields V_a in a prolongation space. These vector fields V_a are written as

$$V_a = V_a^{(m)}(q)\frac{\partial}{\partial q^m}, \tag{A1}$$

and their commutator products are defined by

$$[V_a, V_b] = (V_a^{(n)}\partial_n V_b^{(m)} - V_b^{(n)}\partial_n V_a^{(m)})\frac{\partial}{\partial q^m}, \tag{A2}$$

where $\partial_n = \partial/\partial q^n$. With the definition of the commutator product (A2) the set of vector fields $\{V_a\}$ can be shown to be a Lie algebra. The functions $V_a^{(m)}(q)$ are called components of V_a.

Now we will discuss a Lie algebra of vector fields in an infinite-dimensional prolongation space that has coordinate variables $\{q_i^{(m)};\ i = 1,2,\ldots,r,\ m = 0\pm 1, \pm 2\ldots, \pm\infty\}$. These vector fields $\{\overset{(m)}{A}{}_{ij}\}$ are assumed to have the following forms:

$$\overset{(m)}{A}{}_{ij} = \sum_{n=-\infty}^{\infty} q_i^{(n+m)}\frac{\partial}{\partial q_j^{(n)}}. \tag{A3}$$

It can be shown that they satisfy commutator products

$$\left[\overset{(m)}{A}{}_{ij}, \overset{(n)}{A}{}_{kl}\right] = \delta_{jk}A_{il}^{(n+m)} - \delta_{il}A_{kj}^{(m+n)}. \tag{A4}$$

This fact indicates that the set of vector fields $\{\overset{(m)}{A}_{ij}\}$ constitute an infinite-dimensional graded algebra.

In the following we will consider a duplicated infinite-dimensional prolongation space $\{q_i^{(m)}, i = 1,2, m = 0, \pm 1,..., \pm \infty\}$ and define a set of vector fields $\overset{(m)}{T}_a$ $(a = 1,2,3,\ m = 0, \pm 1,..., \pm \infty)$ that are linear combinations of $\overset{(m)}{A}_{ij}$:

$$\overset{(m)}{T}_1 = \frac{1}{2} \sum_{n=-\infty}^{\infty} \left\{ q_1^{(n+m)} \frac{\partial}{\partial q_1^{(n)}} - q_2^{(n+m)} \frac{\partial}{\partial q_2^{(n)}} \right\},$$
$$\overset{(m)}{T}_2 = \frac{1}{2} \sum_{n=-\infty}^{\infty} \left\{ q_2^{(n+m)} \frac{\partial}{\partial q_1^{(n)}} + q_1^{(n+m)} \frac{\partial}{\partial q_2^{(n)}} \right\}, \quad \text{(A5)}$$
$$\overset{(m)}{T}_3 = \frac{i}{2} \sum_{n=-\infty}^{\infty} \left\{ q_2^{(n+m)} \frac{\partial}{\partial q_1^{(n)}} - q_1^{(n+m)} \frac{\partial}{\partial q_2^{(n)}} \right\}.$$

It can be easily shown that $\overset{(m)}{T}_i$ satisfies commutator products

$$\left[\overset{(m)}{T}_i, \overset{(n)}{T}_j\right] = i\epsilon_{ijk} \overset{(m+n)}{T_k}, \quad \text{(A6)}$$

where ϵ_{ijk} is a completely antisymmetric structure constant and $\epsilon_{123} = 1$. If we define vectors $\overset{(m)}{T}_{\pm} = \overset{(m)}{T}_1 \pm i\overset{(m)}{T}_2$, then they satisfy commutator brackets

$$\left[\overset{(m)}{T}_3, \overset{(n)}{T}_{\pm}\right] = \overset{(m+n)}{T}_{\pm}, \quad \left[\overset{(m)}{T}_+, \overset{(n)}{T}_-\right] = 2\overset{(m+n)}{T}_3. \quad \text{(A7)}$$

Next we will define new vectors $\overset{(m)}{D}$ $(m = 0, \pm 1,..., \pm \infty)$ by

$$\overset{(m)}{\mathrm{D}} = \sum_{n=-\infty}^{\infty} (n+m)\left\{ q_1^{(n+m)} \frac{\partial}{\partial q_1^{(n)}} + q_2^{(n+m)} \frac{\partial}{\partial q_2^{(n)}} \right\}, \quad \text{(A8)}$$

then we find that they satisfy

$$\left[\overset{(m)}{D}, \overset{(n)}{T}_i\right] = n\overset{(m+n)}{T}_i,$$
$$\left[\overset{(m)}{D}, \overset{(n)}{D}\right] = (n-m)\overset{(m+n)}{D}. \quad \text{(A9)}$$

The commutator products (A6) and (A9) show that both sets of vectors $\{\overset{(m)}{T}_i\}$ and $\{\overset{(m)}{T}_i, \overset{(n)}{D}\}$ constitute the Kac–Moody algebra and the coupled Kac–Moody and Virasoro algebra. In Secs. II and III we have used the representations of vectors $\overset{(m)}{T}_i$ and $\overset{(m)}{D}$ given by (A5) and (A8).

APPENDIX B: ALGEBRA OF CHIRAL MODEL

In this appendix we will show that vector fields $\overset{(m)}{M}{}^a{}_b$, $\overset{(m)}{N}{}^a{}_b$ given by (4.14) satisfy commutator products (4.13). From the definitions we have

$$\left[\overset{(m)}{M}{}^a{}_b, \overset{(n)}{M}{}^c{}_d\right]$$
$$= (-2)^{m+n+2} \times \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{k-2-m-n} \sum_{s=m}^{k-l-2-n} \binom{s}{m}\binom{k-l-2-s}{n} \times \left\{ \delta^c{}_b q_l{}^a \frac{\partial}{\partial q_k{}^d} - \delta_d{}^a q_l{}^c \frac{\partial}{\partial q_k^b} \right\},$$

$$\left[\overset{(m)}{N}{}^a{}_b, \overset{(n)}{N}{}^c{}_d\right]$$
$$= (-2)^{m+n+2} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{k-2-m-n} (-1)^{k-l} \times \sum_{s=m}^{k-l-2-n} \binom{s}{m}\binom{k-l-2-s}{n} \times \left\{ \delta^c{}_b q_l{}^a \frac{\partial}{\partial q_k{}^d} - \delta_d{}^a q_l{}^c \frac{\partial}{\partial q_K{}^b} \right\},$$

$$\left[\overset{(m)}{M}{}^a{}_b, \overset{(n)}{N}{}^c{}_d\right]$$
$$= -(-2)^{m+n+2} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{k-2-m-n} (-1)^{k-l} \times \sum_{s=m}^{k-l-2-n} (-1)^s \binom{s}{m}\binom{k-l-2-s}{n} \times \left\{ \delta^c{}_b q_l{}^a \frac{\partial}{\partial q_k{}^d} - \delta_d{}^a q_l{}^a \frac{\partial}{\partial q_k{}^b} \right\}, \quad \text{(B1)}$$

which show that traces of $\overset{(m)}{M}{}^a{}_b$ and $\overset{(m)}{\mathrm{N}}{}^a{}_b$ commute with all vectors. Since

$$\sum_{s=m}^{k-l-2-n} \binom{s}{m}\binom{k-l-2-s}{n} = \binom{k-l-1}{m+n+1},$$

the first two commutator products of (B1) become

$$\left[\overset{(m)}{M}{}^a{}_b, \overset{(n)}{M}{}^c{}_d\right] = \delta^c{}_b \overset{(m+n+1)}{m}{}^a{}_d - \delta^a{}_d \overset{(m+n+1)}{M}{}^c{}_b,$$
$$\left[\overset{(m)}{N}{}^a{}_b, \overset{(n)}{N}{}^c{}_d\right] = \delta^c{}_b \overset{(m+n+1)}{N}{}^a{}_d - \delta^a{}_d \overset{(m+n+1)}{N}{}^c{}_b. \quad \text{(B2)}$$

In order to show the last equation of (4.13) we will use relations

$$\overset{(m+1)}{M}{}^a{}_b = \frac{1}{r}\left\{ \sum_c \left[\overset{(m)}{M}{}^a{}_c, \overset{(0)}{M}{}^c{}_b\right] + \delta^a{}_b T_r \overset{(m+1)}{M} \right\},$$
$$\overset{(m+1)}{N}{}^a{}_b = \frac{1}{r}\left\{ \sum_c \left[\overset{(m)}{N}{}^a{}_c, \overset{(0)}{N}{}^c{}_b\right] + \delta^a{}_b T_r \overset{(m+1)}{N} \right\}, \quad \text{(B3)}$$

which can be obtained from (B2). Now we will consider the commutator product $\left[\overset{(m)}{M}{}^a{}_b, \overset{(0)}{N}{}^c{}_d\right]$, and show by the mathematical induction that this commutator product can be given by

$$\left[\overset{(m)}{M}{}^a{}_b, \overset{(0)}{N}{}^c{}_d\right] = \sum_{l=0}^{m} (\delta^c{}_b \overset{(l)}{M}{}^a{}_d - \delta^a{}_d \overset{(l)}{M}{}^c{}_b) + (\delta^c{}_b \overset{(0)}{N}{}^a{}_d - \delta^a{}_d \overset{(0)}{N}{}^c{}_b). \quad \text{(B4)}$$

From (B1) it can be shown that for $m = 0$ (B4) is satisfied. By using (B3) we have

$$\left[\overset{(m+1)}{M}{}^{a}{}_{b}, \overset{(0)}{N}{}^{c}{}_{d}\right] = \frac{1}{r}\sum_{e}\left[\left[\overset{(m)}{M}{}^{a}{}_{e}, \overset{(0)}{M}{}^{e}{}_{b}\right], \overset{(0)}{N}{}^{c}{}_{d}\right]$$
$$= -\frac{1}{r}\sum_{e}\left\{\left[\left[\overset{(0)}{M}{}^{e}{}_{b}, \overset{(0)}{N}{}^{c}{}_{d}\right], \overset{(m)}{M}{}^{a}{}_{e}\right] + \left[\left[\overset{(0)}{N}{}^{c}{}_{d}, \overset{(m)}{M}{}^{a}{}_{e}\right], \overset{(0)}{M}{}^{e}{}_{b}\right]\right\}, \quad \text{(B5)}$$

where we used the Jacobi identity and used the fact that Tr M commutes with all vectors M and N. Assuming (B4) in (B5) we can obtain

$$\left[\overset{(m+1)}{M}{}^{a}{}_{b}, \overset{(0)}{N}{}^{c}{}_{d}\right] = \sum_{l=0}^{m+1}\left(\delta^{c}{}_{b}\overset{(l)}{M}{}^{a}{}_{d} - \delta^{a}{}_{d}\overset{(l)}{N}{}^{c}{}_{b}\right)$$
$$+ \left(\delta^{c}{}_{b}\overset{(0)}{N}{}^{a}{}_{b} - \delta^{a}{}_{d}\overset{(0)}{N}{}^{c}{}_{b}\right). \quad \text{(B6)}$$

Thus we have shown that (B4) is satisfied by an arbitrary m.

Now we will proceed to show the last equation of (4.15) by the mathematical induction again. From the above discussions we know that it can be satisfied for $n = 0$. By using similar arguments used in proving (B6) we can show

$$\left[\overset{(m)}{M}{}^{a}{}_{b}, \overset{(n+1)}{N}{}^{c}{}_{d}\right]$$
$$= \frac{1}{r}\sum_{e}\left[\overset{(m)}{M}{}^{a}{}_{b}, \left[\overset{(n)}{N}{}^{c}{}_{e}, \overset{(0)}{N}{}^{e}{}_{d}\right]\right]$$
$$= -\frac{1}{r}\sum_{e}\left\{\left[\overset{(n)}{N}{}^{c}{}_{e}, \left[\overset{(0)}{N}{}^{e}{}_{d}, \overset{(m)}{M}{}^{a}{}_{b}\right]\right] + \left[\overset{(0)}{N}{}^{e}{}_{d}, \left[\overset{(m)}{M}{}^{a}{}_{b}, \overset{(n)}{N}{}^{c}{}_{e}\right]\right]\right\}$$
$$= \sum_{l=0}^{m}\binom{m+n+1-l}{n+1}\left(\delta^{c}{}_{b}\overset{(l)}{M}{}^{a}{}_{d} - \delta^{a}{}_{d}\overset{(l)}{M}{}^{c}{}_{b}\right)$$
$$+ \sum_{l=0}^{n+1}\binom{m+n+1-l}{m}\left(\delta^{c}{}_{b}\overset{(l)}{N}{}^{a}{}_{d} - \delta^{a}{}_{d}\overset{(l)}{N}{}^{c}{}_{b}\right). \quad \text{(B7)}$$

This shows that (4.13) can be satisfied for every $\overset{(m)}{M}{}^{a}{}_{b}$ and $\overset{(n)}{N}{}^{c}{}_{d}$.

[1]F. B. Estabrook and H. D. Wahlquist, J. Math. Phys. **16**, 1 (1975).
[2]F. B. Estabrook and H. D. Wahlquist, *Nonlinear Evolution Equations, Research Notes in Mathmatics*, No. 26 (Pitman, London, 1978).
[3]D. J. Kaup, Physica D **1**, 391 (1980).
[4]C. Hoenselaers, Prog. Theor. Phys. **74**, 645 (1985).
[5]B. K. Harrison, Phys. Rev. Lett. 41, 1197 (1978).
[6]M. Omote and M. Wadati, J. Math. Phys. **22**, 961 (1981).
[7]M. Omote and M. Wadati, Prog. Theor. Phys. **65**, 1621 (1981).
[8]L. L. Chau, Lect. Notes Phys. **189**, 110 (1982).
[9]V. E. Zakharov and A. V. Kikhailov, Sov. Phys. JETP **47**, 1017 (1978).
[10]H. C. Morris, J. Math. Phys. **17**, 1870 (1976).
[11]H. C. Morris, J. Math. Phys. **21**, 327 (1980).
[12]M. Gürses, Phys. Lett. **A 101**, 388 (1984).

On a new hierarchy of nonlinear evolution equations containing the Pohlmeyer–Lund–Regge equation

F. Pempinelli[a)]
Laboratoire de Physique Mathématique,[b)] Université des Sciences et Techniques du Languedoc, Place Eugène-Bataillon, 34060 Montpellier Cédex, France

S. Potenza
Università di Lecce, Corso di Laurea in Fisica, 73100 Lecce, Italy

(Received 24 October 1985; accepted for publication 11 June 1986)

A hierarchy of local nonlinear evolution equations associated with a new spectral problem is derived. It is shown that each equation is Hamiltonian and that their fluxes commute and a local infinite set of conserved densities is given. An interesting reduction is considered. In this case a hierarchy of local nonlinear evolution equations is generated by a recursion operator and its explicit inverse. Also this hierarchy satisfies a canonical geometrical scheme. It contains as a special case the Pohlmeyer–Lund–Regge equation.

I. INTRODUCTION

Among the nonlinear evolution equations (NLEE's) solved by the method of the inverse spectral transform (IST),[1] there are two relativistic covariant equations, which seem to have a great interest in many physical problems: the well-known sine–Gordon (SG) equation[2] and the Pohlmeyer–Lund–Regge (PLR) equation.[3,4,5]

The PLR equation, which is a generalization of the SG equation and shows soliton soutions,[4,5] arises in a relativistic theory of vortex motion in a superfluid (like He II)[4] and in the related theory of dual strings interacting through a scalar field.[6] It seems also to be very interesting in general relativity[7] and it appears in the one-space-dimensional version of the O(4) nonlinear σ model.[3,8]

In general, the NLEE's appear as members of an infinite hierarchy of partial differential equations generated by an integrodifferential operator, the so-called recursion operator, which is associated with an eigenvalue problem.

In this paper we have succeeded in finding the hierarchy containing the PLR equation in the laboratory coordinates.

Following the AKNS method,[9] we introduce the principal spectral problem

$$\Psi_x = U\Psi, \quad U = U(x,t;\lambda), \tag{1.1}$$

and the auxiliary spectral problem

$$\Psi_t = V\Psi, \quad V = V(x,t;\lambda). \tag{1.2}$$

The compatibility condition $\Psi_{xt} = \Psi_{tx}$ gives the so-called Lax representation

$$U_t - V_x + [U,V] = 0 \tag{1.3}$$

for the NLEE's that we are interested in. In general U and V are matrix operators, and Ψ is a fundamental matrix solution of the spectral problems. One can obtain a hierarchy of equations related to the potential operator U, choosing different V's.

Boiti and Tu[10] proposed an interesting new spectral problem

$$U = i\lambda\sigma_3 + u(x,t)\sigma_1 + (i/\lambda)(s(x,t)\sigma_3 + iv(x,t)\sigma_2), \tag{1.4}$$

where the σ_i's are the 2×2 Pauli matrices and $u(x,t), s(x,t)$, and $v(x,t)$ are the three independent scalar potentials, which are supposed to go to zero as $|x| \to \infty$. The auxiliary spectral operator V was chosen to be a polynomial of positive powers of λ. They found an associated new hierarchy of NLEE's and showed the canonical structure of this scheme.

The most interesting special case of this spectral problem is obtained by taking the reduction (Boiti–Leon–Pempinelli)[11]

$$s^2 - v^2 = s_0^2, \quad s_{0x} = s_{0t} = 0, \tag{1.5}$$

with the following asymptotic behaviors:

$$\begin{aligned} &u(x,t)\to 0,\\ &v(x,t)\to 0, \quad \text{as } |x|\to\infty.\\ &s(x,t)\to s_0, \end{aligned} \tag{1.6}$$

Also in this case there is an associated hierarchy of NLEE's; moreover the condition (1.5) allows us to find a hierarchy also for V chosen to be a polynomial of negative powers of λ. The recurrence operator in this case is the inverse of the recurrence operator of the hierarchy obtained by choosing a polynomial of positive powers of λ. The canonical structure of this scheme is proved by using the method recently proposed by Boiti–Pempinelli–Tu (BPT).[12] Therefore this spectral problem shows the new interesting property that both the recurrence operator and its associated explicit inverse generate two hierarchies of NLEE's, which are both local, in spite of the nonlocal character of these operators.

The most interesting NLEE associated to this spectral problem comes out to be the sine–Gordon equation in the laboratory coordinates.

In the paper[10] already quoted, Boiti and Tu proposed also a more general spectral problem

$$U(x,t;\lambda) = -i\lambda\sigma_3 + P(x,t) + (i/\lambda)Q(x,t), \tag{1.7}$$

where $P(x,t)$ is an off-diagonal 2×2 matrix and $Q(x,t)$ is a free 2×2 matrix.

[a)] On leave of absence from Dipartimento di Fisica dell' Università, 73100 Lecce, Italy.

[b)] Unit associated with C.N.R.S. N° 040768, cooperative research under Programme N° 080264.

The aim of this paper is to study this spectral problem and a reduced case that furnishes a hierarchy containing the PLR equation in the laboratory coordinates. In both nonreduced and reduced cases the canonical structure is explicitly found using the BPT method and infinitely many conservation laws are derived.

The reduced case shows the nice property that the recurrence operator and its explicit inverse generate two hierarchies of NLEE's, which are both purely differential.

Let us finally note that the PLR equation in the light-cone coordinates has been related to the Zakharov–Shabat–AKNS spectral problem.[13] However, the NLEE's in the hierarchy that can be obtained in this case couple two independent fields and cannot be related to the NLEE's obtained in this paper that couple four independent fields.

II. THE GENERAL CASE

Let us consider the spectral problem[10]

$$U(x,t;\lambda) = -i\lambda\sigma_3 + P(x,t) + i\lambda^{-1}Q(x,t), \tag{2.1}$$

where

$$P=\begin{pmatrix} 0 & q_1 \\ q_2 & 0\end{pmatrix},\quad Q=\begin{pmatrix} q_3 & q_4 \\ q_5 & q_6\end{pmatrix}, \tag{2.2}$$

with the asymptotic behaviors

$$q_i(x,t) \underset{|x|\to\infty}{\to} 0,\quad i=1,\ldots,6. \tag{2.3}$$

As usual, we can choose

$$V=\sum_{j=0}^{n} V_j(x,t)\lambda^{n-j}. \tag{2.4}$$

One can see immediately that both Q and V must be traceless and we put

$$P=\begin{pmatrix} 0 & q_1 \\ q_2 & 0\end{pmatrix},\quad Q=\begin{pmatrix} q_3 & q_4 \\ q_5 & -q_3\end{pmatrix}. \tag{2.5}$$

U takes the form

$$\begin{aligned} U=&(-i\lambda+i\lambda^{-1}q_3)\sigma_3+(q_1+i\lambda^{-1}q_4)\sigma_+ \\ &+(q_2+i\lambda^{-1}q_5)\sigma_-. \end{aligned}\tag{2.6}$$

We can decompose $V_j(x,t)$ as

$$V_j(x,t)=\tfrac{1}{2}(d_j\sigma_3+e_j\sigma_+ + f_j\sigma_-). \tag{2.7}$$

Inserting the expressions (2.6), (2.4), and (2.7) in the Lax representation (1.3), with a convenient choice of the integration constants, we obtain the recurrence relations (introducing for convenience e_{n+1} and f_{n+1})

$$\begin{aligned} &e_0=f_0=0,\quad d_0=-2i,\quad e_1=2q_1,\quad f_1=2q_2,\\ &ie_{j+1}=-\tfrac{1}{2}e_{j,x}-i(q_4d_{j-1}-q_3e_{j-1})-q_1d_j,\\ &if_{j+1}=\tfrac{1}{2}f_{j,x}+i(q_3f_{j-1}-q_5d_{j-1})-q_2d_j,\\ &d_j=I\,[(q_1f_j-q_2e_j)+i(q_4f_{j-1}-q_5e_{j-1})],\\ &\qquad j=1,\ldots,n, \end{aligned}\tag{2.8}$$

where I is the operator defined by

$$I\equiv\frac{1}{2}\left(\int_{-\infty}^{x}-\int_{x}^{\infty}\right)dx, \tag{2.9}$$

and the evolution equations

$$\begin{aligned} q_{1t}&=-ie_{n+1},\\ q_{2t}&=if_{n+1},\\ q_{3t}&=\tfrac{1}{2}(q_5e_n-q_4f_n),\\ q_{4t}&=q_4d_n-q_3e_n,\\ q_{5t}&=q_3f_n-q_5d_n. \end{aligned}\tag{2.10}$$

These can be written in a more convenient form by introducing the vectors

$$q=\begin{pmatrix} q_1\\ q_2\\ q_3\\ q_4\\ q_5\end{pmatrix}, \tag{2.11}$$

$$f(q)=\begin{pmatrix} q_2\\ q_1\\ 2\\ 0\\ 0\end{pmatrix}, \tag{2.12}$$

and

$$\Phi_j=\begin{pmatrix} -ie_j\\ if_j\\ \tfrac{1}{2}(q_5e_{j-1}-q_4f_{j-1})\\ q_4d_{j-1}-q_3e_{j-1}\\ q_3f_{j-1}-q_5d_{j-1}\end{pmatrix},\quad j=1,\ldots,n+1, \tag{2.13}$$

together with the matrix operators

$$L=\begin{pmatrix} -(i/2)D+iq_2Iq_1 & -iq_2Iq_2 & (i/2)q_5 & -iq_3+iq_2Iq_4 & -iq_2Iq_5\\ iq_1Iq_1 & (i/2)D-iq_1Iq_2 & (i/2)q_4 & iq_1Iq_4 & -iq_3-iq_1Iq_5\\ 2iIq_1 & -2iIq_2 & 0 & 2iIq_4 & -2iIq_5\\ i & 0 & 0 & 0 & 0\\ 0 & i & 0 & 0 & 0\end{pmatrix} \tag{2.14}$$

and

$$J=\begin{pmatrix} 0 & -2i & 0 & 0 & 0\\ 2i & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & iq_4 & -iq_5\\ 0 & 0 & -iq_4 & 0 & 2iq_3\\ 0 & 0 & iq_5 & -2iq_3 & 0\end{pmatrix}, \tag{2.15}$$

where $D \equiv \partial/\partial x$.

The recursion relations (2.8) have become

$$\Phi_{j+1} = L^\dagger \Phi_j, \quad j = 1,\ldots,n \tag{2.16}$$

[† means adjoint with respect to the bilinear form (2.18)] and the evolution equations have become

$$q_t = JL^n f(q), \quad n = 0,1,2,\ldots . \tag{2.17}$$

We briefly remember[12,14] that $q(x,t)$ can be regarded as a point in the configuration space $\mathscr{M}$ of vector-valued functions of the real variable x. Associated with each point q of $\mathscr{M}$ there is a tangent space T_q and a cotangent space $T_q^\dagger$ dual to T_q with respect to the symmetric bilinear form

$$\langle \beta,\alpha \rangle = \int_{-\infty}^{+\infty} \sum_j \beta_j(x)\alpha_j(x)dx, \tag{2.18}$$

where $\alpha \in T_q$ and $\beta \in T_q^\dagger$. We suppose $\alpha(x)$, $\beta(x) \to 0$ as $|x| \to \infty$ sufficiently rapidly for the convergence of any integral that may be required.

With respect to this bilinear form (2.18), the operator J is cosymplectic, i.e., it is antisymmetric

$$\langle \beta, J\alpha \rangle = -\langle J\beta,\alpha \rangle \tag{2.19}$$

and satisfies the Jacobi identity

$$\langle \alpha, J'[J\beta]\gamma \rangle + \langle \beta, J'[J\gamma]\alpha \rangle + \langle \gamma, J'[J\alpha]\beta \rangle = 0, \tag{2.20}$$

where $J'[\ ]$ is the Gâteaux (or directional) derivative of J, defined by

$$J'[\rho] = \frac{d}{d\epsilon} J(q+\epsilon\rho)\bigg|_{\epsilon=0}. \tag{2.21}$$

The fact that J is cosymplectic enables us to introduce a Poisson bracket for any pair of functionals F and G

$$\{F,G\} = \left\langle \frac{\delta F}{\delta q}, J\frac{\delta G}{\delta q} \right\rangle \tag{2.22}$$

($\delta/\delta q$ is the variational derivative), which is skew symmetric and satisfies the Jacobi identity.

Moreover J and L satisfy the first coupling condition

$$JL = L^\dagger J. \tag{2.23}$$

It can be shown that $L^\dagger$ is both a hereditary and a strong symmetry for all the equations in the hierarchy (2.17).[15]

III. THE HAMILTONIAN STRUCTURE

We show now that each equation of the hierarchy (2.17) has an infinite set of polynomial conserved quantities.

Let us introduce the projective variable

$$Z = \psi_2/\psi_1, \tag{3.1}$$

where ψ_1 and ψ_2 are the elements of the eigenvector $\Psi = \binom{\psi_1}{\psi_2}$ of the spectral equation (1.1). Here Z satisfies the Riccati equation

$$Z_x = (q_2 + i\lambda^{-1}q_5) + 2(i\lambda - i\lambda^{-1}q_3)Z - (q_1 + i\lambda^{-1}q_4)Z^2 \equiv C - 2AZ - BZ^2. \tag{3.2}$$

One can easily see that the quantity

$$\mathscr{H} = A + BZ \tag{3.3}$$

satisfies the relation

$$\mathscr{H}_t = (A+BZ)_x \tag{3.4}$$

and consequently is a conserved density with flux $A + BZ$.

Expanding Z formally as

$$Z = \sum_{k=0}^{\infty} Z_k \lambda^{-k}, \tag{3.5}$$

we get the recurrence relations

$$\begin{aligned} &Z_0 = 0,\\ &Z_1 = (i/2)q_2\\ &\vdots\\ &Z_{k,x} = iq_5\delta_{k1} + 2iZ_{k+1} - 2iq_3 Z_{k-1} - q_1 \\ &\qquad \times \sum_{j=0}^{k} Z_{k-j}Z_j - iq_4 \sum_{j=0}^{k-1} Z_{k-j-1}Z_j \\ &\qquad (k = 1,2,\ldots). \end{aligned} \tag{3.6}$$

They furnish the coefficients in the expansion

$$\mathscr{H} = -i\lambda + \sum_{k=0}^{\infty} \mathscr{H}_k \lambda^{-k} \tag{3.7}$$

and therefore an infinite set of local conserved densities

$$\mathscr{H}_k = iq_3\delta_{k1} + q_1 Z_k + iq_4 Z_{k-1}, \quad k = 0,1,2,\ldots, \tag{3.8}$$

where for convenience we have put $Z_{-1} = 0$.

We write the first few explicitly

$$\begin{aligned} &\mathscr{H}_0 = 0,\\ &\mathscr{H}_1 = i(q_3 + q_1q_2/2),\\ &\mathscr{H}_2 = \tfrac{1}{4}q_1q_{2x} - \tfrac{1}{2}q_1q_5 - \tfrac{1}{2}q_4q_2. \end{aligned} \tag{3.9}$$

The conserved quantities associated with $\mathscr{H}$ and $\mathscr{H}_k$ are

$$H = \int_{-\infty}^{+\infty} (\mathscr{H}(x) - \mathscr{H}(\infty))dx = \int_{-\infty}^{+\infty} (\mathscr{H}(x) + i\lambda)dx \tag{3.10}$$

and

$$H_k = \int_{-\infty}^{+\infty} \mathscr{H}_k(x)dx, \tag{3.11}$$

respectively, where

$$H = \sum_{k=0}^{\infty} H_k \lambda^{-k}. \tag{3.12}$$

Let us now follow the BPT method[12] and examine the functional derivative of these quantities.

If we define a matrix K such that

$$K_{ij} = \frac{\delta H}{\delta U_{ji}}, \tag{3.13}$$

we find that K satisfies

$$K_x = [U,K] \tag{3.14}$$

and

$$\mathrm{Tr}\{K\} = 1. \tag{3.15}$$

Let us consider the usual Weyl–Cartan basis R

$$\{R_1 = 1, R_2 = \sigma_3, R_3 = \sigma_+, R_4 = \sigma_-\}. \tag{3.16}$$

The operator U takes the form

$$U = \zeta_2\sigma_3 + \zeta_3\sigma_+ + \zeta_4\sigma_-, \tag{3.17}$$

where the components ζ_α ($\alpha = 2,3,4$) are furnished directly by (2.6).

We can decompose K with respect to the dual basis S

$$\{S_1 = \tfrac{1}{2}\mathbb{1}, S_2 = \tfrac{1}{2}\sigma_3, S_3 = \sigma_+, S_4 = \sigma_-\}, \tag{3.18}$$

$$K = \tfrac{1}{2}K_1\mathbb{1} + \tfrac{1}{2}K_2\sigma_3 + K_3\sigma_+ + K_4\sigma_-. \tag{3.19}$$

Condition (3.15) furnishes us

$$1 = \mathrm{Tr}\{K\} = K_1 \tag{3.20}$$

and Eq. (3.14) gives us

$$\begin{aligned} K_{1x} &= 0, \\ K_{2x} &= 2(K_3\zeta_3 - K_4\zeta_4), \\ K_{3x} &= K_2\zeta_4 - 2K_3\zeta_2, \\ K_{4x} &= 2K_4\zeta_2 - K_2\zeta_3. \end{aligned} \tag{3.21}$$

These differential equations can be used to write the BPT isospectral eigenvalue equation[12]

$$L\frac{\delta H}{\delta q} = \mu(\lambda)\frac{\delta H}{\delta q} + \nu(\lambda)\gamma(q), \tag{3.22}$$

where $\gamma(q)$ is a conserved covariant for all the equations of the hierarchy and $\mu(\lambda)$ and $\nu(\lambda)$ are functions of the spectral parameter λ to be determined.

In this case $\delta H/\delta q$ becomes

$$\frac{\delta H}{\delta q} = \begin{pmatrix} K_3 \\ K_4 \\ i\lambda^{-1}K_2 \\ i\lambda^{-1}K_3 \\ i\lambda^{-1}K_4 \end{pmatrix}. \tag{3.23}$$

Applying L to this quantity and using (3.20), (3.21), and the asymptotic behavior of $\delta H/\delta q$,

$$\frac{\delta H}{\delta q} \to \begin{pmatrix} 0 \\ 0 \\ i\lambda^{-1} \\ 0 \\ 0 \end{pmatrix} \quad \text{as } |x| \to \infty, \tag{3.24}$$

we obtain the explicit form of the isospectral equation (3.22)

$$L\frac{\delta H}{\delta q} = \lambda\frac{\delta H}{\delta q} + \frac{1}{2i}f(q), \tag{3.25}$$

where $f(q)$ is given by (2.12).

Inserting the asymptotic expansion (3.12) in (3.25) and equating to zero the coefficients of the successive powers of λ^{-1}, we get the following recursion relations:

$$\begin{aligned} &\frac{\delta H_1}{\delta q} = -1/2i f(q), \\ &\vdots \\ &\frac{\delta H_{k+1}}{\delta q} = L\frac{\delta H_k}{\delta q}, \quad k = 1,2,..., \end{aligned} \tag{3.26}$$

which can be rewritten as

$$L^k f(q) = -2i\frac{\delta H_{k+1}}{\delta q}, \quad k = 0,1,2,... . \tag{3.27}$$

Equations (3.27) prove the Hamiltonian character of the NLEE's (2.17). Therefore the general IST-solvable NLEE related to the spectral operator U (2.6) can be written as

$$\begin{aligned} q_t &= \sum_{j=0}^{n} \mu_j(t) J L^j f(q) \\ &= -2i\sum_{j=0}^{n} \mu_j(t) J \frac{\delta H_{j+1}}{\delta q}, \end{aligned} \tag{3.28}$$

where the coefficients $\mu_j(t)$ are arbitrary functions of t.

Let us note that from (3.25) we can easily prove, following the BPT method,[12] that the fluxes commute and the NLEE's (3.28) can be identified with the group of motion of a special geometrical structure called a symplectic Kähler manifold.[14]

Let us finally write explicitly the first equation in the hierarchy (3.28) ($n = 1$)

$$\begin{aligned} q_{1t} &= q_{1x} + 2q_4, \\ q_{2t} &= q_{2x} - 2q_5, \\ q_{3t} &= q_1q_5 - q_2q_4, \\ q_{4t} &= -2q_3q_1, \\ q_{5t} &= 2q_3q_2. \end{aligned} \tag{3.29}$$

We can still note that all the results of Boiti and Tu[10] are recovered in the limit

$$q_1, q_2 \to u, \quad q_3 \to s, \quad q_4 \to v, \quad q_5 \to -v. \tag{3.30}$$

IV. THE REDUCED CASE

We can now go on to examine an interesting reduced case. The first problem is the existence of a compatible reduction condition of the same kind as (1.5).[11]

We introduce an expansion for the traceless operator V in both negative and positive powers of λ:

$$V = \sum_{j=0}^{n} V_j\lambda^{n-j} + \sum_{j=0}^{p} W_j\lambda^{j-p-1}. \tag{4.1}$$

We can make two alternative choices of V:

$$\begin{aligned} &W_j(x,t) = \tfrac{1}{2}(a_j\sigma_3 + b_j\sigma_+ + c_j\sigma_-), \\ &V_j(x,t) \equiv 0 \end{aligned} \tag{4.2a}$$

or

$$\begin{aligned} &W_j(x,t) \equiv 0, \\ &V_j(x,t) = \tfrac{1}{2}(d_j\sigma_3 + e_j\sigma_+ + f_j\sigma_-). \end{aligned} \tag{4.2b}$$

One can see immediately that, for the second choice (4.2b), Eqs. (2.10) satisfy the condition

$$(q_3^2 + q_4q_5)_t = 0. \tag{4.3}$$

So we can consider the reduction

$$q_3^2 + q_4q_5 = \gamma^2, \quad \gamma_x = \gamma_t = 0, \tag{4.4}$$

with the following asymptotic behaviors for the fields:

$$\begin{aligned} q_1, q_2, q_4, q_5 &\to 0 \\ q_3 &\to \gamma \end{aligned} \quad \text{as } |x| \to \infty. \tag{4.5}$$

In the limit (3.30) we recover the condition (1.5).

When (4.4) is imposed, the recurrence relations (2.16) take the form (with a convenient choice of the constants of integration)

$$\hat{\Phi}_{j+1} = \hat{L}^{\dagger}\hat{\Phi}_j, \quad j=1,\ldots,n, \qquad e_0 = f_0 = 0, \quad d_0 = -2i, \tag{4.6}$$

where $\hat{\Phi}_j$ is the vector

$$\hat{\Phi}_j = \begin{pmatrix} -ie_j \\ if_j \\ q_4 d_{j-1} - q_3 e_{j-1} \\ q_3 f_{j-1} - q_5 d_{j-1} \end{pmatrix}, \quad j=1,\ldots,n+1, \tag{4.7}$$

and $\hat{L}^{\dagger}$ is the adjoint of the recurrence operator $\hat{L}$

$$\hat{L} = \begin{pmatrix} -\frac{1}{2}D + iq_2Iq_1 & -iq_2Iq_2 & -iq_3 + iq_2Iq_4 & -iq_2Iq_5 \\ iq_1Iq_1 & (i/2)D - iq_1Iq_2 & iq_1Iq_4 & -iq_3 - iq_1Iq_5 \\ i-(iq_5/q_3)Iq_1 & i(q_5/q_3)Iq_2 & -i(q_5/q_3)Iq_4 & (iq_5/q_3)Iq_5 \\ -i(q_4/q_3)Iq_1 & i+(iq_4/q_3)Iq_2 & -(iq_4/q_3)Iq_4 & (iq_4/q_3)Iq_5 \end{pmatrix}. \tag{4.8}$$

Let us introduce a new vector $\hat{q}$

$$\hat{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_4 \\ q_5 \end{pmatrix}. \tag{4.9}$$

The evolution equations now take the standard form in the case (4.2b)

$$\hat{q}_t = \hat{J}\hat{L}^{n} g(\hat{q}), \quad n=0,1,\ldots, \tag{4.10}$$

where

$$g(\hat{q}) = \begin{pmatrix} q_2 \\ q_1 \\ -q_5/q_3 \\ -q_4/q_3 \end{pmatrix}, \tag{4.11}$$

$$\hat{J} = \begin{pmatrix} 0 & -2i & 0 & 0 \\ 2i & 0 & 0 & 0 \\ 0 & 0 & 0 & 2iq_3 \\ 0 & 0 & -2iq_3 & 0 \end{pmatrix}. \tag{4.12}$$

Also in this case $\hat{J}$ is a cosymplectic operator, $\hat{L}$ is both a hereditary symmetry and a strong symmetry for all the equations in the hierarchy (4.10), and the first coupling condition

$$\hat{J}\hat{L} = \hat{L}^{\dagger}\hat{J} \tag{4.13}$$

is satisfied.

Let us now consider the first choice (4.2a). If we put expansion (4.1) with the coefficients given by (4.2a) in the Lax representation and we equate to zero the coefficients of the powers of λ, we can see by inspection that the recurrence relations can be solved only in the case when the reduction condition (4.4) holds and $\gamma \neq 0$. In this case we have (with a convenient choice of the constants of integration and introducing for convenience a_{p+1}, b_{p+1}, c_{p+1}, and $b_{-1} = c_{-1} = 0$)

$$\tilde{\Phi}_j = M^{\dagger}\tilde{\Phi}_{j-1}, \quad j=0,1,\ldots,p, \qquad a_0 = -2iq_3, \quad b_0 = -2iq_4, \quad c_0 = -2iq_5, \tag{4.14}$$

where $\tilde{\Phi}_j$ is the vector

$$\tilde{\Phi}_j = \begin{pmatrix} ib_j \\ -ic_j \\ q_3 b_{j+1} - q_4 a_{j+1} \\ q_5 a_{j+1} - q_3 c_{j+1} \end{pmatrix} \tag{4.15}$$

and $M^{\dagger}$ is the adjoint of the recurrence operator

$$M = -\frac{i}{2\gamma^2} \begin{pmatrix} 2q_5Iq_4 & -2q_5Iq_5 & -q_5w_1q_3 - 2\gamma^2 & q_5w_2q_3 \\ 2q_4Iq_4 & -2q_4Iq_5 & -q_4w_1q_3 & q_4w_2q_3 - 2\gamma^2 \\ 2\gamma^2/q_3 - w_2^{\dagger}q_4 & w_2^{\dagger}q_5 & \frac{1}{2}w_2^{\dagger}Dw_1q_3 + (\gamma^2/q_3)D & -\frac{1}{2}w_2^{\dagger}Dw_2q_3 \\ -w_1^{\dagger}q_4 & 2\gamma^2/q_3 + w_1^{\dagger}q_5 & \frac{1}{2}w_1^{\dagger}Dw_1q_1 & -\frac{1}{2}w_1^{\dagger}Dw_2q_3 - (\gamma^2/q_3)D \end{pmatrix}, \tag{4.16}$$

with

$$w_1 \equiv -q_4/q_3 + I(q_{4x}/q_3 + 2q_1) \tag{4.17}$$

and

$$w_2 \equiv q_5/q_3 + I(-q_{5x}/q_3 + 2q_2). \tag{4.18}$$

The evolution equations look like

$$\hat{q}_t = \tilde{\Phi}_p \tag{4.19}$$

and they can be put in the standard form

$$\hat{q}_t = \gamma\hat{J}M^{p+1}g(\hat{q}), \quad p=0,1,2,\ldots, \tag{4.20}$$

where $\hat{J}$ and $g(\hat{q})$ are given by (4.12) and (4.11), respectively.

Furthermore $M^{\dagger}$ is both a hereditary and a strong symmetry for all the equations in the hierarchy (4.20) and the first coupling condition

$$\hat{J}M = M^{\dagger}\hat{J} \tag{4.21}$$

is satisfied.

One can check directly that

$$\hat{L}M = M\hat{L} = 1, \tag{4.22}$$

and consequently (4.20) can be rewritten as

$$\hat{q}_t = \gamma\hat{J}\hat{L}^{-p-1}g(\hat{q}), \quad p=0,1,\ldots. \tag{4.23}$$

Therefore the general IST-solvable NLEE related to the spectral operator U (2.6) with the reduction (4.4) ($\gamma\neq 0$) can be written as

$$\hat{q}_t = \sum_{j=0}^{n} \mu_j^{(+)}(t)\widehat{\mathscr{J}L}^{\,j} g(\hat{q}) + \sum_{j=0}^{p} \mu_j^{(-)}(t)\widehat{\mathscr{J}L}^{\,-j-1} g(\hat{q}), \tag{4.24}$$

where the coefficients $\mu_j^{(\pm)}(t)$ are arbitrary functions of t.

V. THE HAMILTONIAN STRUCTURE IN THE REDUCED CASE

Let us consider again the Riccati equation (3.2) and both the asymptotic expansion of Z as $\lambda\to\infty$ and as $\lambda\to 0$:

$$Z^{(+)} = \sum_{k=0}^{\infty} Z_k^{(+)}\lambda^{-k} \tag{5.1}$$

and

$$Z^{(-)} = \sum_{k=0}^{\infty} Z_k^{(-)}\lambda^{k}. \tag{5.2}$$

Inserting (5.1) in (3.2) we reobtain the recurrence relations (3.6)

$$\begin{aligned} &Z_0^{(+)} = 0,\\ &Z_1^{(+)} = (i/2)q_2,\\ &\vdots\\ &Z_{k,x}^{(+)} = iq_5\delta_{k1} + 2iZ_{k+1}^{(+)} - 2iq_3 Z_{k-1}^{(+)} - q_1 \\ &\qquad\times \sum_{j=0}^{k} Z_{k-j}^{(+)}Z_j^{(+)} - iq_4 \sum_{j=0}^{k-1} Z_{k-j-1}^{(+)}Z_j^{(+)} \\ &\qquad (k=1,2,\ldots), \end{aligned} \tag{5.3}$$

and inserting (5.2) in (3.2) we now obtain

$$\begin{aligned} &Z_0^{(-)} = q_5/(q_3+\gamma),\\ &Z_1^{(-)} = (i/2\gamma)(Z_{0,x}^{(-)} + q_1 Z_0^{(-)2} - q_2),\\ &\vdots\\ &Z_{k,x}^{(-)} = 2iZ_{k-1}^{(-)} - 2iq_3 Z_{k+1}^{(-)} - q_1 \sum_{j=0}^{k} Z_{k-j}^{(-)}Z_j^{(-)} \\ &\qquad - iq_4 \sum_{j=0}^{k} Z_{k-j+1}^{(-)}Z_j^{(-)} \\ &\qquad (k=1,2,\ldots). \end{aligned} \tag{5.4}$$

We can introduce

$$\mathscr{H}^{(+)} = -i\lambda + \sum_{k=0}^{\infty} \mathscr{H}_k^{(+)}\lambda^{-k} \tag{5.5}$$

and

$$\mathscr{H}^{(-)} = i\lambda^{-1}\gamma + \sum_{k=0}^{\infty} \mathscr{H}_k^{(-)}\lambda^{k}. \tag{5.6}$$

Using (3.3) and the recurrence relations (5.3) and (5.4) we get two infinite sets of conserved densities:

$$\mathscr{H}_k^{(+)} = iq_3\delta_{k1} + q_1 Z_k^{(+)} + iq_4 Z_{k-1}^{(+)}, \quad k=1,2,\ldots, \tag{5.7}$$

and

$$\mathscr{H}_k^{(-)} = q_1 Z_k^{(-)} + iq_4 Z_{k+1}^{(-)} - i\delta_{k1}, \quad k=0,1,2,\ldots. \tag{5.8}$$

The conserved quantities associated with $\mathscr{H}^{(+)}$, $\mathscr{H}^{(-)}$, $\mathscr{H}_k^{(+)}$, and $\mathscr{H}_k^{(-)}$ are

$$H^{(\pm)} = \int_{-\infty}^{+\infty} dx(\mathscr{H}^{(\pm)}(x) - \mathscr{H}^{(\pm)}(\infty)) \tag{5.9}$$

and

$$H_k^{(\pm)} = \int_{-\infty}^{+\infty} dx(\mathscr{H}_k^{(\pm)}(x) - \mathscr{H}_k^{(\pm)}(\infty)), \tag{5.10}$$

respectively, where

$$H^{(+)} = \sum_{k=0}^{\infty} H_k^{(+)}\lambda^{-k} \tag{5.11}$$

and

$$H^{(-)} = \sum_{k=0}^{\infty} H_k^{(-)}\lambda^{k}. \tag{5.12}$$

Let us stress that we still have not used the reduction condition (4.4) and consequently the two sets $\{\mathscr{H}_k^{(\pm)}\}$ furnish the conserved densities for NLEE's related to the spectral operator U (2.6) in reduced and nonreduced cases.

In order to relate the conserved quantities $H_k^{(\pm)}$ to the Hamiltonians of NLEE's (4.24) we need to prove that both $\delta H^{(+)}/\delta\hat{q}$ and $\delta H^{(-)}/\delta\hat{q}$ satisfy a BPT equation (3.22). One can follow the same method used in Ref. 11 for the reduced case. Precisely we prove, in the nonreduced case, that $\delta H^{(-)}/\delta q = \delta H^{(+)}/\delta q$ and, because the equality keeps its validity in the reduced case, we write the eigenvalue equation (3.25) for the reduced operator $\hat{L}$ as

$$\hat{L}\frac{\delta H^{(\pm)}}{\delta\hat{q}} = \frac{\delta H^{(\pm)}}{\delta\hat{q}} + \frac{1}{2i}g(\hat{q}), \tag{5.13}$$

where

$$\frac{\delta}{\delta\hat{q}} = \begin{pmatrix} \dfrac{\delta}{\delta q_1} \\ \dfrac{\delta}{\delta q_2} \\ \dfrac{\delta}{\delta q_4} - \dfrac{q_5}{2q_3}\dfrac{\delta}{\delta q_3} \\ \dfrac{\delta}{\delta q_5} - \dfrac{q_4}{2q_3}\dfrac{\delta}{\delta q_3} \end{pmatrix}, \tag{5.14}$$

which is satisfied by both $\delta H^{(+)}/\delta\hat{q}$ and $\delta H^{(-)}/\delta\hat{q}$.

Inserting the asymptotic expansions (5.5) and (5.6) in (5.13) and equating to zero the coefficients of the successive powers of λ^{-1} and λ, we get the recursion relations

$$\begin{aligned} &\frac{\delta H_1^{(+)}}{\delta\hat{q}} = -\frac{1}{2i}g(\hat{q}),\\ &\frac{\delta H_{k+1}^{(+)}}{\delta\hat{q}} = \hat{L}\frac{\delta H_k^{(+)}}{\delta\hat{q}}, \quad k=1,2,\ldots, \end{aligned} \tag{5.15}$$

and

$$\begin{aligned} &\frac{\delta H_0^{(-)}}{\delta\hat{q}} = \frac{1}{2i}\hat{L}^{-1}g(\hat{q}),\\ &\frac{\delta H_{k+1}^{(-)}}{\delta\hat{q}} = \hat{L}^{-1}\frac{\delta H_k^{(-)}}{\delta\hat{q}}, \quad k=0,1,2,\ldots. \end{aligned} \tag{5.16}$$

Equations (5.15) and (5.16) can be rewritten as

$$\hat{L}^{k} g(\hat{q}) = -2i \frac{\delta H^{(+)}_{k+1}}{\delta \hat{q}}, \quad k = 0,1,2,..., \tag{5.17}$$

and

$$\hat{L}^{-k} g(\hat{q}) = 2i \frac{\delta H^{(-)}_{k-1}}{\delta \hat{q}}, \quad k = 1,2,... . \tag{5.18}$$

In this way we have proved the Hamiltonian character of the NLEE's (4.24).

Therefore the general IST-solvable NLEE related to the spectral operator U (2.6) with the condition (4.4) and $\gamma \neq 0$ can be written as

$$\begin{aligned}\hat{q}_t &= \sum_{j=0}^{n} \mu_j^{(+)}(t) \widehat{J} \widehat{L}^{j} g(\hat{q}) \\ &\quad + \sum_{j=0}^{p} \mu_j^{(-)}(t) \widehat{J} \widehat{L}^{-j-1} g(\hat{q}) \\ &= -2i \sum_{j=0}^{n} \mu_j^{(+)}(t) \hat{J} \frac{\delta H^{(+)}_{j+1}}{\delta \hat{q}} \\ &\quad + 2i \sum_{j=0}^{p} \mu_j^{(-)}(t) \hat{J} \frac{\delta H^{(-)}_{j}}{\delta \hat{q}}.\end{aligned} \tag{5.19}$$

Also in this case one can easily prove from (5.13) that the fluxes commute and the NLEE's (5.19) can be identified with the group of motion of a symplectic Kähler manifold.

Let us once more stress that, for this spectral problem with the reduction condition (4.4), together with the subcase considered by Boiti–Leon–Pempinelli[11] (and only in these two cases at our knowledge), we have the nice property that the recurrence operator $\hat{L}$ has an explicit inverse $\hat{L}^{-1}$, which is also a recurrence operator, and *both* the hierarchies of NLEE's associated with $\hat{L}$ and $\hat{L}^{-1}$ are purely local.

VI. THE POHLMEYER–LUND–REGGE EQUATION

Let us now write explicitly one interesting equation of the hierarchy (5.19). We choose $n = 1$, $p = 0$, and $\mu_0^{(+)} = 0$, $\mu_1^{(+)} = 1$, and $\mu_0^{(-)} = \gamma$ and we get

$$\begin{aligned} &q_{1t} = q_{1x} + 4q_4, \quad q_{2t} = q_{2x} - 4q_5, \\ &q_{4t} = -q_{4x} - 4q_3 q_1, \quad q_{5t} = -q_{5x} + 4q_3 q_2. \end{aligned} \tag{6.1}$$

Now we can introduce the two new functions $\omega(x,t)$ and $\chi(x,t)$ in the following way:

$$\begin{aligned} &q_3 = \gamma \cos \omega, \\ &q_4 = \gamma e^{i\chi} \sin \omega, \\ &q_5 = \gamma e^{-i\chi} \sin \omega, \end{aligned} \tag{6.2}$$

which satisfy the reduction condition (4.4)

$$q_3^2 + q_4 q_5 = \gamma^2, \quad \gamma_x = \gamma_t = 0, \quad \gamma \neq 0.$$

The first two equations (6.1) give

$$\begin{aligned} &q_{1t} - q_{1x} = 4\gamma e^{i\chi} \sin \omega, \\ &q_{2t} - q_{2x} = -4\gamma e^{-i\chi} \sin \omega, \end{aligned} \tag{6.3}$$

and the two last equations (6.1) take the form, after some algebraic manipulations,

$$\begin{aligned} &\omega_{tt} - \omega_{xx} + 16\gamma \sin \omega - \tan \omega (\chi_t^2 - \chi_x^2) = 0, \\ &\chi_{tt} - \chi_{xx} + 2 \cot \omega (\omega_t \chi_t - \omega_x \chi_x) \\ &\quad + \tan \omega (\chi_t + \chi_x)(\omega_t - \omega_x) = 0. \end{aligned} \tag{6.4}$$

In the limit $\chi \to$ const, we reobtain, as expected, the sine-Gordon equation in the laboratory coordinates.

Equations (6.4) can be easily transformed into the Pohlmeyer–Lund–Regge equation. In fact, if we introduce the new fields θ and λ as

$$\begin{aligned} &\omega = 2\theta + \pi, \\ &\chi_x = \lambda_x - \tan^2(\omega/2) \lambda_t, \\ &\chi_t = \lambda_t - \tan^2(\omega/2) \lambda_x, \end{aligned} \tag{6.5}$$

we get immediately

$$\begin{aligned} &\theta_{xx} - \theta_{tt} + 16\gamma \sin \theta \cos \theta \\ &\quad + (\lambda_x^2 - \lambda_t^2)(\cos \theta / \sin^3 \theta) = 0, \\ &(\cot^2 \theta \lambda_x)_x = (\cot^2 \theta \lambda_t)_t, \end{aligned} \tag{6.6}$$

which is just the equation given by Lund and Regge in their paper.[4]

Therefore we have succeeded in getting the PLR equation in the laboratory coordinates as a member of a hierarchy of purely local NLEE's that satisfy a canonical geometrical scheme.

The study of the Bäcklund transformations of this hierarchy and the case $\gamma = 0$ is deferred to a forthcoming paper.

ACKNOWLEDGMENTS

We are very grateful to Professor M. Boiti, who called our attention on this spectral problem, for many useful discussions and comments. Thanks are also due to Dr. C. Laddomada for some preliminary computations. One of us (F.P.) is grateful to Professor P. C. Sabatier and the Laboratoire de Physique Mathématique, where this work has been completed, for the kind hospitality.

This work was supported in part by MPI, Italy and INFN, Sezione de Bari, Italy.

[1]See, for instance, M. J. Ablowitz and J. Segur, "Solitons and the inverse scattering transform," in *Studies in Applied Mathematics* (SIAM, Philadelphia, 1981); F. Calogero and A. De Gasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982), Vol. 1; S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons—The Inverse Scattering Method* (Plenum, New York, 1984).

[2]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973); L. A. Takhtadzhyan and L. Faddeev, Theor. Mat. Phys. USSR **21**, 1046 (1974); D. J. Kaup, Stud. Appl. Math. **54**, 165 (1975).

[3]K. Pohlmeyer, Commun. Math. Phys. **46**, 207 (1976).

[4]F. Lund and T. Regge, Phys. Rev. D **14**, 1524 (1976).

[5]F. Lund, Phys. Rev. Lett. **38**, 1175 (1977).

[6]H. Nielsen and P. Olesen, Nucl. Phys. B **57**, 367 (1973); Y. Nambu, Phys. Rep. **23**, 250 (1976).

[7]J. L. Gervais and A. Neveu, Phys. Rep. **23**, 240 (1976); in *Quantum Gravity—an Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford U. P., Oxford, 1975).

[8]F. Lund, Phys. Rev. D **15**, 1540 (1977).

[9]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).

[10]M. Boiti and G. Z. Tu, Nuovo Cimento B **75**, 145 (1983).

[11]M. Boiti, J. J. P. Leon, and F. Pempinelli, J. Math. Phys. **25**, 1725 (1984).

[12]M. Boiti, F. Pempinelli, and G. Z. Tu, Nuovo Cimento B **79**, 231 (1984).

[13]M. Jimbo, Prog. Theor. Phys. **61**, 359 (1979).

[14]F. Magri, "A geometrical approach to the nonlinear solvable equations," in *Lecture Notes in Physics*, Vol. 120, edited by M. Boiti, F. Pempinelli, and G. Soliani (Springer, Berlin, 1980).

[15]A. S. Fokas and B. Fuchssteiner, Physica D **4**, 47 (1981); R. L. Anderson and A. S. Fokas, J. Math. Phys. **23**, 1066 (1982).

Some remarks on the nonlinear integral equation in Kirkpatrick's theory of glass transition

Tetz Yoshimura
Paris, France[a)]

(Received 18 February 1986; accepted for publication 28 July 1986)

The nonlinear singular integral equation for "self-energy" $\Sigma(k,z)$ arising in Kirkpatrick's mode-coupling theory of glass transition is analyzed without suppressing the k dependence. An equation that is equivalent to Kirkpatrick's equation and suitable for high density is set up. Applicability of Lika's generalization of the Newton–Kantorovich successive approximation is discussed. The possibility of solutions that cannot be found by iteration is pointed out.

I. INTRODUCTION

Recently, Kirkpatrick[1] proposed a mode-coupling theory of glass transition. The correlation function $\Phi(k,z)$ is defined as

$$\Phi(k,z) = \int_0^\infty dt\, e^{-zt}\{ns(k)\Omega(k)\}^{-1} \times\Big(\sum_j e^{-ikr_j(0)} \sum_l e^{ikr_l(t)}\Big), \tag{1}$$

and the relation between $\Phi(k,z)$ and the "self-energy" $\Sigma(k,z)$ is

$$\Phi(k,z) = [z + \Omega^2(k)\{\gamma(k) + \Sigma(k,z)\}^{-1}]^{-1}, \tag{2}$$

or equivalently

$$\Sigma(k,z) = \Omega^2(k)\{(\Phi(k,z))^{-1} - z\}^{-1} - \gamma(k). \tag{3}$$

In this paper, we investigate the mathematical structures of singular nonlinear integral equations arising in the Kirkpatrick model without simplifications. Earlier works by Leutheusser[2] and Bergtzelius *et al.*[3] also suppressed k dependence. We apply nonlinear operator theory to these integral equations, regarding Φ and Σ as complex-valued functions of real variables, not as complex analytic functions.

II. INTEGRAL EQUATIONS AND THEIR SOLVABILITY

The equation for the "self-energy" $\Sigma(k,z)$ is

$$\begin{aligned}\Sigma(k,z) = {} & \int_{c-i\infty}^{c+i\infty} \frac{dz_1}{2\pi i}\int_0^\infty dq \int_{-1}^{1} d\mu\, V(q,k,\mu) \\ & \times [z_1 + \Omega^2((q^2+kq\mu+\tfrac14 k^2)^{1/2})\{\gamma((q^2+kq\mu+\tfrac14 k^2)^{1/2}) + \Sigma((q^2+kq\mu+\tfrac14 k^2)^{1/2},z_1)\}^{-1}]^{-1} \\ & \times [z - z_1 + \Omega^2((q^2-kq\mu+\tfrac14 k^2)^{1/2})\{\gamma((q^2-kq\mu+\tfrac14 k^2)^{1/2}) + \Sigma((q^2-kq\mu+\tfrac14 k^2)^{1/2},z-z_1)\}^{-1}]^{-1},\end{aligned} \tag{4}$$

where

$$\begin{aligned} & \Omega^2(k) = k^2(\beta m S(k))^{-1}, \quad \beta = (k_B T)^{-1}, \\ & V(q,k,\mu) = \frac{nq^2 S(q_+)S(q_-)}{2(2\pi)^2\beta m}\Big[\Big(\frac12 k + q\mu\Big)c(q_+) + \Big(\frac12 k - q\mu\Big)c(q_-) + nkc^{(3)}(q,k,\mu)\Big]^2, \\ & q_\pm = (q^2 \pm kq\mu + \tfrac14 k^2)^{1/2}, \quad \gamma(k) = n\beta^{-1/2}m^{-1/2}f(k)g(n), \\ & f(k) \simeq O(k^\delta), \quad 1<\delta<2, \quad \text{for } k\to 0, \quad f(k)\simeq O(k^\epsilon), \quad \epsilon<0, \quad \text{for } k\to\infty, \\ & 0<g(0)<\infty, \quad 0<g(\infty)<\infty, \quad s(q) = 1 + nh(q). \end{aligned} \tag{5}$$

Here $h(q)$, $c(q)$, and $c^{(3)}(q,k,\mu)$ are Fourier transforms of the total correlation function, direct two-particle correlation function, and direct three-particle correlation functions, respectively. (For definitions of these functions, see, e.g., Cole[4] and Resibois and DeLeener.[5]) As the structure of Eq. (4) is different from those of the equations in quantum field theory, etc., it is a challenging problem to deal with this equation. First, let us consider the case of small n. The equation of Σ can be rewritten as follows, showing parameters n and β more explicitly:

$$\begin{aligned}\Sigma(k,z) &= \mathfrak{S}(\Sigma;k,z) \\ &=: \frac{n}{\beta}\int_{-i\infty+c}^{i\infty+c}\frac{dz_1}{2\pi i}\int_0^\infty dq \int_{-1}^{1} d\mu\, \tilde V(q,k,\mu)\{z_1 + \beta^{-1}\omega^2(q_+)(\Sigma(q_+,z_1) + n\beta^{-1/2}\tilde\gamma(q_+))^{-1}\}^{-1} \\ &\quad \times\{z - z_1 + \beta^{-1}\omega^2(q_-)(\Sigma(q_-,z-z_1) + n\beta^{-1/2}\tilde\gamma(q_-))^{-1}\}^{-1},\end{aligned} \tag{6}$$

where

[a)] Unemployed. Postal contact address: c/o Professor A. Vančura, FB Physik, Universität Kaiserslautern, D6750 Kaiserslautern, West Germany.

 0022-2488/86/122868-04$02.50

$$\widetilde{V}(q,k,\mu) = n^{-1}\beta V(q,k,\mu), \quad \omega(k) = \beta^{1/2}\Omega(k), \quad \tilde{\gamma}(k) = n^{-1}\beta^{1/2}\gamma(k), \tag{7}$$

and n plays the role of a sort of coupling constant.

The denominator

$$\{z_1 + \beta^{-1}\omega^2(q_+)(\Sigma(q_+,z_1) + n\beta^{-1/2}\tilde{\gamma}(q_+))^{-1}\}$$

vanishes for $z_1 = 0, \mu = \pm 1, q = \pm \frac{1}{2}k$. For $k \neq 0$, one can take the principal value. For $k = 0$, this denominator vanishes at the end point $q = 0$, but the integral converges because the kernel $\widetilde{V}$ contains the factor q^2. The situation is similar for $\{z - z_1 + \cdots\}$.

So, we replace $\int_{c-i\infty}^{c+i\infty}$ by $\int_{-i\infty}^{i\infty}$ and regard $\Sigma(k,z)$ as a complex-values function of real variables k and $\zeta = i^{-1}z$ and put in a Banach space. It can be easily seen that

$$\mathfrak{S}(0;k,z) = \frac{n}{\beta}\int dq \int d\mu\, \widetilde{V}(q,k,\mu)\left\{z + \frac{\beta^{-1/2}}{n}[\omega^2(q_+)(\tilde{\gamma}(q_+))^{-1} + \omega^2(q_-)(\tilde{\gamma}(q_-))^{-1}]\right\}^{-1} \tag{8}$$

converges if

$$c(q) = o(q^{-3/2-\epsilon/2}), \quad c^{(3)}(q,k,\mu) = o(q^{-1/2-\epsilon/2}), \quad \text{for } q \to \infty. \tag{9}$$

It is an interesting fact that the conditions (9) are not satisfied for hard spheres. Because of the poles of the integrand, the mapping $\mathfrak{S}$ is not Fréchet differentiable, so that conventional Newton–Kantorovich-type successive approximations are not applicable. But do not worry. We have the following theorems due to Lika.[6]

Theorem 1: Suppose the following conditions are satisfied: (1) the Fréchet derivative Q' of the Lipschitz approximation Q to P satisfies the condition

$$\|Q'(x_1) - Q'(x_2)\| < \kappa\|x_1 - x_2\|, \quad \forall x_1,x_2 \in M;$$

(2) $\|P(x_1) - Q(x_1) - P(x_2) + Q(x_2)\| < \lambda\, \|x_1 - x_2\|, \quad \forall x_1,x_2 \in M;$

(3) $H_0 = [I - Q'(x_0)]^{-1}$ exists and $\|H_0\| < B_0, \quad \|H_0(x_0 - P(x_0))\| < \xi_0;$

(4) $B_0\lambda < 1;$

(5) $h_0 = B_0\xi_0\kappa < \frac{1}{2}(1 - B_0\lambda)^2;$

(6) the closed ball $\bar{S}(x_0,r_0) \subseteq M$, where $r_0 = [1 - B_0\lambda - \{(1 - B_0\lambda)^2 - 2h_0\}^{1/2}]\xi_0 h_0^{-1}$.

Then the equation $x - P(x) = 0$ has a solution $x^* \in S(x_0,r_0)$, to which the sequence defined by

$$x_{n+1} = x_n - [I - Q'(x_n)]^{-1}(x_n - P(x_n))$$

converges, and the error estimate is

$$\|x^* - x_n\| < [1 - B_n\lambda - \{(1 - B_n\lambda)^2 - 2h_n\}^{1/2}](B_n\kappa)^{-1}.$$

Theorem 2: Suppose the conditions (1)–(4) of Theorem 1 are satisfied, as well as the condition (5′) there exists a number $N \in ((1 - B_0\lambda)^{-1}, 2(1 - B_0\lambda)^{-1})$ such that

$$2\{(1 - B_0\lambda)N - 1\}N^{-2} > h_0.$$

Then, the sequence defined by

$$x_{n+1} = x_n - [I - Q'(x_0)]^{-1}(x_n - P(x_n))$$

converges to the unique solution x^* in $S(x_0,N\xi_0)$ of the equation $x - P(x) = 0$ and the error estimate reads

$$\|x^* - x_n\| \leqslant (Nh_0 - B\lambda_0)\xi_0(1 - Nh_0 - B_0\lambda)^{-1}.$$

If $N > 2(1 - B_0\lambda)^{-1}$ but $h_0 \leqslant \frac{1}{2}(1 - B_0\lambda)^2$, a solution exists but uniqueness is not guaranteed.

As for a Lipschitz approximation to $\mathfrak{S}$, one can take

$$\begin{aligned}
&\mathfrak{T}(\Sigma;k,z)\\
&=: \frac{n}{\beta}\int_{-i\infty}^{i\infty}\frac{dz_1}{2\pi_2}\int_0^\infty dq\int_{-1}^{1}d\mu\,\widetilde{V}(q,k,\mu)[\{z_1 + \omega^2(q_+)(n\beta^{-1/2}\tilde{\gamma}(q_+) + \Sigma(q_+,z_1))^{-1}\}^{-1}\\
&\quad\times\theta_\rho(\{(q + \tfrac{1}{2}k\mu)^2 + (|\mu| - 1)^2\tau_1^2 + |z_1^2|\tau_2^2\}^{1/2} - \nu) + \{z_1 + \omega^2(q_+)(n\beta^{-1/2}\tilde{\gamma}(q_+) + \Sigma_{(0)}(q_+,z_1))^{-1}\}^{-1}\\
&\quad\times\{1 - \theta_\rho(\{(q + \tfrac{1}{2}k\mu)^2 + (|\mu| - 1)^2\tau_1^2 + |z_1^2|\tau_2^2\}^{1/2} - \nu)\}][\{z - z_1 + \omega^2(q_-)(n\beta^{-1/2}\tilde{\gamma}(q_-) + \Sigma(q_-,z - z_1))^{-1}\}^{-1}\\
&\quad\times\theta_\rho(\{(q - \tfrac{1}{2}k\mu)^2 + (|\mu| - 1)^2\tau_1^2 + |(z - z_1)^2|\tau_2^2\}^{-1/2} - \nu) + \{z - z_1 + \omega^2(q_-)(n\beta^{-1/2}\tilde{\gamma}(q_-)\\
&\quad+ \Sigma_{(0)}(q_-,z - z_1))^{-1}\}^{-1}\{1 - \theta_\rho(\{(q - \tfrac{1}{2}k\mu)^2 + (|\mu| - 1)^2\tau_1^2 + |(z - z_1)^2|\tau_2^2\}^{1/2} - \nu)\}],
\end{aligned} \tag{10}$$

$$\theta_\rho(x) = \int_{-\infty}^{x}\varphi_\rho(y)dy, \quad \varphi_\rho(y) = \begin{cases}0, & |y| \geqslant \rho,\\ \exp\{-y^2(\rho - y)^{-2}\}, & |y| < \rho,\end{cases} \tag{11}$$

and choose parameters ν,ρ,τ_1, and τ_2 and $\Sigma_{(0)}$ so as to satisfy the conditions of Theorem 1 or 2.

This technique can be applied also when it happens that

$$\mathrm{Re}\,\Sigma(k,i\zeta) + n\beta^{-1/2}\tilde{\gamma}(k) = 0, \quad \omega^2(k)\left[\mathrm{Im}\,\Sigma(k,i\zeta)\{(\mathrm{Re}\,\Sigma(k,i\zeta) + n\beta^{-1/2}\tilde{\gamma}(k))^2 + (\mathrm{Im}\,\Sigma(k,i\zeta))^2\}^{-1}\right] = \beta\zeta, \tag{12}$$

for some combinations of k, $\zeta\in\mathbb{R}$ after certain steps of successive approximation.

Alternatively, one can write an integral equation for $\Xi(k,z) = :z\Phi(k,z)$:

$$\Xi(k,z) = z\Big\{z + n\omega^2(k)\Big[\beta^{1/2}\tilde{\gamma}(k) + \int_{-i\infty}^{i\infty}\frac{dz_1}{2\pi i}\int_0^\infty dq \times\int_{-1}^{1} d\mu\, \tilde{V}(q,k,\mu)\Xi(q_+,z_1)\Xi(q_-,z-z_1)z_1^{-1}(z-z_1)^{-1}\Big]^{-1}\Big\}^{-1} = :\mathfrak{X}(\Xi;k,z). \tag{13}$$

(Φ is not normable.)

Obviously $\Xi = 1$ if $n = 0$. So, one might think that one could begin with the zeroth approximation $\Xi_o = 1$, but it does not work unless

$$c(q) = o(q^{-5/2}), \quad c^{(3)}(q,k,\mu) = o(q^{-3/2}), \quad \text{for } q\to\infty, \tag{14}$$

because the integral $\int_0^\infty dq\, V(q,k,\mu)$ diverges otherwise. Moreover $\Xi = 1$ gives $\Sigma = \infty$. Therefore, Ξ, and consequently Φ, cannot be expanded in powers of n, unless condition (14) is satisfied.

Now we define the norm of Ξ by

$$\|\Xi\| = c_1 \sup_{k\in\mathbb{R}^+,\ \zeta\in\mathbb{R}} |\Xi(k,i\zeta)| + c_2 \sup_{k\in\mathbb{R}^+,\ \zeta\in\mathbb{R}} |k^2\Xi(k,i\zeta)|, \quad c_1,c_2\in\mathbb{R}^+. \tag{15}$$

Then the map $\mathfrak{X}$ is Fréchet differentiable and one can begin with the zeroth approximation

$$\Xi_0(k,z) = z\{z + n\beta^{-1/2}\omega^2(k)(\tilde{\gamma}(k))^{-1}\}^{-1}.$$

However, the situation (12) cannot be dealt with by Eq. (13), because the norm of corresponding Ξ is not bounded. On the other hand, by virtue of the Fréchet differentiability, bifurcation theory is applicable. In order to find bifurcation, one has, at first, to solve Eq. (13) and find a solution $\Xi(k,z',n,\beta)$ as a function of k,z,n, and β. Then a necessary condition for bifurcation becomes a nonlinear (with respect to n and β) eigenvalue problem

$$\begin{aligned}\Xi^{\blacktriangledown}(k,z) = {}& 2zn\omega^2(k)\Big\{\beta^{1/2}\tilde{\gamma}(k) + \int_{-i\infty}^{i\infty} dz_1\int_0^\infty dq\int_{-1}^{1} d\mu\, \tilde{V}(q,k,\mu)\Xi(q_+,z_1;\beta)\Xi(q_-,z-z_1;n,\beta)z_1^{-1}(z-z_1)^{-1}\Big\}^{-2}\\ &\times\Big[z + n\omega^2(k)\Big\{\int_{-i\infty}^{i\infty} dz_1\int_0^\infty dq\int_{-1}^{1} du\, \tilde{V}(q,k,\mu)\Xi(q_+,z_1;n,\beta)\Xi(q_-,z-z_1;n,\beta)z_1(z-z_1)^{-1}\Big\}^{-1}\Big]^{-2}\\ &\times\int_{-i\infty}^{i\infty} dz_1\int_{-1}^{1} d\mu\, \tilde{V}(q,k,\mu)\Xi(q_+,z_1;n,\beta)z_1^{-1}(z-z_1)^{-1}\Xi^{\blacktriangledown}(q_-,z-z_1).\end{aligned} \tag{16}$$

If Eq. (16) admits a nonzero solution for some combination of values of n and β, say (n_c,β_c), there may be a bifurcation. However, in order to know whether a bifurcation actually occurs, one has to investigate the equation

$$\mathfrak{X}(\Xi(\cdot,\cdot;n,\beta);k,z) - \mathfrak{X}(\Xi(\cdot,\cdot;n,\beta) + \Xi^{\blacktriangle}(\cdot,\cdot);k,z) = \Xi^{\blacktriangle}(k,z), \tag{17}$$

for values of n and β near (n_c,β_c). It should be noticed that bifurcation and the glass transition are different. However, neither Eq. (4) nor Eq. (13) is suitable for a search for critical values of n and β, because the norms of Σ and Ξ are not bounded at a critical point.

Let us assume asymptotic behaviors of $\gamma(k)$,

$$\gamma(k\to 0) \simeq O(k^\delta), \quad 1\leqslant\delta\leqslant 2, \quad \gamma(k\to\infty) \simeq O(k^{-\epsilon}), \quad \epsilon\geqslant 0,$$

and define a new unknown function Λ by

$$\Lambda(k,z) = k^\delta(k^2+\lambda)^{-\epsilon'/2}(\gamma(k) + \Sigma(k,z))^{-1}, \quad \lambda\in\mathbb{R}^+, \qquad \epsilon' = \mathrm{Max}(\delta,\epsilon). \tag{18}$$

Then the equation for Λ is

$$\begin{aligned}\Lambda(k,z) = {}& \frac{k^\delta}{n^3\beta^{1/2}(k^2+\lambda)^{\epsilon'/2}}\Big[n^{-2}\tilde{\gamma}(k) + \beta^{-1/2}\int_{-i\infty}^{i\infty}\frac{dz_1}{2\pi i}\int_0^\infty dq\int_{-1}^{1} d\mu\\ &\times V^{\blacktriangle}(q,k,\mu)\{z_1 + \beta^{-1}\omega^2(q_+)\Lambda(q_+,z_1)(q_+^2+\lambda)^{\epsilon'/2}q_+^{-\delta}\}^{-1}\\ &\times\{z - z_1 + \beta^{-1}\omega^2(q_-)\Lambda(q_-,z-z_1)(q_-^2+\lambda)^{\epsilon'/2}q_-^{-\delta}\}^{-1}\Big]^{-1} = :\mathfrak{L}(\Lambda;k,z),\end{aligned} \tag{19}$$

where

$$V^{\blacktriangle}(q,k,\mu) = n^{-3}\beta^{-1/2}V(q,k,\mu). \tag{20}$$

This equation has the trivial solution $\Lambda = 0$ corresponding to $\Sigma = \infty$ and $\Xi = 1$, unless condition (14) is satisfied.

But this does not imply the absence of a nontrivial solution, because for the trivial solution, ξ_0 of Theorem 2 is 0, so that $S(0,N\xi_0) = \{0\}$. Though $\mathfrak{L}$ is not Fréchet differentiable, one can take the Lipschitz approximation

$$\begin{aligned}\mathfrak{M}(\Lambda;k,z) =: &\frac{k^{\delta}}{n^3\beta^{1/2}(k^2+\lambda)^{\epsilon/2}}\left[n^{-2}\tilde{\gamma}(k) + \beta^{-1/2}\int_{-i\infty}^{i\infty}\frac{dz_1}{2\pi i}\int_0^{\infty} dq\int_{-1}^{1} d\mu\, V^{\blacktriangle}(q,k,\mu)\right.\\ &\times[\{z_1+\beta^{-1}\omega^2(q_+)\Lambda(q_+,z_1)\ \ (q_+^2+\lambda)^{\epsilon/2}q_+^{-\delta}\}^{-1}\theta_\rho(\{(q+\tfrac{1}{2}k\mu)^2+(|\mu|-1)^2\tau_1^2+|z_1^2|\tau_2^2\}^{1/2}-\nu)\\ &+\{z_1+\beta^{-1}\omega^2(q_+)\Lambda_{(0)}(q_+,z)(q_+^2+\lambda)^{\epsilon/2}q_+^{-\delta}\}^{-1}\\ &\times\{1-\theta_\rho(\{(q+\tfrac{1}{2}k\mu)^2+(|\mu|-1)^2\tau_1^2+|z_1^2|\tau_2^2\}^{1/2}-\nu)\}]\\ &\times[\{z-z_1+\beta^{-1}\omega^2(q_-)\Lambda(q_-,z-z_1)(q_-^2+\lambda)^{\epsilon/2}q_-^{-\delta}\}^{-1}\\ &\times\theta_\rho(\{(q-\tfrac{1}{2}k\mu)^2+(|\mu|-1)^2\tau_1^2+|(z-z_1)^2|\tau_2^2\}^{1/2}-\nu)\\ &+\{z-z_1+\beta^{-1}\omega^2(q_-)\,\Lambda_{(0)}(q_-,z-z_1)(q_-^2+\lambda)^{\epsilon/2}q_-^{-\delta}\}\\ &\left.\times\{1-\theta_\rho(\{(q-\tfrac{1}{2}k\mu)^2+(|\mu|-1)^2\tau_1^2+|(z-z_1)^2|\tau_2^2\}^{1/2}-\nu)\}]\right]^{-1},\end{aligned}\tag{21}$$

with a suitable $\Lambda_{(0)}$ and apply Theorem 1 and 2 with a zeroth approximation

$$\Lambda^{(0)}(k,z) = n^{-1}\beta^{-1/2}k^{\delta}(k^2+\lambda)^{-\epsilon/2}\tilde{\gamma}(k).$$

In order to find the critical values of n and β, one has to investigate $\Lambda(k,0)$ as a function of n, β, and k.

If Eqs. (6) and (19) admit a solution, then

$$\lim_{n\to 0}\Sigma(k,z) = -\gamma(k) = -n\beta^{-1/2}\tilde{\gamma}(k),\quad \lim_{n\to\infty}\Sigma(k,z)=\infty,\quad \lim_{\beta\to 0}\Sigma(k,z)=\infty,\quad \lim_{\beta\to\infty}\Sigma(k,z)=0,\tag{22}$$

should hold.

It should be noticed that the unsimplified equation (6) cannot have a solution of the form $\Sigma(k,z) = z^{-1}\Sigma_1(k) - \Sigma_2(k,z)$, $\Sigma_1(k)\not\equiv 0$. Should such a solution exist, Σ_1 would have to satisfy the equation

$$\Sigma_1(k) = \frac{n}{\beta}\int_0^{\infty} dq\int_{-1}^{1} d\mu\,\frac{\tilde{V}(q,k,\mu)\Sigma_1(q_+)\Sigma_1(q_-)}{(\omega^2(q_+)+\Sigma_1(q_+))(\omega^2(q_-)+\Sigma_1(q_-))} \equiv \mathfrak{G}(\Sigma_1,k).\tag{23}$$

Obviously this equation has trivial solution $\Sigma_1\equiv 0$. Because the Fréchet derivative of $\mathfrak{G}_{(1)}(\Sigma_1;k,z)$ at $\Sigma_1 = 0$ is identically zero, a nontrivial solution cannot exist. On the other hand, bifurcation theory is not applicable to Eqs. (6) and (19), because the mappings $\mathfrak{G}$ and $\mathfrak{X}$ are not Fréchet differentiable. However, if $c(q)$ and $c^{(3)}(q,k,\mu)$ behave asymptotically for $q\to\infty$,

$$c(q) = O(q^{-\eta}),\quad \eta>\tfrac{5}{2},\quad c^{(3)}(q,k,\mu) = O((k^2+q^2)^{-1/4})\quad\text{not}\quad o((k^2+q^2)^{-1/4}),\tag{24}$$

asymptotic behavior $\Sigma(k,z)\sim k^{\alpha}$, $0<\alpha<2$, is consistent with Eq. (6), but a solution with such asymptotic behavior cannot be found by the Newton–Kantorovich–Lika (NKL) algorithm with zeroth approximation $\Sigma^{(0)} = 0$. In order to find such solutions, one has to rewrite Eq. (6) in the form

$$\begin{aligned}\tilde{\Sigma}(k,z)\equiv k^{-\alpha}\Sigma(k,z) = &\frac{nk^{-\alpha}}{\beta}\int\frac{dz_1}{2\pi i}\int dq\int d\mu\,\tilde{V}(q,k,\mu)\{z_1+\omega^2(q_+)\beta^{-1}(q_+^{\alpha}\,\tilde{\Sigma}(q_+,z_1)+n\beta^{-1/2}\tilde{\gamma}(q_+))^{-1}\}^{-1}\\ &\times\{z-z_1+\omega^2(q_-)\beta^{-1}(q_-^{\alpha}\,\tilde{\Sigma}(q_-,z-z_1)+n\beta^{-1/2}\tilde{\gamma}(q_-))^{-1}\}^{-1},\end{aligned}\tag{25}$$

and fix a zeroth approximation so as to make the left-hand side and the right-hand side coincide for certain discrete values of the variables k and z and apply the NKL algorithm. Similarly, Eq. (19) may have a solution with asymptotic behavior

$$\Lambda(k,z)\sim k^{-\alpha'+\delta-\epsilon'},\quad 0<\alpha' = \mathrm{Max}(\alpha,\delta)<2,\quad k\to\infty.$$

III. CONCLUDING REMARKS

Though without actually (numerically) solving Eq. (6) and/or Eq. (19) one cannot definitely conclude whether the Kirkpatrick model without simplication describes glass transition, these equations deserve further investigation because they have many interesting mathematical features. Another question to be investigated is under what conditions solutions of simplified equations give reliable conclusions about critical phenomena.

[1] T. R. Kirkpatrick, Phys. Rev. A **31**, 939 (1985).
[2] E. Leutheusser, Phys. Rev. A **29**, 2765 (1984).
[3] U. Bergtzelius, W. Götze, and A. Sjölander, J. Phys. C **17**, 5915 (1984).
[4] G. H. A. Cole, *An Introduction to the Statistical Theory of Classical Simple Dense Fluids* (Pergamon, London, 1967).
[5] R. Resibois and M. DeLeener, *Classical Kinetic Theory of Fluids* (Wiley, New York, 1977).
[6] D. K. Lika, *Analiza funktsionale shi rezolvarya aproksimative a ekuatsiilor neliniare* (*Functional Analysis and Approximate Solutions of Nonlinear Equations*) (Editura Shtiintsa, Kishineu, 1975).

A class of continuum models with no phase transitions

David Klein
Department of Mathematics, University of California, Los Angeles, California 90024

(Received 2 March 1986; accepted for publication 28 July 1986)

For a restricted family of classical grand canonical continuum interactions, it is proved that the Gibbs state is unique at all temperatures and fugacities. The interactions considered are not translation invariant except in the one-dimensional case.

I. INTRODUCTION

More than ten years ago, Dobrushin developed a probabilistic technique for proving uniqueness of Gibbs states and applied it to one-dimensional classical lattice and hard-core systems to establish the absence of phase transitions for those systems.[1] In this paper, we apply the same technique to a restricted class of continuum interactions in arbitrary dimension. Since one expects phase transitions in dimensions greater than one for physically realistic interactions, the restrictions needed for our results must be strong. We consider positive, superstable interactions with the additional requirement that, loosely put, in dimension d the range of the interaction in the radial direction decreases for some constant c, as $(cr+c)^{1/d}-(cr)^{1/d}$, when the distance r of particles from the origin becomes sufficiently large. For the precise requirements, see Condition 2.1 in Sec. II below. In spite of the artificiality of the models considered here, we believe that our results can shed some light on the conditions which make phase transitions possible in more realistic models. In addition the results obtained here show, in a sense, what is really going on in the one-dimensional case, where the restrictions on the interaction potential are natural. The specialization to the one-dimensional case is made in Sec. IV, where we also give some further discussion. For positive results on the existence of phase transitions for lattice and some simple continuum systems, we refer the reader to a paper of Bricmont, Kuroda, and Lebowitz[2] and the references contained therein. More general background information can be found in Refs. 3 and 4.

II. NOTATION AND DEFINITIONS

For a Borel measurable subset $\Lambda\subset\mathbb{R}^d$, let $X(\Lambda)$ denote the set of all locally finite subsets (configurations) of Λ. Here B_Λ denotes the σ field on $X(\Lambda)$ generated by all sets of the form $\{s\in X(\Lambda): |s\cap B|=m\}$, where B runs over all bounded Borel subsets of Λ, m runs over the set of nonnegative integers, and $|\cdot|$ denotes cardinality. For later convenience we let $(\Omega,S)=(X(\mathbb{R}^d),B_{\mathbb{R}^d})$. Let Ω_F be the set of configurations in Ω of finite cardinality and $X_N(\Lambda)$ the set of configurations in $X(\Lambda)$ of cardinality N.

Let $T:\Lambda^N\to X_N(\Lambda)$ be the map which takes the ordered N tuple $(x_1,\dots,x_N)$ to be the unordered set $\{x_1,\dots,x_N\}$. For $N=1,2,3,\dots$, let d^Nx be the projection of Nd-dimensional Lebesgue measure onto $X_N(\Lambda)$ under the map T. The measure d^0x assigns mass 1 to $X_0(\Lambda)=\{\varnothing\}$. When Λ is bounded, define as in Refs. 3 and 5,

$$\nu_\Lambda(dx)=\sum_{N=0}^{\infty}\frac{z^N}{N!}d^Nx, \tag{2.1}$$

where z is chemical activity. For any bounded disjoint Borel sets Λ and Λ', there is a natural isomorphism between

$$(X(\Lambda),B_\Lambda,\nu_\Lambda)\times(X(\Lambda'),B_{\Lambda'},\nu_{\Lambda'})$$

and

$$(X(\Lambda\cup\Lambda'),B_{\Lambda\cup\Lambda'},\nu_{\Lambda\cup\Lambda'}).$$

We will identify these spaces and write

$$\begin{aligned}(X(\Lambda\cup\Lambda'),&B_{\Lambda\cup\Lambda'},\nu_{\Lambda\cup\Lambda'})\\&=(X(\Lambda),B_\Lambda,\nu_\Lambda)\times(X(\Lambda'),B_{\Lambda'},\nu_{\Lambda'}).\end{aligned} \tag{2.2}$$

We will consider S-measurable many-body interactions $V:\Omega_F\to(-\infty,\infty]$ of the form

$$V(x)=\sum_{N=1}^{\infty}\sum_{\substack{y\subset x\\|y|=N}}\phi_N(y), \tag{2.3}$$

where $\phi_N: X_N(\Lambda)\to(-\infty,\infty]$ is an N-body interaction. For a bounded Borel set Λ of positive Lebesgue measure, we define as in Preston[3] the S-measurable set $R_\Lambda\subset\Omega$ so that $V(x|s)$ represents the energy of the configuration $x\in X(\Lambda)$ in Λ, assuming the configuration $s\in R_\Lambda\cap X(\Lambda^c)$ outside of Λ. The finite volume Gibbs state $\mu_\Lambda(dx|s)$ for Λ,V, inverse temperature β, chemical activity z, and external configuration s is given by

$$\mu_\Lambda(dx|s)=\exp[-\beta V(x|s)]/Z_\Lambda(s), \tag{2.4}$$

where $Z_\Lambda(s)$ makes $\mu_\Lambda(dx|s)$ a probability measure. Under general conditions on V, in particular Condition 2.1 below, $1\leqslant Z_\Lambda(s)<\infty$. If $s\notin R_\Lambda$, define $\mu_\Lambda(dx|s)$ to be the zero measure.

Henceforth let a positive constant c be fixed and let Λ_k denote the hypersphere of volume ck centered at the origin in $\mathbb{R}^d$. Let $A_k=\Lambda_k\setminus\Lambda_{k-1}$. If $d=1$, then Λ_k is the interval of length ck centered at the origin.

Condition 2.1: For an interaction V of the form (2.3):

(a) V is superstable[6] and ϕ_N is positive for each $N\geqslant 2$.

(b) There exists some positive integer k_0 such that for all $N\geqslant 2$ and any $x=(x_1,\dots,x_N)\in\Omega_F$,

$$\phi_N(x)=0$$

whenever $x_i\in\Lambda_k$, $x_j\in\Lambda_{k+1}^c$ for some pair i,j with $1\leqslant i,j\leqslant N$ and some $k\geqslant k_0$.

Remark 2.1: The requirement that ϕ_N be positive in Condition 2.1(a) can be removed if V has an nth body hard-core restriction for some n, analogous to that imposed in Refs. 5 and 6. Condition 2.1(b) excludes translation-invariant interactions except in dimension $d=1$. For $d\geqslant 2$, the

 0022-2488/86/122872-04$02.50

range in the radial direction of the interaction must decrease as the distance of particles from the origin increases.

Let $\{\pi_\Lambda\}$ denote the specification associated with β, z, and V (see Ref. 3, p. 16) defined by

$$\pi_\Lambda(A,s) = \int_{A'} \mu_\Lambda(dx|s\cap\Lambda^c), \tag{2.5}$$

where $s\in\Omega$, $A\in S$, and $A' = \{x\in X(\Lambda): x\cup(s\cap\Lambda^c)\in A\}$.

Definition 2.1: A probability measure on (Ω,s) is a Gibbs state for the specification $\{\pi_\Lambda\}$ if

$$\sigma(\pi_\Lambda(A,s)) = \sigma(A)$$

for every $A\in S$ and every bounded Borel set $\Lambda\subset\mathbf{R}^d$ of positive measure.

Definition 2.2: For an interaction V, bounded Borel sets $\Lambda\subset\widetilde{\Lambda}$ with positive Lebesgue measure, and $s\in R_{\widetilde{\Lambda}}$, the finite volume Gibbs density $r_{\widetilde{\Lambda}}^{\Lambda}(x|s)$ is given by

$$r_{\widetilde{\Lambda}}^{\Lambda}(x|s) = \int_{X(\widetilde{\Lambda}\setminus\Lambda)} \frac{\exp[-\beta V(x\cup y|s\cap\widetilde{\Lambda}^c)]}{Z_{\widetilde{\Lambda}}(s\cap\widetilde{\Lambda}^c)}\, \nu_{\widetilde{\Lambda}\setminus\Lambda}(dy). \tag{2.6}$$

Note that if f is S measurable and $f(x) = f(x\cap\Lambda)$ for all $x\in\Omega$, then

$$\pi_{\widetilde{\Lambda}}(f,s) \equiv \int f(x)\mu_{\widetilde{\Lambda}}(dx|s) = \int_{X(\Lambda)} f(x) r_{\widetilde{\Lambda}}^{\Lambda}(x|s)\nu_\Lambda(dx). \tag{2.7}$$

III. UNIQUENESS FOR ARBITRARY DIMENSION

In this section we prove, for any β and z, uniqueness of the Gibbs state correspondng to an interaction V satisfying Condition 2.1. The method of proof is based on an application of a result of Dobrushin (Ref. 1, Lemma 1) which we restate below for the convenience of the reader as Lemma 3.1.

Let (X,B_X) be a measurable space and let μ_1 and μ_2 be probability measures on (X,B_X). The variation distance between the measures μ_1 and μ_2 is defined as

$$\rho(\mu_1,\mu_2) = \sup_{A\in B_X} |\mu_1(A) - \mu_2(A)|. \tag{3.1}$$

If μ_1 and μ_2 have respective densities p_1 and p_2 with respect to a finite measure ν on X, then defining $\rho(p_1,p_2) = \rho(\mu_1,\mu_2)$, we have

$$\rho(p_1,p_2) = \frac{1}{2}\int_X |p_1(x) - p_2(x)|\nu(dx)$$
$$= 1 - \int_X \min(p_1(x),p_2(x))\nu(dx). \tag{3.2}$$

Uniqueness of the Gibbs state for V, β, and z follows provided

$$\lim_{n\to\infty} \sup_{s,t\in\Omega} |\pi_{\Lambda_n}(f,s) - \pi_{\Lambda_n}(f,t)| = 0 \tag{3.3}$$

for any bounded S-measurable function f which satisfies $f(x) = f(x\cap\Lambda)$ for some bounded set $\Lambda\subset\mathbf{R}^d$ and all $x\in\Omega$ (see Lemma 9.3, Ref. 3). From (2.7) and (3.2), Eq. (3.3) holds as long as

$$\lim_{n\to\infty} \sup_{t,s\in\Omega} \rho\left[r_{\Lambda_n}^{\Lambda_k}(\cdot|s), r_{\Lambda_n}^{\Lambda_k}(\cdot|t)\right] = 0 \tag{3.4}$$

for each fixed positive integer $k > k_0$, where k_0 is given by Condition 2.1.

Lemma 3.1 (Dobrushin): Let (X_j, B_j, ν_j) be a measure space for $j = 1,2,3$, and let

$$(X,B_X,\nu) = \prod_{j=1}^{3} (X_j, B_j, \nu_j)$$

be the product measure space with measure $\nu = \nu_1\times\nu_2\times\nu_3$. Let $p^1(\cdot)$ and $p^2(\cdot)$ be densities with respect to ν for probability measures on (X,B_X). Consider the marginal densities

$$p_1^i(x_1) = \iint p^i(x_1,x_2,x_3)\nu_2(dx_2)\nu_3(dx_3),$$

$$p_{1,2}^i(x_1,x_2) = \int p^i(x_1,x_2,x_3)\nu_3(dx_3), \quad \text{for} \quad i = 1,2,$$

and the similarly defined densities $p_2^i(x_2)$, $p_3^i(x_3)$, $p_{1,3}^i(x_1,x_3)$, and $p_{2,3}^i(x_2,x_3)$ for $i = 1,2$. Suppose there exist conditional densities $p_1^i(x_1|x_2,x_3)$ and $p_{1/2}^i(x_1|x_2)$ for which

$$p^i(x_1,x_2,x_3) = p_1^i(x_1|x_2,x_3)p_{2,3}^i(x_2,x_3),$$
$$p_{1,2}^i(x_1,x_2) = p_{1/2}^i(x_1|x_2)p_2^i(x_2) \quad (i = 1,2).$$

Then

$$\rho(p_1^1,p_1^2) \leqslant \alpha\rho(p_2^1,p_2^2) + \overline{\alpha}(1 - \rho(p_2^1,p_2^2)),$$

where

$$\alpha = \sup_{\substack{x_j,\tilde{x}_j\in X_j\\ j=1,2}} \rho(p_1^1(\cdot|x_2,x_3), p_1^2(\cdot|\tilde{x}_2,\tilde{x}_3)),$$

$$\overline{\alpha} = \sup_{\substack{x_2\in X_2\\ x_3,\tilde{x}_3\in X_3}} \rho(p_1^1(\cdot|x_2,x_3), p_1^2(\cdot|x_2,\tilde{x}_3)).$$

Let some integer $k > k_0$ be fixed and suppose $n > k+1$. In the language of Dobrushin's lemma, we let

$$\begin{aligned}(X_1,B_1) &= (X(\Lambda_k),B_{\Lambda_k}),\\ (X_2,B_2) &= (X(A_{k+1}),B_{A_{k+1}}),\\ (X_3,B_3) &= (X(\Lambda_n\setminus\Lambda_{k+1}),B_{\Lambda_n\setminus\Lambda_{k+1}}).\end{aligned} \tag{3.5}$$

For a configuration $x\in\Omega$, let

$$x_1 = x\cap\Lambda_k, \quad x_2 = x\cap A_{k+1}, \quad x_3 = x\cap(\Lambda_n\setminus\Lambda_{k+1}).$$

The measure ν of Lemma 3.1 is identified as $\nu = \nu_{\Lambda_k}\times\nu_{A_{k+1}}\times\nu_{\Lambda_n\setminus\Lambda_{k+1}} = \nu_{\Lambda_n}$ because of (2.2). For $i = 1,2$ let

$$p^i(x_1,x_2,x_3) = \frac{\exp[-\beta V(x_1\cup x_2\cup x_3|s_i)]}{Z_{\Lambda_n}(s_i)}, \tag{3.6}$$

where $s_1,s_2\in X(\Lambda_n^c)$ are chosen arbitrarily. It follows that

$$p_1^i(x_1) = r_{\Lambda_n}^{\Lambda_k}(x_1|s_i), \tag{3.7}$$

$$p_2^i(x_2) = r_{\Lambda_n}^{A_{k+1}}(x_2|s_i), \tag{3.8}$$

$$p_1^i(x_1|x_2,x_3) = \frac{\exp[-\beta V(x_1|x_2\cup x_3\cup s_i)]}{Z_{\Lambda_k}(x_2\cup x_3\cup s_i)}. \tag{3.9}$$

Lemma 3.2: Let V satisfy Condition 2.1. Then for any $\beta,z > 0$, $s_1,s_2\in X(\Lambda_k^c)$, $k > k_0$, and $y\in(\Lambda_{k-1})$,

$$\frac{\exp[-\beta V(\varnothing_k \cup y|s_2)]}{Z_{\Lambda_k}(s_2)}$$

$$\geqslant \exp(-cz)\frac{\exp[-\beta V(\varnothing_k \cup y|s_1)]}{Z_{\Lambda_k}(s_1)}, \tag{3.10}$$

where $\varnothing_k$ is the empty configuration in A_k and c is given in the definition of Λ_k.

Proof: Since V is positive, $V(\varnothing_k \cup y|s_2) < V(y_k \cup y|s_2)$ for any $y_k \in X(A_k)$. It follows that

$$\int_{X(\Lambda_{k-1})}\int_{X(A_k)} \exp[-\beta V(\varnothing_k \cup y|s_2)]\nu_{A_k}(dy_k)\nu_{\Lambda_{k-1}}(dy)$$
$$> \int_{X(\Lambda_{k-1})}\int_{X(A_k)} \exp[-\beta V(y_k \cup y|s_2)] \times \nu_{A_k}(dy_k)\nu_{\Lambda_{k-1}}(dy).$$

From (2.2) it follows that

$$\int_{X(\Lambda_{k-1})} \exp[-\beta V(\varnothing_k \cup y|s_2)]\nu_{\Lambda_{k-1}}(dy)$$
$$\geqslant [\nu_{A_k}(X(A_k))]^{-1} Z_{\Lambda_k}(s_2) = \exp(-cz)Z_{\Lambda_k}(s_2). \tag{3.11}$$

Also

$$Z_{\Lambda_k}(s_1) = \int_{X(\Lambda_{k-1})}\int_{X(A_k)} \exp[-\beta V(\varnothing_k \cup y|s_1)] \times \exp[-\beta V(y_k|y \cup s_1)]\nu_{A_k}(dy_k)\nu_{\Lambda_{k-1}}(dy)$$
$$\geqslant \int_{X(\Lambda_{k-1})} \exp[-\beta V(\varnothing_k \cup y|s_1)]\nu_{\Lambda_{k-1}}(dy), \tag{3.12}$$

because

$$\int_{X(A_k)} \exp[-\beta V(y_k|y \cup s_1)]\nu_{A_k}(dy_k) \geqslant \nu_{A_k}(\varnothing_k) = 1.$$

From Condition 2.1,

$$V(\varnothing_k \cup y|s_1) = V(\varnothing_k \cup y|s_1) = V(y). \tag{3.13}$$

Combining (3.11)–(3.13) gives

$$Z_{\Lambda_k}(s_1) \geqslant \int_{X(\Lambda_{k-1})} \exp[-\beta V(\varnothing_k \cup y|s_2)]\nu_{\Lambda_{k-1}}(dy)$$
$$\geqslant \exp(-cz)Z_{\Lambda_k}(s_2). \tag{3.14}$$

Now combining (3.13) and (3.14) yields the desired result. This completes the proof.

Lemma 3.3: Let V satisfy Condition 2.1. Then for any $\beta, z > 0$ there exists an $h > 0$, independent of k such that

$$\rho[p^1(\cdot|x_2,x_3), p^2(\cdot|\tilde{x}_2,\tilde{x}_3)] \leqslant 1 - h, \tag{3.15}$$

for all $s_1,s_2,x_2,\tilde{x}_2,x_3,\tilde{x}_3$.

Proof: Since V is positive and A_k has Lebesgue measure c for each k,

$$\sup_{\substack{k>k_0\\ s\in\Omega}} \int_{X(A_k)} \exp[-\beta V(x|s \cap \Lambda_k^c)]\nu_{A_k}(dx) < \infty. \tag{3.16}$$

For any interaction V and any $s\in\Omega$,

$$\exp[-\beta V(\varnothing_k|s \cap \Lambda_k^c)] = 1,$$

where $\varnothing_k$ is the empty configuration in A_k. Thus

$$\inf_{\substack{k>k_0\\ t\in\Omega}} \pi_{A_k}(\{\varnothing_k\},t) \equiv h_1 > 0. \tag{3.17}$$

From the consistency[3] of the specification,

$$\pi_{\Lambda_k}(\{\varnothing_k\},s) = \int \pi_{A_k}(\{\varnothing_k\},t)\pi_{\Lambda_k}(dt,s) \geqslant h_1, \tag{3.18}$$

for any $s\in X(\Lambda_k^c)$. Hence,

$$\pi_{\Lambda_k}(\{\varnothing_k\},s) = \int_{\{\varnothing_k\}\cup X(\Lambda_{k-1})} \frac{\exp[-\beta V(\varnothing_k \cup (y \cap \Lambda_{k-1})|s)]}{Z_{\Lambda_k}(s)} \times \nu_{\Lambda_k}(dy) \geqslant h_1. \tag{3.19}$$

Combining (3.10) with (3.19) and (3.9) yields

$$\int_{\{\varnothing_k\}\cup X(\Lambda_{k-1})} \min[p^1(x_1|x_2,x_3), p^2(x_1|\tilde{x}_2,\tilde{x}_3)]\nu_{\Lambda_k}(dx_1)$$
$$\geqslant h_1 \exp(-cz), \tag{3.20}$$

for all $s_1,s_2,x_2,\tilde{x}_2,x_3,\tilde{x}_3$. Now combining (3.20) with (3.2) gives (3.15) with $h = h_1\exp(-cz)$. This concludes the proof.

Lemma 3.4: Let V satisfy Condition 2.1 and let $\beta, z > 0$. For any integers n and k with $n \geqslant k + 3 > k_0 + 3$ and any $s,t\in\Omega$,

$$\rho[r_{\Lambda_n}^{\Lambda_k}(\cdot|s), r_{\Lambda_n}^{\Lambda_k}(\cdot|t)]$$
$$\leqslant (1-h)\rho[r_{\Lambda_n}^{\Lambda_{k+1}}(\cdot|s), r_{\Lambda_n}^{\Lambda_{k+1}}(\cdot|t)], \tag{3.21}$$

where h is given by Lemma 3.3.

Proof: With the identifications (3.5)–(3.9), the term $\bar{\alpha}$ in Lemma 3.1 is zero for any choices of $k \geqslant k_0$ and $n \geqslant k + 3$ because of Condition 2.1(b).

From Lemma 3.3 we obtain $\alpha \leqslant 1 - h$, where h is independent of k and n. Applying Lemma 3.1 together with (3.7) and (3.8) then gives

$$\rho(r_{\Lambda_n}^{\Lambda_k}(\cdot|s), r_{\Lambda_n}^{\Lambda_k}(\cdot|t))$$
$$\leqslant (1-h)\rho(r_{\Lambda_n}^{A_{k+1}}(\cdot|s), r_{\Lambda_n}^{A_{k+1}}(\cdot|t)). \tag{3.22}$$

But

$$\rho(r_{\Lambda_n}^{A_{k+1}}(\cdot|s), r_{\Lambda_n}^{A_{k+1}}(\cdot|t)) \leqslant \rho(r_{\Lambda_n}^{\Lambda_{k+1}}(\cdot|s), r_{\Lambda_n}^{\Lambda_{k+1}}(\cdot|t)), \tag{3.23}$$

because $A_{k+1} \subset \Lambda_{k+1}$. The proof is completed by combining (3.22) and (3.23).

Theorem 3.1: Let V satisfy Condition 2.1. Then for any $\beta, z > 0$, there is at most one Gibbs state for V.

Proof: A simple induction argument together with (3.21) shows that given any positive integer l,n can be chosen large enough so that

$$\sup_{s,t\in\Omega} \rho[r_{\Lambda_n}^{\Lambda_k}(\cdot|s), r_{\Lambda_n}^{\Lambda_k}(\cdot|t)]$$
$$\leqslant (1-h)^l \sup_{s,t\in\Omega} \rho[r_{\Lambda_n}^{\Lambda_{k+l}}(\cdot|s), r_{\Lambda_n}^{\Lambda_{k+l}}(\cdot|t)].$$

Equation (3.4) now follows by letting n and l go to infinity. This concludes the proof.

Remark 3.1: Existence of Gibbs states for interactions of the type considered here can be established using the methods of Refs. 3 and 7.

IV. THE ONE-DIMENSIONAL CASE

In this section we consider the question of uniqueness of the Gibbs state for all $\beta,z > 0$ in the special case $d = 1$. When $d = 1$, Theorem 3.1 specializes to the following result for translation-invariant interactions.

Corollary 4.1: Let $d = 1$ and suppose V is a translation-invariant, finite-range, positive, superstable interaction. Then V has a unique Gibbs state for all positive β and z.

Proof: From the definition of Λ_k in Sec. I, Λ_k is the interval of length ck centered at the origin of $\mathbb{R}^1$. Choose c to be a number at least twice as large as the range of V. With this choice of c, V satisfies Condition 2.1 and the result follows from Theorem 3.1.

Remark 4.1: Using different methods, this result was already established as a special case in Ref. 8 and subsequently in Sec. II of Ref. 6, where it was proved but not explicitly stated.

Conditions on the interaction other than those imposed by Corollary 4.1 have been studied in regard to the question of uniqueness of the Gibbs state for $d = 1$. In Ref. 5 uniqueness was proved for infinite range, translation-invariant (not necessarily positive), superstable interactions for which there is an n-body hard-core restriction for some $n \geqslant 2$. Previously, Dobrushin[1] proved uniqueness and related results with similar restrictions for the important case $n = 2$.

In Ref. 6 an incorrect proof was given for uniqueness of the Gibbs state for finite-range, superstable interactions (not necessarily positive). With minor modifications, the proof establishes the following. For any $\beta,z > 0$, let V be an interaction, at least one of whose Gibbs states is supported on

$$U_\infty = \bigcup_{m \geq 1} U_m,$$

where

$$U_m = \{s \in \Omega: |s \cap [n,n+1)| < m \text{ for all integers } n\}.$$

Then exactly one Gibbs state for V is supported on U_∞ for each $\beta,z > 0$.

For translation-invariant interactions, the condition that a Gibbs state for V is supported on U_∞ evidently requires a hard-core restriction of the type considered in Refs. 5 and 6. But when translation invariance of the interaction V is not required, then one would expect some Gibbs state for V to be supported on U_∞ provided an external repulsive force (i.e., one-body interaction) increases sufficiently rapidly as the distance from a particle to the origin increases. Alternatively if V is such that the repulsion between nearby particles increases sufficiently rapidly as their distance from the origin increases, one might also expect some Gibbs state for V to be supported on U_∞. As a simple example of a nontranslation-invariant interaction whose unique Gibbs state is supported on U_∞, but which is not of a type considered in any of the previous references, we give the following.

Example 4.1: Let V_1 be any finite-range, superstable interaction. For any fixed positive integers k and N and any $x = (x_1,\dots,x_N) \in \Omega_F$, let

$$\Phi_N^k(x) = \begin{cases} \infty, & \text{if } \max\limits_{ij} |x_i - x_j| < 1 \text{ and } x \not\subset [-k,k], \\ 0, & \text{otherwise.} \end{cases}$$

For any $x \in \Omega_F$ define

$$V(x) = V_1(x) + \sum_{\substack{y \subset x \\ |y| = N}} \Phi_N^k(y).$$

Then any Gibbs state for V is supported on U_∞, and consequently by the method of Ref. 6, the Gibbs state for V is unique for any $\beta,z > 0$.

[1]R. L. Dobrushin, "Conditions for the absence of phase transitions on one-dimensional classical systems," Math. Sb. **93**, 29 (1974).

[2]J. Bricmont, K. Kuroda, and J. Lebowitz, "First order phase transitions in lattice and continuous systems: Extension of Pirogov-Siani theory," Commun. Math. Phys. **101**, 501 (1985).

[3]C. Preston, "Random fields," in *Lecture Notes in Mathematics*, Vol. 534 (Springer, Berlin, 1976).

[4]L. Gross, "Thermodynamics, statistical mechanics and random fields. Ecole d'Ete' de Probabilites de Saint-Flour X—1980," in *Lecture Notes in Mathematics*, Vol. 929 (Springer, Berlin, 1982).

[5]D. Klein, "Uniqueness of continuum one-dimensional Gibbs states for slowly decaying interactions," J. Stat. Phys. **43**, 207 (1986).

[6]D. Klein, "Uniqueness of the one-dimensional continuum Gibbs states," Commun. Math. Phys. **97**, 595 (1985).

[7]D. Ruelle, "Superstable interactions in classical statistical mechanics," Commun. Math. Phys. **18**, 127 (1970).

[8]Yu. Sukov, "Random point processes and DLR equations," Commun. Math. Phys. **50**, 113 (1976).

Central-limit theorems on groups

P. A. Mello[a)]
Instituto de Física, Universidad Nacional Autonoma de México, Mexico[b)]

(Received 4 March 1986; accepted for publication 28 July 1986)

The probability density p_N of the product of N statistically independent (and identically distributed, each with probability density p_1) elements of a group is studied in the limit $N \to \infty$. It is shown, for the compact groups $R(2)$ and $R(3)$, that $p_N \to 1$ as $N \to \infty$, independently of p_1. It is made plausible that a similar behavior is to be expected for other compact groups. For noncompact groups, the case of SU(1,1), which is of interest to the physics of disordered conductors, is studied. The case in which p_1 is isotropic, i.e., independent of the phases, is analyzed in detail. When p_1 is fixed and $N \gg 1$, a Gaussian distribution in the appropriate variable is found. When the original variables are rescaled by $1/N$ and the limit $N \to \infty$ is taken, keeping the ratio of the length of the conductor to the localization length fixed, an explicit integral representation for the resulting probability density is found. It is also exhibited that the latter satisfies a "diffusion" equation on the group manifold.

I. INTRODUCTION

The central-limit theorem[1] (CLT) is one of the most powerful results in the theory of statistics and has very important consequences in many physical problems. It studies the distribution of the sum x of N statistically independent random variables x_i,

$$x = \sum_{i=1}^{N} x_i , \tag{1.1}$$

in the limit when $N \to \infty$. With only very general assumptions on the individual distributions, it states that $\xi = (x - \bar{x})/(\operatorname{var} x)^{1/2}$ (with $\bar{x} = \Sigma_i \bar{x}_i$, $\operatorname{var} x = \Sigma_i \operatorname{var} x_i$) becomes a Gaussian variable with zero centroid and variance 1. An alternative statement is that, under successive convolutions, a distribution finally approaches a Gaussian limit. The power of the theorem lies in the insensitivity of the asymptotic distribution to the distributions of the individual x_i's.

The standard proof of the CLT uses the fact that the Fourier transform (FT) of a convolution is the product of the individual FT's. One then finds, for the limit of that product when the number N of factors grows, a Gaussian in k (the variable conjugate to x), which implies a Gaussian in x.

Alternatively, we can rephrase the above problem in the following language. Consider, along some axis, successive translations $T(x_i)$ by the amount x_i, and designate by $T(x)$ the resulting translation by the amount $x = \Sigma_{i=1}^{N} x_i$. Equation (1.1) is then equivalent to

$$T(x) = T(x_1) \cdots T(x_N) . \tag{1.2}$$

The $T(x_i)$'s are the elements of the *translation group* T_1. If D denotes any representation of T_1, (1.2) implies

$$D(x) = D(x_1) \cdots D(x_N) . \tag{1.3}$$

Call $p_N(x)$ the probability density of x and assume for simplicity that all the x_i's have the same probability density $p_1(x_i)$. Averaging both sides of (1.3) we then have

$$\langle D \rangle_N = \langle D \rangle_1^N , \tag{1.4}$$

where we have used the notation

$$\langle D \rangle_1 = \int D(x) p_1(x)\, dx , \tag{1.5a}$$

$$\langle D \rangle_N = \int D(x) p_N(x)\, dx . \tag{1.5b}$$

In particular, if D is the unitary irreducible representation e^{ikx}, the problem reduces to finding the limiting form of $\langle e^{ikx} \rangle_1^N$, and then inverting the FT to obtain $p_N(x)$.

There is a problem of current physical interest that can be formulated precisely in the above language and that motivated the present investigation. It is the one-dimensional problem of scattering by a succession of N random scatterers.[2] The scattering by the ith center can be represented by a 2×2 transfer matrix R_i, which, from flux conservation and time-reversal invariance, must be pseudounitary and unimodular. The collection of matrices R_i forms the group SU(1,1) (see Refs. 3 and 4), which is a noncompact group homomorphic to the Lorentz group SO(2,1). The resulting transfer matrix R can be written as

$$R = R_1 R_2 \cdots R_N . \tag{1.6}$$

This equation has the same structure as (1.2). Suppose that we consider the individual transfer matrices R_i ($i = 1,...,N$) as statistically independent and that we propose, for each one of them, the same probability density $p_1(R_i)$. The problem is to find the distribution $p_N(R)$ of the resulting R of (1.6) and study whether in the large-N limit the answer is independent of the input $p_1(R_i)$. If this were so, we would have a CLT on the group SU(1,1), just as the ordinary CLT can be associated with the group T_1.

The above concepts can actually be studied in connection with an arbitrary group, so that, besides its physical applications, the problem has an intrinsic mathematical interest.

The above questions have previously been addressed in

[a)] Also at Departamento de Física, U.A.M.-Iztapalapa, Mexico, and fellow of the Sistema Nacional de Investigadores, Mexico.
[b)] Mailing address: Instituto de Física, UNAM, Apartado Postal 20-364, Delegación Alvaro Obregón, 01000 México D.F., Mexico.

the literature by several authors.[5–9] For instance, in Ref. 7(a) it is shown that a strong law of large numbers is valid for the elements R_{ij} of the R of (1.6), in the sense that $N^{-1}\ln R_{ij}\to\alpha$ (where α is a constant) with probability 1, provided the elements of the R_n's ($n=1,\ldots,N$) are positive and bounded away from ∞ and 0 in an appropriate sense. In Ref. 7(b) a law of large numbers is shown in connection with matrices belonging to a noncompact semisimple Lie group. In Ref. 8 a CLT is proved for the amplitudes of plane waves traveling in a semi-infinite isotopically disordered harmonic chain; the theorem is then applied to the problem of heat conduction in disordered harmonic chains.

Despite of the extensive literature in connection with this problem, we thought that the simplicity and appeal of the spectral analysis[4] employed here, as well as the direct applicability to physical problems[10] of some of the *explicit* results that we obtain, made it worthwhile to publish the present article.

We thought that it would be easier to first study a CLT in the case of compact Lie groups. The groups $R(2)$ and $R(3)$ are studied in detail in Secs. II and III, respectively, following a line of thought that parallels very closely the one given above for T_1, Eqs. (1.2)–(1.5). The analysis is actually generalizable to other compact groups, and this is indicated at the end of Sec. III.

In Sec. IV we make some considerations for the ordinary CLT. The study of a CLT on SU(1,1), the main contribution of the present paper, is carried out in Sec. V, again in close parallel to that on T_1 outlined above. This is the only nontrivial noncompact group that we have studied so far.

Finally, we give in Sec. VI a summary of the results and some perspectives for future investigations.

II. THE CLT AND THE GROUP $R(2)$

The group elements are of the form

$$R(\phi)=\begin{pmatrix}\cos\phi & \sin\phi\\ -\sin\phi & \cos\phi\end{pmatrix}. \tag{2.1}$$

We construct the product of N such group elements $R_i=R(\phi_i)$:

$$R=R_1R_2\cdots R_N\,. \tag{2.2}$$

Suppose that $\phi_1,\phi_2,\ldots,\phi_N$ are statistically independent, all distributed with the same differential probability $dP_1(\phi_i)=p_1(\phi_i)d\phi_i/2\pi$, where the probability density $p_1(\phi_i)$ is a non-negative function defined on the unit circle. We consider the resultant R as defined by (2.2) (*without the normalization factor* $1/\sqrt{N}$ of the familiar CLT) and inquire about the resulting probability density $p_N(\phi)$ in the limit $N\to\infty$.

We can perform a Fourier decomposition of $p_1(\phi)$ and $p_N(\phi)$ in terms of the unitary irreducible representations $e^{im\phi}$ of $R(2)$:

$$p_1(\phi)=\sum_{m=-\infty}^{\infty}e^{im\phi}\langle e^{-im\phi}\rangle_1\,, \tag{2.3a}$$

$$p_N(\phi)=\sum_{m=-\infty}^{\infty}e^{im\phi}\,\langle e^{-im\phi}\rangle_N\,, \tag{2.3b}$$

where

$$\langle e^{-im\phi}\rangle_1=\int e^{-im\phi}p_1(\phi)\,\frac{d\phi}{2\pi}\,, \tag{2.4a}$$

$$\langle e^{-im\phi}\rangle_N=\int e^{-im\phi}p_N(\phi)\,\frac{d\phi}{2\pi}\,. \tag{2.4b}$$

If we know $\langle e^{-im\phi}\rangle_N$, we can thus evaluate $p_N(\phi)$. Similarly to (1.3) we can write

$$e^{im\phi}=e^{im\phi_1}\cdots e^{im\phi_N}\,. \tag{2.5}$$

Averaging both sides of (2.5) we obtain, just as in (1.4),

$$\langle e^{im\phi}\rangle_N=\langle e^{im\phi}\rangle_1^N\,. \tag{2.6}$$

From Appendix A we have the strict inequality

$$|\langle e^{im\phi}\rangle_1|<1\,,\quad m\neq 0\,. \tag{2.7}$$

Thus, for $m\neq 0$, $\langle e^{im\phi}\rangle_1^N\to 0$ as $N\to\infty$. Only $m=0$ survives, giving

$$\langle e^{im\phi}\rangle_N\underset{N\to\infty}{\to}\delta_{m0}\,. \tag{2.8}$$

The resulting probability density $p_N(\phi)$ is then

$$p_N(\phi)\underset{N\to\infty}{\to}1\,, \tag{2.9}$$

regardless of the individual distribution $p_1(\phi_i)$.

The differential probability

$$dP_N=d\phi/2\pi \tag{2.10}$$

is thus proportional to the *invariant* or *Haar*'s measure of $R(2)$, in agreement with the results of Refs. 5 and 6.

Since the group $R(2)$ is compact, repeated convolutions of $p_1(\phi_i)$ with itself thus tend to fill the whole space available, with a resulting probability density that is proportional to the invariant measure of the group, regardless of the initial $p_1(\phi_i)$.

We would like to describe an alternative point of view that emphasizes the idea of convolution.

Suppose that we have convoluted $p_1(\phi)$ with itself N times, giving a result $p_N(\phi)$. If we convolute once more, we have

$$p_{N+1}(\phi)=\int p_1(\phi_1)p_N(\phi-\phi_1)\,\frac{d\phi_1}{2\pi}\,. \tag{2.11}$$

Suppose that $p_N(\phi)$ indeed approaches a limit as $N\to\infty$, i.e.,

$$\lim_{N\to\infty}p_N(\phi)=p(\phi)\,. \tag{2.12}$$

Then Eq. (2.11) gives

$$p(\phi)=\int p_1(\phi_1)p(\phi-\phi_1)\,\frac{d\phi_1}{2\pi}\,,\quad\text{for all } p_1(\phi_1)\,. \tag{2.13}$$

This integral equation has to be satisfied by the limiting form $p(\phi)$ in question, independently of $p_1(\phi_1)$. Result (2.9), i.e., $p(\phi)=1$, is certainly such a solution, because $\int p_1d\phi_1/2\pi=1$. To convince ourselves that this is the only solution, we Fourier transform both sides of (2.13):

$$\langle e^{im\phi}\rangle=\langle e^{im\phi}\rangle_1\langle e^{im\phi}\rangle\,. \tag{2.14}$$

If $\langle e^{im\phi}\rangle$ were nonzero for a set of m's, we could divide both sides of (2.14) by $\langle e^{im\phi}\rangle$, with the result that

$\langle e^{im\phi}\rangle_1 = 1$ for all those m's and for arbitrary p_1. But according to (2.7) this can only happen for $m=0$, so that $\langle e^{im\phi}\rangle$ can be nonzero only for $m=0$, i.e.,

$$\langle e^{im\phi}\rangle = \delta_{m0}\,, \tag{2.15}$$

and, from (2.3b), $p(\phi) = \text{const}$.

For a presentation of a CLT in the language of Eq. (2.13), the reader is referred to Ref. 11, where generalizations to dependent variables are studied.

III. THE CLT AND THE GROUP $R(3)$. MORE GENERAL COMPACT GROUPS

We denote by R_i ($i=1,\ldots,N$), N elements of the group $R(3)$ and construct the product

$$R = R_1 R_2 \cdots R_N\,. \tag{3.1}$$

We suppose that $R_1,\ldots,R_N$ are statistically independent and all distributed with the same differential probability, which we write as

$$dP_1(R_i) = p_1(R_i)\,d\mu(R_i)\,. \tag{3.2a}$$

In (3.2a), $p_1(R_i)$ is a non-negative function defined on the group manifold and $d\mu(R_i)$ is the invariant or Haar's measure of $R(3)$.

Again we inquire whether the resulting differential probability

$$dP_N(R) = p_N(R)\,d\mu(R) \tag{3.2b}$$

has a limiting form as $N\to\infty$.

We can perform a "Fourier decomposition" of $p_N(R)$ and $p_1(R)$ in terms of the unitary irreducible representations $D^l_{mm'}(R)$ of $R(3)$:

$$p_1(R) = \sum_{lmm'} D^l_{mm'}(R)\langle D^l_{mm'}\rangle_1^*\,, \tag{3.3a}$$

$$p_N(R) = \sum_{lmm'} D^l_{mm'}(R)\langle D^l_{mm'}\rangle_N^*\,, \tag{3.3b}$$

where

$$\langle D^l_{mm'}\rangle_1 = \int D^l_{mm'}(R)p_1(R)\,d\mu(R)\,, \tag{3.4a}$$

$$\langle D^l_{mm'}\rangle_N = \int D^l_{mm'}(R)p_N(R)\,d\mu(R)\,. \tag{3.4b}$$

The $D^l_{mm'}$ are orthogonal under the measure $d\mu(R)$, which we assume normalized, i.e., $\int d\mu(R) = 1$.

For the l representation we can write, from (3.1),

$$D^l(R) = D^l(R_1)\cdots D^l(R_n)\,. \tag{3.5}$$

We now multiply both sides of (3.5) by $p_1(R_1)d\mu(R_1)\cdots p_1(R_N)d\mu(R_N)$ and integrate. On the rhs we obtain $\langle D^l\rangle_1^N$, using the notation of (3.4a). On the other hand, the lhs can be considered as defining the integral $\int D^l(R)p_N(R)d\mu(R)$ [where $p_N(R)$ is our unknown], which we denote as $\langle D^l\rangle_N$, using the notation of (3.4b). We thus have

$$\langle D^l\rangle_N = \langle D^l\rangle_1^N\,. \tag{3.6}$$

Knowing $\langle D^l\rangle_N$ $\forall\, l$, we can recover $p_N(R)$ through (3.3b).

It is proved in Appendix B that the matrix $\langle D^l\rangle$ is strictly "subunitary" for $l\neq 0$, in the sense that for an arbitrary normalized $(2l+1)$-dimensional vector u, we have

$$u^\dagger\langle D^l\rangle^\dagger\langle D^l\rangle u < 1\,,\quad l\neq 0\,. \tag{3.7}$$

It is also proved that $\langle D\rangle^2$ is even "more subunitary" than $\langle D\rangle$, in the sense that

$$u^\dagger[\langle D^l\rangle^2]^\dagger[\langle D^l\rangle^2]u < u^\dagger\langle D^l\rangle^\dagger\langle D^l\rangle u\,,\quad l\neq 0\,. \tag{3.8}$$

If we choose $u_j = \delta_{ji}$, (3.8) gives

$$\sum_k |[\langle D^l\rangle^2]_{ki}|^2 < \sum_k |\langle D^l\rangle_{ki}|^2\,,\quad l\neq 0\,. \tag{3.9}$$

Thus the quantity $\Sigma_k|[\langle D^l\rangle^N]_{ki}|^2$, $l\neq 0$, decreases with N and tends to zero as $N\to\infty$. Since it is a sum of squares, each term tends to zero, which means

$$\langle D^l\rangle_N = \langle D^l\rangle_1^N \underset{N\to\infty}{\to} 0\,,\quad l\neq 0\,. \tag{3.10}$$

The only l that survives is thus $l=0$, i.e.,

$$\langle D^0\rangle_N = 1\,, \tag{3.11}$$

so that (3.3b) gives the probability density

$$p_N(R) \underset{N\to\infty}{\to} 1\,, \tag{3.12}$$

and hence the differential probability

$$dp_N(R) \underset{N\to\infty}{\to} d\mu(R)\,. \tag{3.13}$$

Just as in the previous section, we reach the conclusion that successive convolutions of a given probability density finally end up with the invariant measure of the group.[5,6]

We would now like to describe the situation in a language that emphasizes the idea of convolution, just as we did for $R(2)$.

The convolution $h(R)$ of two functions $f(R), g(R)$ defined on R, will be taken as

$$h(R) = \int f(R_1)g(R_1^{-1}R)\,d\mu(R_1)\,, \tag{3.14}$$

which is a generalization of the ordinary concept. If D is a representation of the group, one can easily show from (3.14) the "convolution theorem"

$$\langle D\rangle_h = \langle D\rangle_f\,\langle D\rangle_g\,. \tag{3.15}$$

Here

$$\langle D\rangle_f = \int D(R)\,f(R)\,d\mu(R)\,, \tag{3.16}$$

and similarly for $\langle D\rangle_g$ and $\langle D\rangle_h$.

Let us recall Eq. (3.5) for the particular case $N=2$. As we mentioned right after that equation, the lhs of (3.5), multiplied by $p_1(R_1)d\mu(R_1)p_1(R_2)d\mu(R_2)$ and integrated, was taken there as defining $\int D^l(R)p_2(R)d\mu(R)$. We now see that this is equivalent to defining $p_2(R)$ as the convolution (3.14) of p_1 with itself.

Let us thus assume that we have convoluted $p_1(R)$ with itself N times, giving a result $p_N(R)$. If we convolute once more we have

$$p_{N+1}(R) = \int p_1(R_1)p_N(R_1^{-1}R)\,d\mu(R_1)\,. \tag{3.17}$$

Suppose that $p_N(R)$ indeed approaches a limit as $N\to\infty$, i.e.,

$$\lim_{N\to\infty} p_N(R) = p(R) . \tag{3.18}$$

Then Eq. (3.17) gives

$$p(R) = \int p_1(R_1) p(R_1^{-1}R) d\mu(R_1) , \quad \text{for all } p_1(R_1) . \tag{3.19}$$

This integral equation has to be satisfied by the limiting form in question, independently of $p_1(R_1)$. Result (3.12), i.e., $p(R)=1$, is certainly such a solution, because $\int p_1 d\mu(R_1) = 1$.

In this case we have not succeeded in actually finding the solution of Eq. (3.19) by "Fourier transforming" both sides of the equation, i.e., by multiplying (3.19) by $D^l(R)$ and integrating over R, as we did for $R(2)$. If we perform that operation, we obtain

$$\langle D^l\rangle = \langle D^l\rangle_1 \langle D^l\rangle . \tag{3.20}$$

If, for a set of l's, $\langle D^l\rangle$ is nonsingular, then $\langle D^l\rangle_1 = I$ for those l's and for all p_1's. But, according to (3.7), this can only happen for $l=0$, so that $\langle D^l\rangle$ can be nonsingular only for $l=0$. We thus reach the conclusion that

$$p(R) = 1 + \sum_{\substack{l=1\\ m,m'}}^{\infty} D^l_{m,m'}(R) \langle D^l_{m,m'}\rangle^* , \tag{3.21}$$

with $\det\langle D^l_{mm'}\rangle = 0$, $l>1$. From this argument we cannot conclude that $\langle D^l_{mm'}\rangle = 0$, $l>1$.

The above ideas can be generalized to other compact groups, at least the unitary and the orthogonal ones, if l is taken to represent the index of the unitary irreducible representation and m its row. A spectral analysis, as in Eq. (3.3), can be done quite in general. The sum is over all unitary irreducible representations, including the "scalar one" $l=0$, that assigns 1 to every element of the group. One again finds, as in (3.10) and (3.11), that the scalar representation is the only one that survives as $N\to\infty$, and hence the conclusion (3.13) that the differential probability tends to the invariant measure of the group. Equation (3.19) is valid more generally, too. Again we see that $p(R)=1$ is a solution of that equation, but we have not been able to prove that that solution is unique.

IV. THE CLT AND THE GROUP T_1

This is the standard CLT that was outlined in the Introduction. The purpose of this section is to indicate a few considerations that we want to contrast with the analysis presented in Secs. II and III on compact groups and in the next section on the noncompact group SU(1,1).

Calling, as before, $p_N(x)$ the probability density of $x = x_1 + \cdots + x_N$, we first look for the asymptotic form of $p_N(x)$ for $N \gg 1$. We assume $p_1(x_i)$ to be centered at the origin, and use the notation

$$\langle e^{-ikx}\rangle_1 = e^{\psi_1(k)} \tag{4.1}$$

for the complex conjugate of the quantity occurring in (1.5a); $\psi_1(k)$ is the "cumulant generating function" [1] of $p_1(x)$, which, when $\langle x\rangle_1 = 0$, can be written as

$$\psi_1(k) = -\frac{k^2}{2!}\sigma^2 + \frac{ik^3}{3!}\kappa_3 - \cdots , \tag{4.2}$$

where $\sigma^2, \kappa_3, \ldots$, are the second, third,..., cumulants of $p_1(x)$.

The asymptotic form of

$$p_N(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx + N\psi_1(k)} dk , \tag{4.3}$$

for $N \gg 1$ can be easily found using the saddlepoint approximation. A similar procedure will be followed in Sec. V for the group SU(1,1). Calling

$$\omega(k) = ikx + N\psi_1(k) , \tag{4.4}$$

the saddle k_0 satisfies $\omega'(k_0) = 0$, so that

$$\psi_1'(k_0) = -ix/N . \tag{4.5}$$

Suppose $|x| \lesssim O(\sqrt{N})$. Then $|\psi_1'(k_0)| \ll 1$ if $N \gg 1$; from (4.2) we then find

$$k_0 \approx i x / N\sigma^2 , \tag{4.6}$$

so that the saddle point is very close to the origin. The integral through the saddle is then approximately

$$p_N(x) \approx \frac{1}{2\pi} e^{\omega(k_0)} \left[\frac{2\pi}{|\omega''(k_0)|}\right]^{1/2} = \frac{e^{-x^2/2N\sigma^2}}{(2\pi N\sigma^2)^{1/2}} , \tag{4.7}$$

the familiar result.

As we increase the number N of convolutions, the resulting probability density $p_N(x)$ stays centered at the origin, but since T_1 is noncompact, $p_N(x)$ becomes wider and wider and tends to zero at every x as $N\to\infty$. However, if we normalize $p_N(x)$ to its value at the origin, i.e.,

$$q_N(x) \equiv \frac{p_N(x)}{p_N(0)} = e^{-x^2/2N\sigma^2} , \tag{4.8}$$

we obtain a function that in any fixed interval Δx becomes as close to 1 as we please by taking N large enough. The reason for this behavior is that only a small region around the saddle k_0 of (4.6) contributes effectively to the integral, if $N \gg 1$; since $|k_0| \ll 1$, $e^{ik_0 x} \approx$ const gives the dominant contribution. This result, which implies that the differential probability is proportional to dx, which is Haar's measure in this case (in great similarity with the behavior of compact groups studied in the previous sections), is, however, a peculiarity of T_1 and is by no means a general behavior of noncompact groups. In the next section we shall see the corresponding behavior for the group SU(1,1).

We can gain more insight into the above situation by describing it from the point of view of convolutions, just as we did in the previous sections for compact groups. At the $(N+1)$st convolution we have

$$p_{N+1}(x) = \int_{-\infty}^{\infty} p_1(x_1) p_N(x - x_1) dx_1 . \tag{4.9}$$

We divide both sides by $p_N(0)$ and define $q_N(x)$ as in (4.8):

$$\frac{p_{N+1}(0)}{p_N(0)} q_{N+1}(x) = \int_{-\infty}^{\infty} p_1(x_1) q_N(x - x_1) dx_1 . \tag{4.10}$$

Assuming that, as $N\to\infty$, the limits

$$\lim_{N\to\infty} q_N(x) = q(x)\,, \tag{4.11}$$

$$\lim_{N\to\infty} \frac{p_{N+1}(0)}{p_N(0)} = K \tag{4.12}$$

exist, then $q(x)$ must satisfy the integral equation

$$Kq(x) = \int_{-\infty}^{\infty} p_1(x_1) q(x - x_1)\, dx_1\,. \tag{4.13}$$

We know *a posteriori*, from (4.7), that $K = 1$: then $q(x) \equiv 1$ indeed satisfies (4.13) for arbitrary $p_1(x)$. That $K = 1$ and $q(x) \equiv 1$ is the only possibility can indeed be seen by assuming that the limit (4.11) is independent[12] of $p_1(x_1)$. Fourier transforming both sides of (4.13) we have

$$K\langle e^{ikx}\rangle = \langle e^{ikx}\rangle_1 \langle e^{ikx}\rangle\,, \tag{4.14a}$$

where

$$\langle e^{ikx}\rangle = \int q(x) e^{ikx}\, dx\,, \tag{4.14b}$$

$$\langle e^{ikx}\rangle_1 = \int p_1(x) e^{ikx}\, dx\,. \tag{4.14c}$$

If there were a range Δk for which $\langle e^{ikx}\rangle \neq 0$, we would have $K = \langle e^{ikx}\rangle_1$ for that Δk. If p_1 is arbitrary, its Fourier transform does not have to be constant over that Δk. Then $\langle e^{ikx}\rangle$ can only be nonzero at one isolated point k_0, i.e.,

$$\langle e^{ikx}\rangle = q_0 \delta(k - k_0)\,. \tag{4.15}$$

Integrating (4.14) in a small region around k_0 we then have

$$K = \langle e^{ik_0 x}\rangle_1\,. \tag{4.16}$$

[Notice that if (4.15) were replaced by the sum $\Sigma_i q_i \delta(k - k_i)$, $\langle e^{ikx}\rangle_1$ would have again to take the same value K at all those k_i's.] The inverse Fourier transform of (4.15), $q(x)$, is then

$$q(x) = (q_0/2\pi) e^{ik_0 x}\,. \tag{4.17}$$

But $q(x)$ must be real and positive, so that $k_0 = 0$ and thus $K = 1$ and $q(x) = \text{const}$.

Qualitatively it is very clear why $K = 1$: since $p_N(x)$ has a variance $N\sigma^2$ (and hence a width $\sqrt{N\sigma^2}$), $p_N(0) \sim 1/\sqrt{N\sigma^2}$, and

$$K = \lim_{N\to\infty} [(N+1)\sigma^2/N\sigma^2]^{1/2} = 1\,.$$

This result depends very directly on the structure of the group. If the "expansion" of $p_N(x)$ as N grows were different, K might differ from 1, and hence $q = \text{const}$ would *not* be a solution of (4.13). This will indeed be the case for the group SU(1,1) to be studied in the next section.

Let us now go back to the result (4.7). That result implies immediately that, if instead of x we consider the renormalized variable

$$\xi = x/\sqrt{N}\,, \tag{4.18}$$

its probability density, as $N \to \infty$, is

$$p(\xi) = e^{-\xi^2/2\sigma^2}/\sqrt{2\pi\sigma^2}\,. \tag{4.19}$$

The standard proof of (4.19) uses the fact that the Fourier transform of $p(\xi)$ can be obtained from that of $p_N(x)$ by replacing k by $k/\sqrt{N}$; (4.2) then implies

$$\exp N\psi(k/\sqrt{N}) \underset{N\to\infty}{\to} -k^2\sigma^2/2\,,$$

whose inverse Fourier transform is (4.19).

V. THE CLT AND THE GROUP SU(1,1)

We present in Appendix C an outline of the properties of the group SU(1,1) that will be needed in what follows. We thus refer the reader to that appendix for the relevant concepts and for the notation employed in this section.

The CLT problem can again be posed in the language used in Sec. III, Eqs. (3.1) and (3.2). Now R_i ($i = 1,...,N$) and R denote elements of SU(1,1) [Eqs. (C2)] and $d\mu(R)$ the invariant measure of Eqs. (C12) and (C15).

Just as we did in Eqs. (3.3), using (C9) and (C28) we can now spectral analyze $p_1(R)$ and $p_N(R)$ in terms of the unitary irreducible representations $D^k_{mm'}(R)$ of SU(1,1):

$$p_1(R) = \sum_{D^\pm} \sum_{\substack{k=1,3/2,...\\ mm'}} D^k_{mm'}(R) \langle D^k_{mm'}\rangle_1^* w_k + \sum_{C^{0,1/2}} \sum_{mm'} \int_0^\infty D^{1/2+is}_{mm'}(R) \langle D^{1/2+is}\rangle_1^* w(s) ds\,, \tag{5.1a}$$

$$p_N(R) = \sum_{D^\pm} \sum_{\substack{k=1,3/2,...\\ mm'}} D^k_{mm'}(R) \langle D^k_{mm'}\rangle_N^* w_k + \sum_{C^{0,1/2}} \sum_{mm'} \int_0^\infty D^{1/2+is}_{mm'}(R) \langle D^{1/2+is}\rangle_N^* w(s) ds\,, \tag{5.1b}$$

where we have used the symbols

$$\langle D^k_{mm'}\rangle_1 = \int D^k_{mm'}(R) p_1(R) d\mu(R)\,, \tag{5.2a}$$

$$\langle D^k_{mm'}\rangle_N = \int D^k_{mm'}(R) p_N(R) d\mu(R)\,, \tag{5.2b}$$

which play the role of the "characteristic function" in the usual probability theory.[1]

As noted in Appendix C, below Eq. (C8), these expansions do not contain the trivial one-dimensional unitary representation that assigns 1 to every group element.

For the kth representation we have the equivalent of Eqs. (3.5) and (3.6), i.e.,

$$D^k(R) = D^k(R_1) \cdots D^k(R_N)\,, \tag{5.3}$$

$$\langle D^k\rangle_N = \langle D^k\rangle_1^N\,. \tag{5.4}$$

Introducing (5.4) in (5.1b) we have, in principle, the exact answer for $p_N(R)$, for any given $p_1(R)$.

We have been able to study in detail the case in which the initial probability density $p_1(R)$ is *isotropic*, i.e., independent of the angles μ, ν of Eq. (C10), i.e.,

$$p_1(R) = p_1(\rho)\,. \tag{5.5}$$

This is the case that we analyze in what follows.

Due to the structure (C16) of $D^k_{mm'}$ and the fact that $-\pi < \mu$, $\nu < \pi$, an isotropic $p_1(R)$ implies that the expectation values $\langle D^k_{mm'}\rangle_1$ of (5.2a) are all zero, except when $m = m' = 0$. From the classification given in (C4)–(C8) we thus see that the only representations that give a nonzero

$\langle D^k_{mm'}\rangle_1$ are those that belong to the continuous class, for the integral case C^0_k (only the principal interval $k=\frac{1}{2}+is$ is relevant): for them, the only nonzero matrix element is the $m=0$, $m'=0$ one and, from (5.4), its Nth power is the only nonzero matrix element of $\langle D^k_{mm'}\rangle_N$. All the other representations give identically zero matrices for $\langle D^k_{mm'}\rangle_1$ and $\langle D^k_{mm'}\rangle_N$. In conclusion, for the only relevant representations $k=\frac{1}{2}+s$ we have

$$\langle D^{1/2+is}_{mm'}\rangle_1 = \delta_{m0}\delta_{m'0}\langle D^{1/2+is}_{00}\rangle_1 , \tag{5.6a}$$

$$\langle D^{1/2+is}_{mm'}\rangle_N = \delta_{m0}\delta_{m'0}\langle D^{1/2+is}_{00}\rangle_1^N , \tag{5.6b}$$

showing that if $p_1(R)$ is isotropic [Eq. (5.5)], the resulting $p_N(R)$ also is [see also Appendix C, Eq. (C39)].

From (C16)–(C18) we write the relevant representation as

$$D^{1/2+is}_{00}(\mu,\rho,\nu) = {}_2F_1(\tfrac{1}{2}+is,\tfrac{1}{2}-is;1;-\rho)\equiv f_s(\rho) , \tag{5.7a}$$

and we now indicate some of its properties.

The function $f_s(\rho)$ satisfies the differential equation[3]

$$\frac{\partial}{\partial\rho}\left[\rho(1+\rho)\frac{\partial f_s(\rho)}{\partial\rho}\right] + \left(s^2+\frac{1}{4}\right)f_s(\rho)=0 . \tag{5.7b}$$

Due to the symmetry of the hypergeometric function ${}_2F_1$ under the interchange of its first two indices, it is clear from (5.7a) that $f_s(\rho)$ is real. Its series expansion in powers of ρ (for $\rho<1$) is[13]

$$f_s(\rho) = 1 - [s^2+(\tfrac{1}{2})^2]\rho + [s^2+(\tfrac{1}{2})^2][s^2+(\tfrac{3}{2})^2] \times\rho^2/(2!)^2 - \cdots . \tag{5.8}$$

The above series, or one of the integral representations of ${}_2F_1$ valid for all ρ, shows that $f_s(\rho)$ is actually a function of s^2. For $s=0$ one can show that

$$f_0(\rho) = [(2/\pi)(1+\rho)^{-1/2}]\,K(\rho/(1+\rho)), \tag{5.9a}$$

where K is the elliptic integral defined in Ref. 13, so that $f_0(\rho)$ is always positive. The asymptotic behavior of $f_0(\rho)$ for $\rho\gg 1$ is

$$f_0(\rho)\sim(\ln\rho)/(\pi\rho^{1/2}) ,\quad \rho\gg 1 . \tag{5.9b}$$

From the asymptotic form (C19) we see that for $s\neq 0$ $f_s(\rho)$ oscillates around the value zero, with a wavelength that decreases as s increases. The behavior of $f_s(\rho)$ has been plotted for various values of s in Fig. 1.

In terms of the functions $f_s(\rho)$ we can write the expansion (5.1) as

$$p_1(\rho) = \int_0^\infty f_s(\rho)\langle f_s\rangle_1 w(s)\,ds , \tag{5.10a}$$

$$p_N(\rho) = \int_0^\infty f_s(\rho)\langle f_s\rangle_N w(s)\,ds , \tag{5.10b}$$

with $w(s)$ given by (C21b), i.e.,

$$w(s) = 2s\tanh\pi s , \tag{5.11}$$

and $\langle f_s\rangle_N = \langle f_s\rangle_1^N$, where

$$\langle f_s\rangle_1 = \int_0^\infty f_s(\rho)p_1(\rho)d\rho \equiv F_1(s)\equiv e^{\psi_1(s)} , \tag{5.12a}$$

$$\langle f_s\rangle_N = \int_0^\infty f_s(\rho)p_N(\rho)d\rho \equiv F_N(s)\equiv e^{\psi_N(s)} . \tag{5.12b}$$

The analysis that follows parallels very closely the one given in the last section for the usual CLT, starting with Eq. (4.3), which has its equivalent, for SU(1,1), in Eq. (5.10b).

Before proceeding, though, we shall use the functions $F(s)$ and $\psi(s)$ defined as in Eq. (5.12) for an arbitrary $p(\rho)$, to define the equivalent of moments and cumulants of the usual probability theory.[1]

It will prove convenient to expand F and ψ in terms of

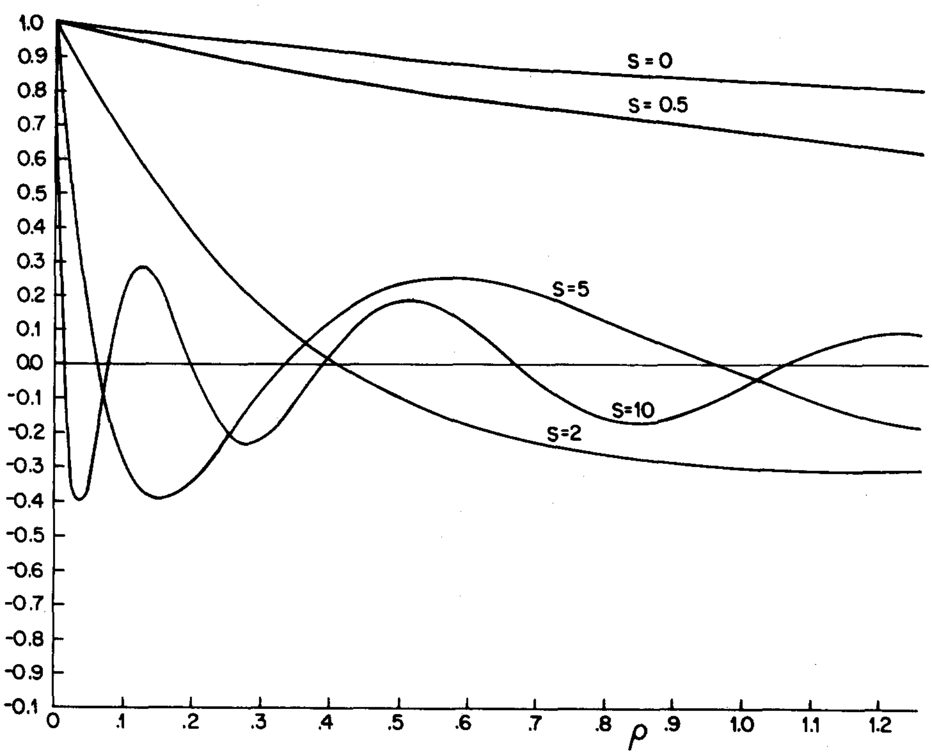


FIG. 1. The function $f_s(\rho)$ of Eq. (5.7a) as a function of ρ, for $s=0, 0.5, 2, 5, 10$.

the variable q [defined in Eq. (C3)], which, in the present case, is given by

$$q = s^2 + \tfrac{1}{4}. \tag{5.13}$$

We notice from (5.8) that

$$F(q=0) = 1, \quad \psi(q=0) = 0. \tag{5.14}$$

We can reexpress the expansion (5.8) of $f_s(\rho)$ as a power series in q, to get

$$f_q(\rho) = \sum_{n=0}^{\infty} \phi_n(\rho) \frac{q^n}{n!}, \tag{5.15}$$

where, by using the procedure of Appendix F, one can show

$$\phi_0(\rho) = 1, \tag{5.16a}$$

$$\phi_1(\rho) = -\ln(1+\rho), \tag{5.16b}$$

$$\phi_2(\rho) = 2\int_0^{\rho} \frac{\ln(1+\rho')}{\rho'} d\rho' - 2\ln(1+\rho), \tag{5.16c}$$

$$\vdots \quad .$$

Averaging both sides of (5.15) with $p(\rho)$ we can thus write

$$F(q) = \sum_{n=0}^{\infty} \mu_n \frac{q^n}{n!}, \tag{5.15$'$}$$

where the "moments" μ_n are defined as

$$\mu_n = \langle \phi_n(\rho) \rangle. \tag{5.17}$$

We can also expand $\ln F$ as

$$\psi(q) = \ln F(q) = \sum_{n=1}^{\infty} \kappa_n \frac{q^n}{n!}, \tag{5.18}$$

where the "cumulants" κ_n are defined as

$$\kappa_1 = \mu_1, \tag{5.19a}$$

$$\kappa_2 = \mu_2 - \mu_1^2, \tag{5.19b}$$

$$\vdots \quad .$$

The cumulants κ_n of Eq. (5.19) *add under the convolution of independent variables.* In particular, since $\kappa_1 = -\langle \ln(1+\rho) \rangle = -\langle z \rangle$ [see definition (C13b)], we have, for the probability densities $p_N(\rho)$ and $p_1(\rho)$ of Eq. (5.10),

$$\langle z \rangle_N = N \langle z \rangle_1. \tag{5.20a}$$

The average of the variable z thus scales with N. The quantity $\langle z \rangle_N$ is usually written in the literature as[14]

$$\langle z \rangle_N = \langle -\ln T \rangle = 2L/L_c, \tag{5.20b}$$

where L is the length of the conductor and L_c its localization length.

A. The asymptotic form of $p_N(\rho)$ for $p_1(\rho)$ fixed and $N \gg 1$

Since $p_1(\rho)$ is kept fixed, $\langle z \rangle_1$ [see definition (C13b)] stays fixed, so that, from (5.20), $L/L_c \gg 1$.

We write $p_N(\rho)$ of (5.10b) as

$$p_N(\rho) = \frac{1}{2}\int_{-\infty}^{\infty} f_s(\rho)\, e^{N\psi_1(s)} w(s) ds, \tag{5.21}$$

where we have used (5.12a) and the fact that the integrand is an even function of s.

We consider two cases.

(a) ρ fixed, independent of N: Since $N \gg 1$, the exponential in (5.21) varies rapidly with s, while the other factors vary relatively smoothly. The integral is done in Appendix D by the saddle-point approximation.

The saddle s_0 is real and close to the origin ($s_0 \sim 1/N$) and the resulting $p_N(\rho)$ is

$$p_N(\rho) = \langle f_0 \rangle_1^N \left(\frac{\pi \langle f_0 \rangle_1}{N |\langle g \rangle_1|} \right)^{3/2} f_0(\rho), \tag{5.22}$$

where g is defined, in an expansion of $f_s(\rho)$ around $s=0$, as

$$f_s(\rho) = f_0(\rho) + g(\rho) s^2 + \cdots. \tag{5.23}$$

Of course, (5.22) is not valid for arbitrarily large ρ: $p_N(\rho)$ eventually decays below the value given by (5.22)—in order that $p_N(\rho)$ be normalizable [see also Eq. (5.9b)]—and becomes wider and wider, thus approaching (5.22), as N increases. It is thus clear why, for all ρ, $p_N(\rho)$ of (5.22) tends to zero as $N \to \infty$ (notice that $f_0 < 1$, from Fig. 1). The decrease to zero of (5.22) is exponential in N. However, if we normalize $p_N(\rho)$ to its value at the origin, we obtain a function $q_N(\rho)$ that in any fixed interval $\Delta\rho$ becomes as close to $f_0(\rho)$ as we please by taking N large enough; i.e.,

$$q_N(\rho) \equiv p_N(\rho)/p_N(0) \underset{N\to\infty}{\to} f_0(\rho). \tag{5.24}$$

The reason for this behavior is that for $N \gg 1$ only a small region around the saddle s_0 contributes effectively to the integral (5.21); since $|s_0| \ll 1$, $f_0(\rho)$ gives the dominant contribution. Now we do not obtain a constant for $q_N(\rho)$ as in the case of T_1 (Sec. IV) because, as we noted right after Eq. (5.2), the one-dimensional representation that assigns 1 to every group element does not enter in the harmonic expansion. What is common to the cases of T_1 and SU(1,1) is that it is the "smoothest" of the functions forming the complete set that dominates the behavior of q_N in a fixed region, in the limit $N \to \infty$.

Again, we can gain more insight by describing the above situation in terms of convolutions.

The convolution of two functions $f(R)$ and $g(R)$ defined on SU(1,1) is given by

$$h(R) \equiv \int f(R_1) g(R_1^{-1} R) d\mu(R_1) \equiv f * g, \tag{5.25}$$

just as in (3.14) for $R(3)$. If the functions involved only depend on ρ, the explicit expression for the convolution is given in Eq. (C39). In the present case, suppose that we have convoluted $p_1(\rho)$ with itself N times, giving $p_N(\rho)$. If we convolute once more we have

$$p_{N+1}(\rho) = p_1(\rho) * p_N(\rho). \tag{5.26}$$

Using the definition of $q_N(\rho)$ given in (5.24) we can write

$$[p_{N+1}(0)/p_N(0)] q_{N+1}(\rho) = p_1(\rho) * q_N(\rho). \tag{5.27}$$

Assuming that, as $N \to \infty$, the following limits exist:

$$\lim_{N\to\infty} q_N(\rho) = q(\rho), \tag{5.28}$$

$$\lim_{N\to\infty} [p_{N+1}(0)/p_N(0)] = K, \tag{5.29}$$

$q(\rho)$ must satisfy the integral equation

$$Kq(\rho) = p_1(\rho) * q(\rho) . \tag{5.30}$$

If K *were* equal to 1, $q(\rho) \equiv 1$ would be a solution of Eq. (5.30), as can be seen from the convolution expression (C39). We know *a posteriori*, from (5.22), the value of K, i.e.,

$$K = \langle f_0 \rangle_1 , \tag{5.31}$$

and since $0 < f_0(\rho) < 1$ for all $\rho > 0$ (see Fig. 1), $0 < K < 1$. Therefore, $q(\rho) \equiv 1$ is *not* a solution of Eq. (5.30).

Result (5.31) for K, as well as the actual form of $q(\rho)$, can actually be found just by assuming that the limit (5.28) is independent[12] of $p_1(\rho)$. Using the "convolution theorem" of Eq. (C42), we can write (5.30) as

$$K \langle f_s \rangle = \langle f_s \rangle_1 \langle f_s \rangle , \tag{5.32a}$$

where

$$\langle f_s \rangle = \int_0^\infty f_s(\rho) q(\rho) d\rho , \tag{5.32b}$$

$$\langle f_s \rangle_1 = \int_0^\infty f_s(\rho) p_1(\rho) d\rho . \tag{5.32c}$$

An argument similar to the one below Eq. (4.14) shows that $\langle f_s \rangle$ must be of the form

$$\langle f_s \rangle = q_0 w^{-1}(s) \delta(s - s_0) , \tag{5.33}$$

where the factor in front of the δ function was introduced for convenience. Multiplying (5.32a) by $w(s)ds$ and integrating over s in a small region around s_0 we then have

$$K = \langle f_{s_0} \rangle_1 . \tag{5.34}$$

We can now calculate $q(\rho)$ knowing the coefficients (5.33) in an expansion in terms of $f_s(\rho)$, as

$$q(\rho) = \int_0^\infty f_s(\rho) \langle f_s \rangle w(s) ds = q_0 f_{s_0}(\rho) . \tag{5.35}$$

From Fig. 1 we see that the only $f_s(\rho)$ that is positive for all ρ is $f_0(\rho)$. Therefore, from (5.34) and (5.35) we have

$$K = \langle f_0 \rangle_1 , \tag{5.36}$$

$$q(\rho) \propto f_0(\rho) , \tag{5.37}$$

just as in (5.31) and (5.24), respectively.

We would now like to verify in a simple example that (5.36) and (5.37) are indeed solutions of (5.30). Suppose we take $p_1(\rho) = \delta(\rho - a)$. This is not a trivial case at all, as it would be for the normal CLT, because now we still have the angles μ, ν, in which $p_1(\rho) d\mu(R)$ is isotropic. Indeed, the repeated convolution of p_1 with itself [Eq. (C39)] is a nontrivial operation. Let us substitute, in Eq. (5.30), $p_1(\rho) = \delta(\rho - a)$, $q(\rho) = f_0(\rho)$, $K = f_0(a)$, making use of (C39). We obtain

$$f_0(a) f_0(\rho) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_0 [\rho + a + 2a\rho - 2\sqrt{a(a+1)}\,\sqrt{\rho(\rho+1)} \cos 2\mu) \, d\mu , \tag{5.38}$$

which, from (C45), is indeed an identity.

(b) $1 \ll \ln \rho \sim N$: In the integral (5.21) we shall employ the asymptotic form (for $\rho \gg 1$) of $f_s(p)$ given in (C19b) and (C19c) that we reproduce here

$$f_s(\rho) = 2 \, \mathrm{Re}[g(s) e^{is \ln \rho} / \sqrt{\rho}] , \quad \rho \to \infty , \tag{5.39a}$$

where

$$g(s) = \Gamma(2is) / [\Gamma(\tfrac{1}{2} + is)]^2 . \tag{5.39b}$$

Again, $w(s)$ is given by (5.11). Since $\ln \rho \sim N$, the integrand in (5.21) oscillates rapidly as a function of s, so that we shall do the integral by the saddle-point approximation. The details are shown in Appendix E. Here we reproduce only a few salient features.

Since we want $\ln \rho \sim N \gg 1$, we write

$$z = aN + x , \tag{5.40}$$

where z is related to ρ as in (C13), a is fixed and $x \ll N$.

From (5.21) and (5.39) we can write the probability density of x as

$$q_N(x) = \frac{1}{2} e^{(aN+x)/2} \, \mathrm{Re} \int_{-\infty}^{\infty} e^{\omega(s)} \, ds . \tag{5.41}$$

The exponent $\omega(s)$ in (5.41) is defined as

$$\omega(s) = \delta(s) + is(aN + x) + N\psi_1(s) , \tag{5.42}$$

where

$$e^{\delta(s)} = \Gamma(1 + is) \tanh \pi s / [\Gamma(\tfrac{1}{2} + is) i \sqrt{\pi}] \tag{5.43}$$

and $\psi_1(s)$ is given in (5.12a).

The saddle point s_0 is given by the solution to the equation

$$\omega'(s_0) = 0 , \tag{5.44a}$$

or

$$\psi_1'(s_0) = -ia - (\delta'(s_0) + ix)/N . \tag{5.44b}$$

Equation (5.44b) can be solved iteratively starting from the approximate solution s_a defined as

$$\psi_1'(s_a) = -ia . \tag{5.45}$$

A few iterations are shown in Eq. (E1).

In the complex s plane one deforms the path of integration, which is along the real s axis in Eq. (5.41), in order to pass through the saddle. It is easy to show that $e^{\delta(s)}$ of (5.43) is an entire function of s. We also assume that $p_1(\rho)$ is such that the resulting $\psi_1(s)$ of Eq. (5.12a) is analytic in the region enclosed between the two paths of integration. We shall see later on examples in which this assumption holds. We can write (5.41) as

$$q_N(x) = \tfrac{1}{2} e^{(aN+x)/2} \, \mathrm{Re}(e^{\omega(s_0)} [2\pi / (-\omega''(s_0))]^{1/2}) , \tag{5.46}$$

which takes the form given in Eq. (E7), once we substitute the values of $\omega(s_0)$ and $\omega''(s_0)$. The choice

$$s_a = i/2 + \delta'(s_a) / N\psi_1''(s_a) \tag{5.47}$$

eliminates the term linear in x in the exponent of (E7), so that the result is a *zero-centered Gaussian distribution* for x. For the variable z [Eq. (5.40)], we thus get the probability density

$$q_N(z) = e^{-(z - aN)^2/2bN} / \sqrt{2\pi b N} , \tag{5.48}$$

a Gaussian, regardless of the shape of the original distribution $p_1(\rho)$. The quantities a and b in (5.48) are given by

$$a = i \left(\frac{\partial \psi_1(s)}{\partial s} \right)_{s = i/2} , \tag{5.49}$$

$$b=-\left(\frac{\partial^2\psi_1(s)}{\partial s^2}\right)_{s=i/2}. \qquad (5.50)$$

In Appendix F it is proved that $\psi_1''(s_a)$ is real and negative. In that appendix, $\psi_1'(s_a)$ and $\psi_1''(s_a)$ are evaluated explicitly in terms of the distribution $p_1(\rho)$ [Eqs. (F18) and (F19)], with the result

$$a=\langle\ln(1+\rho)\rangle_1=\langle z\rangle_1, \qquad (5.51)$$

$$b=2\left\langle\int_0^\rho\frac{\ln(1+\rho')}{\rho'}d\rho'\right\rangle_1-\langle\ln(1+\rho)\rangle_1^2. \qquad (5.52)$$

Equation (5.51) is a particular case of (5.20a).

From (5.48) we find that the probability density of the variable ρ is given by

$$p_N(\rho)=(e^{-a^2N/2b}/2\sqrt{2\pi bN})\,\rho^{(a/2b-1)}e^{-\ln^2\rho/8bN}. \qquad (5.53)$$

B. The limiting form of $p_N(\rho)$ when ρ is rescaled and $N\to\infty$

In the above subsection, $p_1(\rho)$ was kept fixed and an approximate, asymptotic form of $p_N(\rho)$ was found for $N\gg 1$.

In what follows we rescale the variable ρ occurring in $p_1(\rho)$ as

$$\rho=\rho'/N, \qquad (5.54)$$

keeping the distribution for ρ' independent of N. In particular, we define the centroid of the latter distribution as

$$\langle\rho'\rangle_1=l. \qquad (5.55)$$

Using the series expansion (5.8) we can express the $\langle f_s\rangle_1^N$ of (5.12) as

$$\langle f_s\rangle_1=1-\left(s^2+\frac{1}{4}\right)\frac{l}{N}+\left(s^2+\frac{1}{4}\right)\left(s^2+\frac{9}{4}\right)\frac{\langle\rho'^2\rangle}{4N^2}-\cdots. \qquad (5.56)$$

In the limit $N\to\infty$, the quantity $\langle f_s\rangle_N=\langle f_s\rangle_1^N$ occurring in (5.10b) then takes the form

$$\langle f_s\rangle_1^N=\left[1-\left(s^2+\frac{1}{4}\right)\frac{l}{N}+O\left(\frac{1}{N^2}\right)\right]^N\underset{N\to\infty}{\to}e^{-(s^2+1/4)l}. \qquad (5.57)$$

The resulting probability density of (5.10b) is then

$$p_l(\rho)=\int_0^\infty f_s(\rho)e^{-(s^2+1/4)l}w(s)\,ds. \qquad (5.58)$$

The present limit thus gives a *universal form*, valid for all ρ. The original distribution p_1 enters only through the quantity l of (5.55): $p_1(\rho)$ is insensitive to other characteristics of p_1.

In the language of Eqs. (5.13)–(5.20) we can write for $\psi_l(q)$, in the present case

$$\psi_l(q)=-lq, \qquad (5.59)$$

corresponding to the cumulants

$$\kappa_1=-l, \qquad (5.60a)$$

$$\kappa_n=0,\quad n>1, \qquad (5.60b)$$

or, more explicitly, from (5.16) and (5.19) [see also (C13b)],

$$\langle z\rangle_l=\langle\ln(1+\rho)\rangle_l=\langle-\ln T\rangle_l=l, \qquad (5.61a)$$

$$2\left\langle\int_0^\rho\frac{\ln(1+\rho')}{\rho'}d\rho'\right\rangle_l-2\langle\ln(1+\rho)\rangle_l-\langle\ln(1+\rho)\rangle_l^2=0, \qquad (5.61b)$$

where $\langle\ \rangle_l$ indicates an average with the probability density $p_l(\rho)$ of Eq. (5.58). In the present case we thus investigate the mathematical limit in which the number of scatterers grows and at the same time each one becomes weaker, so that the resulting $\langle-\ln T\rangle=2L/L_c$ has the fixed value l, which is now arbitrary, i.e., $0<l<\infty$.

One can find an asymptotic expression for the $p_l(\rho)$ of (5.58) valid for $l\gg 1$ and $1\ll\ln\rho\sim l$, using the saddle-point method, just as in the previous subsection. In terms of the variable z we get the probability density

$$q_l(z)=e^{-(z-l)^2/4l}/\sqrt{4\pi l}, \qquad (5.62)$$

so that

$$\langle z\rangle_l=l, \qquad (5.63)$$

$$(\mathrm{var}\,z)_l=2l. \qquad (5.64)$$

Equation (5.63) is just a particular case of the relation (5.61a), which is valid for all l.

Since the function $f_s(\rho)$ satisfies the differential equation (5.7b), one can easily see that the probability density (5.58) satisfies the "diffusionlike" equation

$$\frac{\partial p_l(\rho)}{\partial l}=\frac{\partial}{\partial\rho}\left[\rho(1+\rho)\frac{\partial p_l(\rho)}{\partial\rho}\right]. \qquad (5.65)$$

The initial condition $p_0(\rho)$ is determined by the integral representation (5.58) as

$$p_0(\rho)=\int_0^\infty f_s(\rho)w(s)ds. \qquad (5.66)$$

Equation (5.66) is the expansion of $p_0(\rho)$ in terms of the complete set $f_s(\rho)$, the expansion coefficients being 1. But this means that

$$\int_0^\infty p_0(\rho)f_s(\rho)d\rho=1. \qquad (5.67)$$

We know that $p_0(\rho)=\delta(\rho)$ [i.e., the "one-sided" δ function, such that $\int_0^\infty\delta(\rho)d\rho=1$] satisfies (5.67). From the uniqueness of the expansion we thus have

$$p_0(\rho)=\delta(\rho). \qquad (5.68)$$

As l increases from zero, $p_l(\rho)$ then "diffuses" according to the differential equation (5.65). We notice the very interesting fact that (5.65) is the same differential equation found in Ref. 15, when the potential felt by the electron is related to a Gaussian random process.

From (5.65) we can easily find the form of the solution for $l\ll 1$, i.e., when $p_l(\rho)$ is still concentrated near $\rho=0$. We can then approximate (5.65) as

$$\frac{\partial p_l}{\partial l}=\frac{\partial}{\partial\rho}\rho\frac{\partial p_l}{\partial\rho}. \qquad (5.69)$$

We can easily check that

$$p_l(\rho)=(1/l)e^{-\rho/l} \qquad (5.70)$$

is the solution of (5.69) that tends to $\delta(\rho)$ as $l\to 0$; it is therefore the approximate form of $p_l(\rho)$ for $l\ll 1$.

If we integrate over ρ both sides of (5.65) we obtain

$$\frac{d}{dl}\int_0^\infty p_l(\rho)d\rho = \left[\rho(1+\rho)\frac{\partial p_l}{\partial\rho}\right]_0^\infty = 0\,, \qquad (5.71)$$

so that the normalization of $p_l(\rho)$ is conserved as l increases.

Just as in Ref. 15, we can multiply Eq. (5.65) by ρ^n and integrate over ρ, to find a recursion relation for the moments $\langle\rho^n\rangle_l$. In particular one finds

$$\langle\rho\rangle_l = \tfrac{1}{2}(e^{2l}-1)\,, \qquad (5.72)$$

$$\langle\rho^2\rangle_l = \tfrac{1}{12}(2e^{6l}-6e^{2l}+4)\,. \qquad (5.73)$$

Equation (5.72) is the well-known exponential increase of the average resistance with the length of the conductor, and Eq. (5.73) indicates that the width of the distribution increases with length more rapidly than the mean.[2,15]

VI. SUMMARY AND CONCLUSIONS

We have analyzed the problem of finding the statistical distribution $dP_N(R) = p_N(R)d\mu(R)$ of the product $R = R_1R_2\cdots R_N$ of N statistically independent elements R_i of a group. The R_i ($i = 1,\ldots,N$) were assumed to be distributed according to the same differential probability $dP_1(R_i) = p_1(R_i)d\mu(R_i)$. Here $d\mu(R)$ is the invariant or Haar's measure of the group in question.

For the compact groups $R(2)$ and $R(3)$ we have been able to show that $p_N(R)\to 1$ as $N\to\infty$, independently of the original $p_1(R_i)$. We made it plausible that a similar behavior is to be expected for other compact groups, at least the unitary and the orthogonal ones.

In connection with noncompact groups, the translation group in one dimension (T_1) gives rise to the usual central-limit theorem. The only nontrivial noncompact group that was analyzed is SU(1,1), which is relevant to the physics of disordered conductors, as was mentioned in the Introduction. We have been able to study in detail the case in which the initial probability density $p_1(R_i)$ is isotropic, i.e., independent of the phases μ,ν of Eq. (C10), i.e., $p_1(R_i) = p_1(\rho_i)$. It was shown that the resulting $p_N(R)$ is also isotropic, i.e., $p_N(R) = p_N(\rho)$, and is given by the exact expression (5.10b). Two limits were studied.

(1) $p_1(\rho)$ is kept fixed and $N\gg 1$. In this case one has $L/L_c \gg 1$, where L is the length of the conductor and L_c the localization length. For ρ fixed, $p_N(\rho)$ is given by (5.22), while for $\ln\rho\sim N$ we obtain a Gaussian distribution in the variable $\ln\rho$ [see Eq. (5.48)].

(2) The variable ρ_i occurring in $p_1(\rho_i)$ is rescaled according to (5.54) and the limit $N\to\infty$ is taken, keeping $\langle -\ln T\rangle = 2L/L_c = l$ fixed. The resulting probability density $p_l(\rho)$ is given by (5.58), which thus reduces the problem to quadratures. It is also shown that $p_l(\rho)$ satisfies the "diffusionlike" equation (5.65) with the initial condition $p_0(\rho) = \delta(\rho)$. From that equation one proves straightforwardly the exponential increase with l of the average value of the resistance.

In a paper to be published elsewhere[10] we generalize the limit (2) above to the case in which the distributions of the various scatterers may be different from one another. That case turns out to be of interest for the description of random conductors placed in an external electric field.

The analysis just described was confined to the isotropic case, i.e., when $p_1(R_i) = p_1(\rho_i)$. The study of the case of a general initial probability density $p_1(R_i)$ would be important in order to complete the description of the problem.

We feel that the analysis of SU(1,1) presented in the present article relies too much upon the specific form of the unitary irreducible representations, a feature that is not needed in the study presented in Secs. II and III in connection with compact groups. One would think that, if the analysis were not tied up to the specific structure of the D's, one should be able to study the general case in a simpler way.

An interesting question about the general case is whether one still has "diffusion" in the full group manifold.

The cumulants κ_n found in Eq. (5.18) for the isotropic case add under the convolution of independent variables, just as normal cumulants do. The relation (5.19) between cumulants and moments is just the standard one; but in Eq. (5.16) we were able to find an explicit expression up to $\phi_2(\rho)$ only: What is the general form of $\phi_n(\rho)$? What is the generalization of these concepts for the nonisotropic problem?

These and other questions (like what happens with other noncompact groups) have to be left for the future.

APPENDIX A: PROOF OF THE INEQUALITY (2.7)

For any probability density $p(\phi)$ we can write

$$\langle e^{im\phi}\rangle = \int e^{im\phi}p(\phi)\frac{d\phi}{2\pi} = \int\left[e^{im\phi}\sqrt{p(\phi)}\,\right]\sqrt{p(\phi)}\,\frac{d\phi}{2\pi}\,. \qquad (A1)$$

We apply Schwartz's inequality to this last integral, the two functions considered by the inequality being

$$f(\phi) = e^{im\phi}\sqrt{p(\phi)} \text{ and } g(\phi) = \sqrt{p(\phi)}\,. \qquad (A2)$$

We have

$$|\langle e^{im\phi}\rangle|^2 < \int |e^{im\phi}\sqrt{p(\phi)}|^2\frac{d\phi}{2\pi}\int|\sqrt{p(\phi)}|^2\frac{d\phi}{2\pi}$$
$$= \left[\int p(\phi)\frac{d\phi}{2\pi}\right]^2 = 1\,. \qquad (A3)$$

The equality sign holds iff

$$f(\phi)\propto g(\phi)\,, \quad \text{for all } \phi\,, \qquad (A4)$$

which can only happen if $m = 0$. Then

$$|\langle e^{im\phi}\rangle| = 1\,, \quad m = 0\,, \qquad (A5a)$$

$$|\langle e^{im\phi}\rangle| < 1\,, \quad m\neq 0\,. \qquad (A5b)$$

APPENDIX B: PROOF OF THE INEQUALITIES (3.7) AND (3.8)

Let u be an arbitrary normalized $(2l+1)$-dimensional vector. Calling

$$u' = \langle D^l\rangle u\,, \qquad (B1)$$

we have

$$u_i' = \int [D^l(R)u]_i\,p(R)d\mu(R)$$
$$= \int\left\{[D^l(R)u]_i\sqrt{p(R)}\right\}\sqrt{p(R)}\,d\mu(R). \qquad (B2)$$

We apply Schwartz's inequality to this last integral, the two functions considered by the inequality being

$$f(R) = [D^l(R)u]_i \sqrt{p(R)} \text{ and } g(R) = \sqrt{p(R)}\,. \qquad \text{(B3)}$$

We have

$$|u_i'|^2 < \int |[D^l(R)u]_i|^2 p(R) d\mu(R) \int p(R) d\mu(R)$$

$$= \int |[D^l(R)u]_i|^2 p(R) d\mu(R)\,, \qquad \text{(B4)}$$

where we have used the normalization of p. Summing over i both sides of (B4) and using the unitarity of D and the fact that the vector u is normalized, and $p(r)$ is normalized, too, we have

$$u'^\dagger u' = \sum_i |u_i'|^2$$

$$< \int \sum_i |[D^l(R)u]_i|^2 p(R) d\mu(R)$$

$$= \int \sum_i |u_i|^2 p(R) d\mu(R) = \int p(R) d\mu(R) = 1\,. \qquad \text{(B5)}$$

The equality sign in Schwartz inequality occurs iff

$$f(R) \propto g(R)\,, \quad \text{for all } R\,, \qquad \text{(B6)}$$

which can only occur for $l = 0$. Then

$$|\langle D^0\rangle| = 1\,, \qquad \text{(B7a)}$$

$$u^\dagger \langle D^l\rangle^\dagger \langle D^l\rangle u < 1\,, \quad l \neq 0\,, \qquad \text{(B7b)}$$

for all normalized u. This proves (3.7).

We can further write

$$u^\dagger[\langle D^l\rangle^2]^\dagger \langle D^l\rangle^2 u$$

$$= \left[\frac{u'}{\sqrt{u'^\dagger u'}} \langle D^l\rangle^\dagger \langle D^l\rangle \frac{u'}{\sqrt{u'^\dagger u'}}\right] u'^\dagger u'\,. \qquad \text{(B8)}$$

If $l \neq 0$, we apply (B7b) to the bracket in (B8), to obtain

$$u^\dagger[\langle D^l\rangle^2]^\dagger \langle D^l\rangle^2 u < u^\dagger \langle D^l\rangle^\dagger \langle D^l\rangle u\,, \qquad \text{(B9)}$$

which is (3.8).

APPENDIX C: REVIEW OF THE GROUP SU(1,1)

We give here an outline of the properties of the group SU(1,1) that are needed in the text.

1. Definition

The group SU(1,1) consists of the collection of all 2×2 matrices R that are pseudounitary and unimodular, i.e.,

$$R^\dagger \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} R = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \det R = 1\,. \qquad \text{(C1)}$$

Now R can be written as

$$R = \begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix}, \qquad \text{(C2a)}$$

with

$$|\alpha|^2 - |\beta|^2 = 1\,. \qquad \text{(C2b)}$$

2. Classification of the unitary irreducible representations

The unitary irreducible representations of SU(1,1) were studied by Bargmann,[3] with the following main results. We call k the index and m the row of the representation. Bargmann also introduces the real Casimir operator eigenvalue

$$q = k(1-k)\,. \qquad \text{(C3)}$$

One has the following classes: (I) continuous class,

(1) integral case (C_k^0):

$$0<q<\infty,\ m = 0, \pm 1, \pm 2, \ldots, \qquad \text{(C4)}$$

(a) exceptional interval:

$$0<k<\tfrac{1}{2},\ \text{i.e., } 0<q<\tfrac{1}{4}\,, \qquad \text{(C5a)}$$

(b) principal interval:

$$[k = \tfrac{1}{2} + is,\ 0<s<\infty,\ \text{i.e., } q>\tfrac{1}{4}]\,; \qquad \text{(C5b)}$$

(2) half-integral case ($C_k^{1/2}$):

$$[k = \tfrac{1}{2} + is,\ 0<s<\infty,\ \text{i.e., } q>\tfrac{1}{4}]\,; \qquad \text{(C6)}$$

(II) discrete class,

$$\text{(1) maximal } m(D_k^-):\ k = \tfrac{1}{2}, [1, \tfrac{3}{2}, 2, \ldots]\,, \quad m = -k, -(k+1), \ldots, \qquad \text{(C7)}$$

$$\text{(2) minimal } m(D_k^+):\ k = \tfrac{1}{2}, [1, \tfrac{3}{2}, 2, \ldots]\,, \quad m = k, k+1, \ldots\,. \qquad \text{(C8)}$$

For $k = 0$ one has the one-dimensional unitary representation that associates 1 to every group element.

The unitary irreducible representations that form a complete set have been enclosed in square brackets in the above scheme. Notice that, in particular, the trivial one-dimensional unitary representation is not a member of the complete set.

3. Spectral analysis

The idea of spectral analysis that was used for compact groups in Secs. II and III and for the translation group in the Introduction can be extended to the group SU(1,1) as well. A function $f(R)$ defined on the group can be spectral-analyzed as

$$f(R) = \sum_{D^\pm} \sum_{\substack{k = 1,3/2,\ldots \\ m,m'}} a_{mm'}^k D_{mm'}^k(R)$$

$$+ \sum_{C^{0,1/2}} \sum_{mm'} \int_0^\infty a_{mm'}^s D_{mm'}^{1/2+is}(R)\, ds\,, \qquad \text{(C9)}$$

where $a_{mm'}^k$, $a_{mm'}^s$ are the expansion coefficients and the sum and the integral run over those unitary irreducible representations $D^\pm$ and $C^{0,1/2}$, respectively, that form a complete set.

4. Parametrizations of SU(1,1)

In the analysis presented in the text we make use of the specific parametrization of SU(1,1) given in terms of the "Euler angles" (μ,ζ,ν) introduced by Bargmann.[3] In terms of them, every group element R can be written as

$$R=\begin{pmatrix} e^{-i\mu} & 0 \\ 0 & e^{i\mu} \end{pmatrix} \begin{pmatrix} \cosh\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix} \begin{pmatrix} e^{-i\nu} & 0 \\ 0 & e^{i\nu} \end{pmatrix}, \tag{C10}$$

where the range of the parameters is specified by

$$0<\zeta<\infty, \quad -\pi<\mu,\nu<\pi. \tag{C11}$$

The invariant measure is given by

$$d\mu(R)=(2\pi)^{-2}\sinh 2\zeta\, d\zeta\, d\mu\, d\nu. \tag{C12}$$

Since the group is noncompact, $d\mu(R)$ is not normalizable.

It is also advantageous to introduce the variables

$$\rho=\sinh^2\zeta, \tag{C13a}$$

$$z=\ln(1+\rho)=-\ln T. \tag{C13b}$$

In the physical applications to disordered conductors that we mentioned in the Introduction, T is the transmission coefficient of the sample and ρ is related to its resistance by[2]

$$\text{resistance}=(\pi\hbar/e^2)\,\rho. \tag{C14}$$

Then ρ is called the "dimensionless resistance." It is really the natural variable of the problem, because in terms of it the invariant measure can be written as

$$d\mu(R)=(2\pi)^{-2}\,d\rho\, d\mu\, d\nu. \tag{C15}$$

5. Explicit form of the unitary irreducible representations

The explicit form of the unitary irreducible representations of SU(1,1) has also been given in Ref. 3. They can be written in a way that resembles those of SU(2). Using the parameters ρ,μ,ν, they take the form

$$D^k_{mm'}(\mu,\rho,\nu)=e^{-2im\mu}d^k_{mm'}(\rho)e^{-2im'\nu}, \tag{C16}$$

where

$$\begin{aligned} d^k_{mm'}(\rho)=\theta_{mm'}(k)&\left[\rho^{|m-m'|/2}/(1+\rho)^{|m+m'|/2}\right]\\ &\times{}_2F_1(k+(|m-m'|-|m+m'|)/2,\\ &\quad 1-k+(|m-m'|-|m+m'|)/2;\\ &\quad 1+|m-m'|;-\rho). \end{aligned} \tag{C17}$$

In (C17), ${}_2F_1$ denotes the usual hypergeometric function and

$$\theta_{mm}=1,$$

$$\theta_{mn}(k)=\frac{1}{|m-n|!}\times\prod_{j=1}^{|m-n|}[k(1-k)+(n+j)(n+j-1)]^{1/2}$$

$$\times\begin{cases}1, & m>n, \quad \text{(C18a)}\\ (-)^{n-m}, & m<n. \quad \text{(C18b)}\end{cases}$$

The asymptotic form (for $\zeta\gg 1$, i.e., $\rho\gg 1$) of (C17) can be shown to be

$$d^k_{mm'}(\rho)\sim\mathrm{Re}[C_{mn}(k)\rho^{-k}]\sim\mathrm{Re}[4C_{mn}(k)e^{-2k\zeta}], \tag{C19a}$$

where $C_{mn}(k)$ is a constant. In particular, for $k=\frac{1}{2}+is$, $m=m'=0$, we have

$$d^{1/2+is}_{00}(\rho)=f_s(\rho)\sim g(s)e^{is\ln\rho}/\sqrt{\rho}+\text{c.c.}, \tag{C19b}$$

where

$$g(s)=\Gamma(2is)/[\Gamma(\tfrac{1}{2}+is)]^2. \tag{C19c}$$

In (C19b) we have used the definition (5.7) of $f_s(\rho)$.

6. Orthogonality of the $D^k_{mm'}$

We observe from (C19) that the representations that form a complete set [those that were enclosed in square brackets in (C4)–(C8)] are square integrable with the measure (C12) or (C15). One can show that the usual orthogonality relation holds for them, either in terms of Kronecker or Dirac delta functions, i.e.,

$$\int D^{k*}_{m_1m_2}(R)D^{k'}_{m_1'm_2'}(R)d\mu(R)=\begin{cases} w_k^{-1}\delta_{kk'}\delta_{m_1m_1'}\delta_{m_2m_2'}, & \text{if } k,k'\in D^\pm, \quad \text{(C20a)}\\ w^{-1}(s)\delta(s-s')\delta_{m_1m_1'}\delta_{m_2m_2'}, & \begin{cases}\text{if } k=\frac{1}{2}+is, k'=\frac{1}{2}+is'\in C^0,\\ \text{or } k=\frac{1}{2}+is, k'=\frac{1}{2}+is'\in C^{1/2},\end{cases} \quad \text{(C20b)}\\ 0, & \text{if } k,k'\in \text{different classes}. \quad \text{(C20c)}\end{cases}$$

In (C20), w_k, $w(s)$ are functions of the index of the representation, i.e.,

$$w_k=2k-1, \quad k\in D^\pm, \quad k>\tfrac{1}{2}, \tag{C21a}$$

$$w(s)=2s\tanh\pi s, \quad k\in C^0, \quad k=\tfrac{1}{2}+is, \tag{C21b}$$

$$w(s)=2s\coth\pi s, \quad k\in C^{1/2}, \quad k=\tfrac{1}{2}+is. \tag{C21c}$$

For the discrete classes $D^\pm_k$ and $k>\frac{1}{2}$, Eq. (C21a) is taken directly from Eq. (12.9) of Ref. 3.

For the continuous classes $C^{0,1/2}_k$ and $k=\frac{1}{2}+is$, we proceed as follows. From (C16) we immediately obtain orthogonality when $m\neq m'$ and $n\neq n'$. When $m=m'$, $n=n'$, we use Eq. (12.25) of Ref. 3:

$$\int_0^\infty B^{(1)*}_{mn}(\zeta)B^{(2)}_{mn}(\zeta)\sinh 2\zeta\, d\zeta=2\pi\int_0^\infty c(s)\psi^{(1)*}(s)\psi^{(2)}(s)\,ds, \tag{C22}$$

where

$$B^{(1,2)}_{mn}=\int\psi^{(1,2)*}(s)d^{(s)}_{mn}(\zeta)\,ds \tag{C23}$$

and

$$c(s)=\begin{cases}(\coth\pi s)/4\pi s, & \text{for } C^0, \quad \text{(C24)}\\ (\tanh\pi s)/4\pi s, & \text{for } C^{1/2}. \quad \text{(C25)}\end{cases}$$

In (C23), the $\psi^{(1,2)}(s)$ construct wave packets out of the $d^{(s)}_{mn}$. The additional factor $\omega_{nn}(s)$ of Eq. (12.18) of Ref. 3 was omitted, since it drops out in the analysis due to the fact that $|\omega_{nn}(s)| = 1$.

We now choose

$$\psi^{(1)}(s) = \delta(s - s_1)\,, \quad \psi^{(2)}(s) = \delta(s - s_2) \tag{C26}$$

and thus obtain from (C22)

$$\int_0^\infty d^{(s_1)*}_{mn}(\zeta) d^{(s_2)}_{mn}(\zeta) \sinh 2\zeta \, d\zeta = 2\pi c(s_1)\delta(s_1 - s_2)\,. \tag{C27}$$

From (C16) and (C27) we then obtain (C20b), with (C21b) and (C21c).

Using the orthogonality relations (C20) we can find the coefficients of the expansion (C9) as

$$a^k_{mm'} = w_k \int f(R)[D^k_{mm'}(R)]^* \, d\mu(R)\,, \tag{C28a}$$

$$a^s_{mm'} = w(s) \int f(R)[D^{1/2+is}_{mm'}(R)]^* \, d\mu(R)\,. \tag{C28b}$$

7. The composition law

Consider two elements R_1, R_2 of SU(1,1),

$$R_1(\mu_1,\zeta_1,\nu_1)\,, \quad R_2(\mu_2,\zeta_2,\nu_2)\,, \tag{C29}$$

and their product R,

$$R_1R_2 = R(\mu,\zeta,\nu)\,. \tag{C30}$$

From the representation (C10) we see that in the middle of the product R_1R_2, the angles ν_1 and μ_2 combine as $\nu_1 + \mu_2$. The resulting ζ depends thus on ζ_1, ζ_2, and $\nu_1 + \mu_2$ only:

$$\cosh^2\zeta = \cosh^2\zeta_1 \cosh^2\zeta_2 + \sinh^2\zeta_1 \sinh^2\zeta_2 + 2\sinh\zeta_1 \times \sinh\zeta_2 \cosh\zeta_1 \cdot \cosh\zeta_2 \cos[2(\nu_1 + \mu_2)]\,. \tag{C31}$$

The resulting ν,μ are

$$\mu = \mu_1 + (\psi - \phi)/2\,, \quad \nu = \nu_2 + (\psi + \phi)/2\,, \tag{C32}$$

where ψ and ϕ are given by

$$\tan\psi = \frac{\cosh(\zeta_1 - \zeta_2)}{\cosh\zeta \cosh(\zeta_1 + \zeta_2)} \tan(\nu_1 + \mu_2)\,, \tag{C33}$$

$$\tan\phi = \frac{\sinh(\zeta_1 - \zeta_2)}{\sinh\zeta \sinh(\zeta_1 + \zeta_2)} \tan(\nu_1 + \mu_2)\,. \tag{C34}$$

In terms of the variable ρ of Eq. (C13) we can also write (C31) as

$$\rho = \rho_1 + \rho_2 + 2\rho_1\rho_2 + 2[\rho_1(1+\rho_1)\rho_2(1+\rho_2)]^{1/2} \times \cos[2(\nu_1 + \mu_2)]\,. \tag{C35}$$

8. The inverse R^{-1}

From the representation (C10) we immediately see that, if we call (μ,ζ,ν) the parameters of R and (μ',ζ',ν') those of R^{-1}, we have the relations

$$\mu' = -\nu - \pi/2\,, \quad \zeta' = \zeta\,, \quad \nu' = -\mu + \pi/2\,. \tag{C36}$$

9. The convolution

Just as in (3.14) for $R(3)$, we define the convolution $h(R)$ of two functions $f(R)$, $g(R)$ defined on R as

$$h(R) \equiv \int f(R_1) g(R_1^{-1}R) d\mu(R_1) \equiv f * g\,. \tag{C37}$$

In the text we are especially interested in the case where f and g are probability densities $p_1(\rho_1)$, $p_2(\rho_2)$, respectively, that do not depend on the angles. Using (C37), (C36), (C35), and (C15) we can write the resulting convolution (μ,ρ,ν) as

$$p(\mu,\rho,\nu) = \int_0^\infty d\rho_1 \iint_{-\pi}^{\pi} \frac{d\mu_1\, d\nu_1}{(2\pi)^2} p_1(\rho_1) p_2(\rho_1 + \rho + 2\rho_1\rho - 2[\rho_1(1+\rho_1)\rho(1+\rho)]^{1/2} \cos 2(\mu - \mu_1))\,. \tag{C38}$$

Integrating over ν_1 and shifting the variable μ_1 to $\mu - \mu_1$, we thus find that the resulting p is given by

$$p(\rho) = \int_0^\infty d\rho_1 p_1(\rho_1) \int_{-\pi}^{\pi} \frac{d\mu_1}{2\pi} p_2(\rho + \rho_1 + 2\rho\rho_1 - 2[\rho_1(1+\rho_1)\rho(1+\rho)]^{1/2} \cos 2\mu_1) \equiv p_1 * p_2 \tag{C39}$$

and is independent of the angles μ,ν. Therefore, the convolution of two isotropic functions is again isotropic. This result was also shown in the text in connection with Eqs. (5.6).

Equation (C39) generalizes to SU(1,1) and for isotropic functions, the usual expression for the convolution.

Equation (C37) implies

$$\langle D^k \rangle_h = \langle D^k \rangle_f \langle D^k \rangle_g\,, \tag{C40}$$

where

$$\langle D^k \rangle_f = \int f(R) D^k(R) d\mu(R)\,, \tag{C41}$$

and similarly for $\langle D^k \rangle_g$, $\langle D^k \rangle_h$.

For the isotropic functions p, p_1, p_2, of Eq. (C39), the argument below (5.5) then shows that

$$\langle f_s \rangle = \langle f_s \rangle_1 \langle f_s \rangle_2\,. \tag{C42}$$

Equations (C40) and (C42) are the equivalent, for SU(1,1), of the familiar convolution theorem.

10. A relation among $f_s(\rho)$'s from the representation property

From the definition of a representation we have the relation

$$\sum_{m_1} D^{1/2+is}_{mm_1}(R_1) D^{1/2+is}_{m_1m'}(R_2) = D^{1/2+is}_{mm'}(R_1R_2)\,. \tag{C43}$$

We use (C16), set $m = m' = 0$, and use (5.7a) and (C35),

$$\sum_{m_1} [d^{1/2+is}_{0m_1}(\rho_1) e^{-2im_1\nu_1}][e^{-2im_1\mu_2} d_{m_10}(\rho_2)] = f_s(\rho_1 + \rho_2 + 2\rho_1\rho_2 + 2[\rho_1(1+\rho_1)\rho_2(1+\rho_2)]^{1/2} \times \cos 2(\nu_1 + \mu_2))\,. \tag{C44}$$

We now integrate over ν_1 and μ_2 and making use of (5.7a) again we get

$$f_s(\rho_1)f_s(\rho_2) = \int_{-\pi}^{\pi} f_s(\rho_1+\rho_2+2\rho_1\rho_2 + 2[\rho_1(1+\rho_1)\rho_2(1+\rho_2)]^{1/2}\cos 2\nu_1)\frac{d\nu_1}{2\pi}. \quad \text{(C45)}$$

APPENDIX D: PROOF OF EQ. (5.22)

Since both $f_s(\rho)$ and $w(s)$ in (5.21) vary slowly, one can look for the saddle s_0 defined by $(\partial\psi_1/\partial s)_{s=s_0}=0$; but, since $\psi=\psi(s^2)$, we have $s_0=0$ and $w(s_0)=0$. It is thus better, before finding the saddle, to write $w(s)$ as

$$w(s) = 2s\tanh\pi s = \frac{s}{\sinh 2\pi s}(e^{2\pi s}+e^{-2\pi s}-2) \quad \text{(D1)}$$

and evaluate the auxiliary integral

$$I_\beta = \frac{1}{2}\int_{-\infty}^{\infty} f_s(\rho)\frac{s}{\sinh 2\pi s}e^{N\psi_1(s)+\beta s}\,ds, \quad \text{(D2)}$$

in terms of which $p_N(\rho)$ is given by

$$p_N(\rho) = I_{2\pi}+I_{-2\pi}-2I_0. \quad \text{(D3)}$$

The saddle s_0 of the integrand in (D2) is now the solution to the equation

$$\psi_1'(s_0) = -\beta/N. \quad \text{(D4)}$$

From (5.12a) we have

$$\psi_1'(s) = F_1'(s)/F_1(s), \quad \text{(D5)}$$

and, using the expansion (5.23), we can write (D4) as

$$\frac{2\langle g\rangle_1 s_0+\cdots}{\langle f_0\rangle_1+\cdots} = -\frac{\beta}{N}. \quad \text{(D6)}$$

Since $\beta/N \ll 1$, we see that an approximate solution to (D6) is

$$s_0 \approx -\frac{\beta}{N}\frac{\langle f_0\rangle_1}{2\langle g_0\rangle_1}, \quad \text{(D7)}$$

which is real and very close to the origin ($|s_0|\ll 1$). The saddle-point approximation to I_β is thus

$$I_\beta = \frac{1}{2}f_{s_0}(\rho)\frac{s_0}{\sinh 2\pi s_0}e^{\beta s_0+N\psi_1(s_0)}\frac{2\pi}{N|\psi_1''(s_0)|}. \quad \text{(D8)}$$

We now expand (D8) in powers of s_0. Recalling that $\psi_1(s)$ is a function of s^2, and hence $\psi'(0) = \psi'''(0) = \cdots = 0$, we have, to first order in s_0,

$$I_\beta \approx \frac{1}{2}f_0(\rho)\frac{1}{2\pi}(1+\beta s_0+\cdots)e^{N\psi_1(0)}\frac{2\pi}{N|\psi_1''(0)|}. \quad \text{(D9)}$$

Substituting (D9) in (D3) we obtain the result (5.22) of the text.

APPENDIX E: PROOF OF EQ. (5.48)

We evaluate $q_N(x)$ of Eq. (5.41). The saddle point is the solution to (5.44), which can be integrated to give

$$s_0 - s_a = -\frac{ix+\delta'(s_a)}{N\psi''(s_a)} + \frac{[ix+\delta'(s_a)]\delta''(s_a) - [\psi_1'''(s_a)/2\psi_1''(s_a)][ix+\delta'(s_a)]^2}{N^2[\psi_1''(s_a)]^2} + \cdots. \quad \text{(E1)}$$

We now expand $\omega(s_0)$ around s_a and keep terms up to the order $1/N$. We then expand $\delta(s_0)$ as

$$\delta(s_0) = \delta(s_a) + (s_0-s_a)\delta'(s_a) + \cdots. \quad \text{(E2)}$$

On the other hand, in the expansion of $\psi_1(s_0)$ we keep one more term, because in (5.42) $\psi_1(s)$ is multiplied by the factor N, i.e.,

$$\psi(s_0) = \psi(s_a) + (s_0-s_a)\psi_1(s_a) + (1/2!)(s_0-s_a)^2\psi''(s_a) + \cdots. \quad \text{(E3)}$$

Substituting (E1) in (E2) and (E3), and these in (5.42), we obtain

$$\omega(s_0) = \frac{x^2}{2N\psi_1''(s_a)} + \left[is_a - i\frac{\delta'(s_a)}{N\psi_1''(s_a)}\right]x + is_a aN - \frac{[\delta'(s_a)]^2}{2N\psi_1''(s_a)} + \delta(s_a) + N\psi_1(s_a). \quad \text{(E4)}$$

We also need $\omega''(s_0)$, which determines the variance of the Gaussian in the saddle-point integral. Only the leading term is needed, i.e.,

$$\omega''(s_0) = N\psi_1''(s_a) + O(N^0) + O(N^{-1}). \quad \text{(E5)}$$

Now $\omega''(s_0)$ can be shown to be real and negative (see Appendix F), thus giving

$$\int_{-\infty}^{\infty} e^{(1/2)\omega''(s_0)(s-s_0)^2}\,ds = \left[\frac{2\pi}{-N\psi''(s_0)}\right]^{1/2}. \quad \text{(E6)}$$

For the $q_N(x)$ of (5.41) we thus have

$$q_N(x) = \frac{1}{2}\mathrm{Re}\left[\frac{2\pi}{-N\psi''(s_a)}\right]^{1/2} \times \exp\left\{\delta(s_a) + \left(is_a+\frac{1}{2}\right)aN + N\psi_1(s_a) + \frac{x^2}{2N\psi_1''(s_a)} + \left[is_a - i\frac{\delta'(s_a)}{N\psi_1''(s_a)}+\frac{1}{2}\right]x\right\}. \quad \text{(E7)}$$

This result is valid for an arbitrary choice of a in Eq. (5.40). We now restrict ourselves to the case in which a is such that $z=aN$ is the centroid of the distribution (E7). This is achieved when the coefficient in front of x in (E7) vanishes, thus giving

$$s_a = i/2 + \delta'(s_a)/[N\psi''(s_a)]. \quad \text{(E8)}$$

We then have

$$\Gamma(1+is_a) \approx \sqrt{\pi}, \quad \text{(E9)}$$

$$\Gamma(\tfrac{1}{2}+is_a) \approx N\psi_1''(s_a)/[i\delta'(s_a)], \quad \text{(E10)}$$

$$\tanh\pi s_a \approx N\psi_1''(s_a)/[\pi\delta'(s_a)], \quad \text{(E11)}$$

which, substituted in (5.43) for $e^{\delta(s_a)}$, gives

$$e^{\delta(s_a)} \approx 1/\pi . \tag{E12}$$

One also finds

$$N\psi_1(s_a) \approx \frac{\psi_1'(s_a)\delta'(s_a)}{\psi_1''(s_a)} . \tag{E13}$$

Substituting (E8), (E12), and (E13) in (E7), and using (5.40) one finally finds the result (5.48) of the text. Equation (5.49) follows from (5.45), with $s_a \approx i/2$ [Eq. (E8)].

APPENDIX F: EVALUATION OF THE PARAMETERS OF THE GAUSSIAN DISTRIBUTION (5.48)

We notice that the quantities in (5.49) and (5.50) are evaluated at $s = i/2$, i.e., at $k = 0$ according to (C5b) or $q = s^2 + \frac{1}{4} = 0$ [see Eq. (C3)]. Since $f_s(\rho)$ of Eq. (5.8) is a function of s^2, it is natural to express $f_s(\rho)$ and $F_1(s)$, $\psi_1(s)$ of (5.12a) as functions of q, which we shall designate with the same symbols, i.e., $f_q(\rho)$, $F_1(q)$, $\psi_1(q)$.

We first write the series (5.8) for $f_s(\rho)$ in terms of q as

$$f_q(\rho) = 1 + \sum_{n=1}^{\infty} (-)^n \left\{ \prod_{m=1}^{n} [q + m(m-1)] \right\} \frac{\rho^n}{(n!)^2} . \tag{F1}$$

It is of great interest to reexpress (F1) as a power series in q, i.e.,

$$f_q(\rho) = \sum_{n=0} \phi_n(\rho) \frac{q^n}{n!} . \tag{F2}$$

We have been able to evaluate $\phi_n(\rho)$ for $n = 0,1,2$ only. This is all we need for our purposes, although it would be interesting to know the $\phi_n(\rho)$ for arbitrary n.

We first have

$$\phi_0(\rho) = 1 . \tag{F3}$$

Differentiating (F1) with respect to q one obtains a series that can be summed to give

$$\phi_1(\rho) = \left(\frac{\partial f_q(\rho)}{\partial q} \right)_{q=0} = -\ln(1+\rho) . \tag{F4}$$

Differentiating (F1) twice one obtains the series

$$\phi_2(\rho) = \left(\frac{\partial^2 f_q(\rho)}{\partial q^2} \right)_{q=0} = 2 \sum_{n=2}^{\infty} (-)^n \sum_{m=2}^{N} \frac{1}{m(m-1)} \frac{\rho^n}{n} . \tag{F5}$$

One can easily show that

$$\sum_{m=2}^{n} \frac{1}{m(m-1)} = \frac{n-1}{n} , \tag{F6}$$

so that

$$\phi_2(\rho) = 2 \sum_{n=1}^{\infty} (-)^n \frac{\rho^n}{n} - 2 \sum_{n=1}^{\infty} (-)^n \frac{\rho^n}{n^2} . \tag{F7}$$

Define

$$g(\rho) = 2 \sum_{n=1}^{\infty} (-)^n \frac{\rho^n}{n} , \tag{F8}$$

$$h(\rho) = 2 \sum_{n=1}^{\infty} (-)^n \frac{\rho^n}{n^2} . \tag{F9}$$

Concerning $g(\rho)$, we see that

$$g(\rho) = -2 \ln(1+\rho) . \tag{F10}$$

We have not succeeded in expressing $\phi_2(\rho)$ in terms of known functions. We first notice that

$$h'(\rho) = \frac{2}{\rho} \sum_{n=1}^{\infty} (-)^n \frac{\rho^n}{n} = -\frac{2}{\rho} \ln(1+\rho) , \tag{F11}$$

so that

$$h(\rho) = -2 \int_0^{\rho} \frac{\ln(1+\rho')}{\rho'} d\rho' , \tag{F12}$$

and thus

$$\phi_2(\rho) = -2 \ln(1+\rho) + 2 \int_0^{\rho} \frac{\ln(1+\rho')}{\rho'} d\rho' . \tag{F13}$$

Results (F3), (F4), and (F13) are the ones quoted in Eq. (5.16) of the text.

Averaging $f_q(\rho)$ with $p_1(\rho)$ we now have

$$F_1(q) = \langle f_q(\rho) \rangle_1 = \sum_{n=0}^{\infty} \langle \phi_n(\rho) \rangle_1 \frac{q^n}{n!} , \tag{F14}$$

where

$$\langle \phi_0(\rho) \rangle_1 = F_1(q=0) = 1 , \tag{F15a}$$

$$\langle \phi_1(\rho) \rangle_1 = F_1'(q=0) = -\langle \ln(1+\rho) \rangle_1 , \tag{F15b}$$

$$\langle \phi_2(\rho) \rangle_1 = F_1''(q=0) = 2 \left\langle \int_0^{\rho} \frac{\ln(1+\rho')}{\rho'} d\rho' \right\rangle_1 - 2 \langle \ln(1+\rho) \rangle_1 , \tag{F15c}$$

where the primes indicate derivatives with respect to q.

The derivatives of $\psi_1(s)$ with respect to s that appear in (5.49) and (5.50) [remember that $\psi_1(s) = \ln F_1(s)$] can now be expressed in terms of derivatives with respect to q as

$$\psi_1'(s) = 2s F_1'(q)/F_1(q) , \tag{F16}$$

$$\psi_1''(s) = (4q-1) \frac{F_1''(q)}{F_1(q)} + 2 \frac{F_1'(q)}{F_1(q)} - (4q-1) \left[\frac{F_1'(q)}{F_1(q)} \right]^2 , \tag{F17}$$

so that, using (F15), we have

$$\psi_1'(s=i/2) = -i \langle \ln(1+\rho) \rangle_1 , \tag{F18}$$

$$\psi_1''(s=i/2) = -2 \left\langle \int_0^{\rho} \frac{\ln(1+\rho')}{\rho'} d\rho' \right\rangle_1 + \langle \ln(1+\rho) \rangle_1^2 . \tag{F19}$$

Substituting the last two equations in Eqs. (5.49) and (5.50) we obtain (5.51) and (5.52).

From (F19) we see that $\psi_1''(s=i/2)$ is real. We can also show that it is negative. Since

$$\langle \ln(1+\rho) \rangle_1^2 < \langle [\ln(1+\rho)]^2 \rangle_1 , \tag{F20}$$

we can write

$$\psi_1''\left(s = \frac{i}{2}\right) < -2 \left\langle \int_0^{\rho} \frac{\ln(1+\rho')}{\rho'} d\rho' \right\rangle_1 + \langle [\ln(1+\rho)]^2 \rangle_1 . \tag{F21}$$

From

$$\frac{\partial}{\partial\rho}\ln^2(1+\rho)=2\,\frac{\ln(1+\rho)}{1+\rho}, \tag{F22}$$

we have

$$\ln^2(1+\rho)=2\int_0^\rho\frac{\ln(1+\rho')}{1+\rho'}\,d\rho', \tag{F23}$$

so that (F19) becomes

$$\begin{aligned}\psi_1''\Big(s=\frac{i}{2}\Big)&\leqslant-2\left\langle\int_0^\rho\frac{\ln(1+\rho')}{\rho'}\,d\rho'\right\rangle_1\\&\quad+2\int_0^\rho\frac{\ln(1+\rho')}{1+\rho'}\,d\rho'\\&=-2\left\langle\int_0^\rho\frac{\ln(1+\rho')}{\rho'(1+\rho')}\,d\rho'\right\rangle_1<0,\end{aligned} \tag{F24}$$

which proves our assertion.

[1]M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Griffin, London, 1952), p. 193.
[2]For a review, see P. Erdös and R. C. Herndon, Adv. Phys. **31**, 63 (1982).
[3]V. Bargmann, Ann. Math. **48**, 568 (1947).
[4]K. B. Wolf, in *Group Theory and its Applications in Physics*, edited by T. H. Seligman (Latin American School of Physics, Mexico City, 1980) [*AIP Conference Proceedings*, No. 71 (AIP, New York, 1981)].
[5]K. Ito and Y. Kawada, Proc. Phys. Math. Soc. Jpn. **22**, 977 (1940).
[6]U. Grenander, in *Probability and Statistics* (Almqvist and Wiksell, Stockholm, 1959); U. Grenander, *Proc. 4th Berkeley Sympos. Math. Statist. Probab.* **2**, 171 (1961); U. Grenander, *Probabilities on Algebraic Structures* (Wiley, New York, 1963).
[7](a) H. Furstenberg and H. Kesten, Ann. Math. Statist. **31**, 335 (1963); (b) H. Furstenberg, Trans. Am. Math. Soc. **108**, 377 (1963).
[8]A. J. O'Connor, Commun. Math. Phys. **45**, 63 (1975).
[9]H. Heyer, *Probability Measures on Locally Compact Groups* (Springer, Berlin, 1977).
[10]J. Flores, P. A. Mello, and G. Monsiváis, "Statistical distribution of the resistance in a one-dimensional disordered electrified chain," Phys. Rev. B (submitted).
[11]M. Cassandro and G. Jona-Lasinio, Adv. Phys. **27**, 913 (1978).
[12]We do not assume such independence for the limit (4.12), since in the important SU(1,1) problem of Sec. V, k depends on p_1.
[13]*Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Natl. Bur. Stand., Washington, D.C., 1965).
[14]P. W. Anderson, D. J. Thouless, E. Abrahams and D. S. Fisher, Phys. Rev. B **22**, 3519 (1980), Eq. (27).
[15]N. Kumar, Phys. Rev. B **31**, 5513 (1985).

On the DLR equation for the two-dimensional sine–Gordon model

R. Gielerak[a)]
Laboratory of Theoretical Physics, JINR, Dubna, USSR

(Received 15 October 1985; accepted for publication 30 April 1986)

The Dobrushin–Lanford–Ruelle equation is studied in a certain space of measures in the case of two-dimensional trigonometric interactions. The uniqueness theorem extending the results of Albeverio and Hoegh-Krohn [S. Albeverio and R. Hoegh-Krohn, Commun. Math. Phys. **68**, 95 (1979)] is proved. The extension is obtained by the application of some correlation inequalities of the Ginibre-type, which reduce the proof of the uniqueness of the translationally invariant, regular, tempered Gibbs states to the question on the independence of the infinite-volume free energy of the boundary conditions. The required independence is proved in this paper.

I. INTRODUCTION. NOTATION AND THE RESULT

The statistical mechanics approach to the two-dimensional superrenormalizable Euclidean scalar (quantum) Euclidean field theory can be formulated in terms of the fundamental notions of the Gibbs states.[1,2]

The standard measure space $\{S'(R^2),B\}$, where $S'(R^2)$ stands for the (real part of) Schwartz's space of the tempered distributions and B for the Borel σ algebra of subsets in $S'(R^2)$, plays the role of the configuration space in this approach. Let $\mu^{\Gamma}_{0,b}$ be the Gaussian measure with the covariance $(-\Delta^{\Gamma}_b+1)^{-1}$ and mean equal to zero. Here Δ^{Γ}_b stands for the two-dimensional Laplace operator with some classical boundary condition b imposed on the given piecewise-C^1 curve Γ. In particular, we will write μ_0 for the measure with the free boundary condition and μ^{Γ}_0 for the Gaussian measure μ_0 with the Dirichlet boundary condition on Γ. Let us denote by $\Sigma(\Lambda)$ the local σ algebras generated by the free Gaussian field μ_0 and by $m_\infty(\Lambda)$, the space of bounded measurable [with respect to $\Sigma(\Lambda)$] functionals of the field φ.

Let $\{U_\Lambda(\varphi)\}$ be an additive functional of the free field such that $\exp(U_\Lambda(\varphi))\in\bigcap_{p>1}L^p(d\mu_0)$. We will say that a probabilistic, Borel cylindric (PBC) measure μ defines a quantum scalar field with the interaction $\{U_\Lambda\}$ iff (i) μ is locally absolutely continuous with respect to μ_0, i.e.,

$$\mu_{|\Sigma(\Lambda)}\ll\mu_{0|\Sigma(\Lambda)};$$

(ii) for any $F\in m_\infty(\Lambda)$ the conditional expectation values of F with respect to the measure μ and the local σ algebra $\Sigma(\Lambda^c)$, $E_\mu\{F|\Sigma(\Lambda^c)\}$, is equal to those computed with respect to the *measure* μ_Λ and the σ algebra $\Sigma(\Lambda^c)$, where μ_Λ is a measure

$$\mu_\Lambda(d\varphi)=(Z_\Lambda)^{-1}\exp(U_\Lambda(\varphi))\mu_0(d\varphi); \tag{1.1}$$

and (iii) the moments of the measure μ exist and obey standard requirements of the Euclidean field theory (such as Glimm–Jaffe axioms,[3] etc.). It is important to note that the family $E_{\mu_\Lambda}\{-|\Sigma(\Lambda^c)\}$ defines on the $\{S'(R^2),B\}$ a local specification in the standard sense.[1] In Refs. 4–6 it was proved that

$$E_{\mu_\Lambda}\{-|\Sigma(\Lambda^c)\}=\frac{E_{\mu_0}\{-e^{U_\Lambda(\varphi)}|\Sigma(\Lambda^c)\}}{E_{\mu_0}\{e^{U_\Lambda(\varphi)}|\Sigma(\Lambda^c)\}}. \tag{1.2}$$

The conditional expectation values $E_{\mu_0}\{-|\Sigma(\Lambda^c)\}(\eta)$ at the randomly chosen point $\eta\in S'(R^2)$ can be written as the solution of the following Dirichlet stochastic problem[4–6]:

$$\begin{aligned}&(-\Delta+1)\Psi^{\partial\Lambda}_\eta(x)=0,\quad x\in\Lambda-\partial\Lambda,\\&\Psi^{\partial\Lambda}_\eta(x)=\eta(x),\quad x\in\partial\Lambda.\end{aligned} \tag{1.3}$$

For a recent, deep discussion of such kinds of stochastic problems, see the paper by Benfatto *et al.*[5] Here, we recall some basic facts concerning the problem (1.3).

For a given additive functional $\{U_\Lambda\}$ let us denote by $\mathscr{G}^t(U_\Lambda)$ the space of all tempered [i.e., supported on $S'(R^2)$ PBC measures] such that

$$\forall_{|\Lambda|<\infty}\ \mu\subset E_{\mu_\Lambda}\{-|\Sigma(\Lambda^c)\}=\mu, \tag{1.4}$$

in the meaning of measures we shall call elements of the set $\mathscr{G}^t(U_\Lambda)$, the tempered Gibbs measures corresponding to the interaction $\{U_\Lambda\}$.

It is not hard to observe that the set $\mathscr{G}^t(U_\Lambda)$ is convex and weakly closed. From the results of Fölmer[7] and the recent, more general results of Winkler[8] and Weizsäcker and Winkler[9] it follows that the set $\mathscr{G}^t(U_\Lambda)$ has a structure similar to that of the Choquet simplex: every $\mu\in\mathscr{G}^t(U_\Lambda)$ may be uniquely represented as a resultant of some probabilistic measure ρ supported on the set $\partial\mathscr{G}^t(U_\Lambda)$ of extremal points of $\mathscr{G}^t(U_\Lambda)$, which exist by the Fölmer–Winkler results.[7,8]

A Gibbs measure $\mu\in\mathscr{G}^t(U_\Lambda)$ is called the regular Gibbs measure iff its two-point moment can be extended continuously to the Sobolev space $\mathscr{H}_{-1}(R^2)$, i.e., there exists a constant c such that

$$\forall_{f\in H_{-1}(R^2)}\int\varphi^2(f)\mu(d\varphi)\leqslant c\|f\|^2_{-1}. \tag{1.5}$$

A Gibbs measure $\mu\in\mathscr{G}^t(U_\Lambda)$ is called the completely regular Gibbs measure iff there exists a constant C such that

$$\forall_{f\in H_{-1}(R^2)}\int e^{\varphi(f)}\mu(d\varphi)\leqslant e^{C\|f\|^2_{-1}}. \tag{1.6}$$

We denote the set of regular Gibbs measures (resp. completely regular Gibbs measures) by $\mathscr{G}^t_r(U_\Lambda)$ [resp. $\mathscr{G}^t_{cr}(U_\Lambda)$]. From the definition it follows that $\mathscr{G}^t_{cr}(U_\Lambda)$

[a)] On leave of absence from the Institute of Theoretical Physics, University of Wroclaw, Poland.

$\subseteq \mathscr{G}^t_r(U_\Lambda) \subseteq \mathscr{G}^t(U_\Lambda)$. From Refs. 5 and 6 we know that for any $\mu \in \mathscr{G}^t_r(U_\Lambda)$ the stochastic Dirichlet problem (1.3) has for almost every η with respect to μ a solution given by the classical Poisson formula:

$$\Psi^{\partial\Lambda}_\eta(x) = \int_{\partial\Lambda} P^{\partial\Lambda}(x,z)\eta(z)dz, \qquad (1.7)$$

valid for $x \notin \partial\Lambda$, where $P^{\partial\Lambda}(x,z)$ is the Poisson kernel for the operator $-\Delta + 1$. These unique solutions have certain local decay properties as $\Lambda \uparrow R^d$ (see Ref. 6 for greater details).

The most important questions in the general theory of Gibbs states are the questions about the existence and detailed topological structure of the sets like $\mathscr{G}^t(U_\Lambda)$. The existence problem for the field-theoretical Gibbs measures has been treated intensively in the seventies. See Refs. 3 and 10 for references. However, till now there did not exist a satisfactory version of the Dobrushin-like theory[11,12] in the field-theoretical context. Let us remark that it is not known whether every $\mu \in \mathscr{G}^t_r(U_\Lambda)$ for a given interaction $\{U_\Lambda\}$ defines a quantum field theory in the sense of (i)–(iii). By experience with the lattice spin systems with noncompact state space of the individual spin,[13,14] we expect that there may exist some spurious solutions of the DLR equations (1.4) in the space $\mathscr{G}^t(U_\Lambda)$.

Essentially important is the question about the cardinality of the set ${}^{\mathrm{QF}}\mathscr{G}^t(U_\Lambda) \cap \mathscr{G}^t(U_\Lambda)$, where we denote by ${}^{\mathrm{QF}}\mathscr{G}^t(U_\Lambda)$ the set of quantum-field theoretical solutions of the DLR equations (1.4). Whenever the above-mentioned set has more than one element, we have to deal with the phenomena of the first-order phase transition. Deep results in this direction have been obtained recently by Jmbrie[15] for the case of polynomial interactions. A detailed description of the set $\mathscr{G}^t(U_\Lambda)$ is, however, a very hard mathematical problem (it is very nontrivial already on the level of the two-dimensional Ising model[16,17]). The analysis of the set $\mathscr{G}^t_r(U_\Lambda)$ for a given $\{U_\Lambda\}$ seems to be a much easier problem. In the case of exponential interactions we have proved in Refs. 18 and 19 that the set $\mathscr{G}^t_{cr}(\{U_\Lambda\})$ reduces exactly to one quantum field theory solution of (1.4). This has been proved also in Ref. 20 extending the ideas taken from Ref. 18.

In this paper we consider the problem of uniqueness [in the space $\mathscr{G}^t_r(U_\Lambda)$] of the solutions of the DLR equations (1.4) for the so-called sine–Gordon interaction. Previously Albeverio and Hoegh-Krohn, using a high-temperature cluster expansion, proved the uniqueness for the weakly coupled sine–Gordon interaction.[6] A similar uniqueness result has been proved[21] for the regularized Yukawa d-dimensional, neutral gas in the region of couplings where the contraction map principle can be applied to the Kirkwood–Salsburg equations.

The sine–Gordon interactions are defined by

$$U_\Lambda(\varphi) = z\int_\Lambda d^2x{:}\cos(\epsilon\varphi(x) + \theta){:}. \qquad (1.8)$$

Here we consider this model in the region

$$z \geqslant 0, \quad \epsilon^2 < 2/(1 - 1/2\pi), \quad \theta = 0.$$

See papers 22–24 for the construction of the corresponding Gibbs measures with $z \in R^1$, $\theta \neq 0$, and $\epsilon^2 < 4\pi$.

The result proved in this paper is the following.

Theorem 1: Whenever the infinite volume pressure $p_\infty(z)$ in the model (1.8) is differentiable at $z = z_0$, the set of regular solutions which have translationally invariant first moment of the corresponding DLR equations at $z = z_0$ consist of exactly one element equals the infinite-volume half-Dirichlet state.

The uniqueness result of Albeverio and Hoegh-Krohn holds only in the region of the convergence of the Glimm–Jaffe Spencer expansion, i.e., for sufficiently small $|z|$. Taking into account that $p_\infty(z)$ as a concave function of z is almost everywhere differentiable (presumably for all z), we have that for almost every z there exists a unique pure Gibbs phase corresponding to the interaction (1.8).

The proof we find seems to be very elementary. We adopt to the present case some correlation inequalities found by Fröhlich and Pfister in their analysis of the DLR equations for Abelian lattice spin systems.[25–27] The additional argument we use for the proof is the independence of the infinite volume pressure of boundary condition (generalizing the known result of Guerra–Rosen–Simon[28] and concerning the independence of the so-called classical boundary conditions).

The ideas used in this paper can be applied to analyze the class of weakly coupled $P(\varphi)_2$ models or the ${:}\varphi^4{:}$ models where the Lee–Yang theorem works. Similar techniques have been applied to the class of charge-symmetric continual systems.[29,30]

Finally, let us say a few words about the organization of this paper. In Sec. II we review some correlation inequalities of Eqs. (1.4) to the statement about independence of the infinite-volume pressure of the boundary conditions. In Sec. III we prove the claimed independence. Section IV contains some techniques necessary to complete the proof of Theorem 1.

II. REDUCTION OF THE PROOF TO THE STATEMENT ABOUT DIFFERENTIABILITY OF THE PRESSURE

The infinite-volume half-Dirichlet sine–Gordon measure corresponding to the interactions (1.8) can be constructed easily by the following correlation inequalities proved in Ref. 31:

$$\langle e^{t\varphi(f)}; {:}\cos\epsilon\varphi{:}(x)\rangle^T_\Lambda(z) \leqslant 0, \qquad (2.1)$$

$$\langle \varphi^2(f); {:}\cos\epsilon\varphi{:}(x)\rangle^T_\Lambda(z) \leqslant 0, \qquad (2.2)$$

where $\langle\ ;\ \rangle^T$ means the truncated expectation value and $\langle\ \rangle^0_\Lambda(z)$ means the expectation with respect to the measures

$$\mu^{\partial\Lambda}_\Lambda(d\varphi) = (Z^0_\Lambda(z))^{-1}\exp\left(z\int_\Lambda {:}\cos\epsilon\varphi{:}(x)d^2x\right)\mu^{\partial\Lambda}_0(d\varphi),$$

$$Z^0_\Lambda(z) = \int \mu^{\partial\Lambda}_0(d\varphi)\exp\left(z\int_\Lambda {:}\cos\epsilon\varphi{:}(x)d^2x\right). \qquad (2.3)$$

From these inequalities it follows easily that the infinite-volume limit $\lim_{\Lambda\uparrow R^2}\langle - \rangle^0_\Lambda = \langle\ \rangle^0_\infty$ exists (independently at how $\Lambda \uparrow R^2$) and fulfills the axioms of Ref. 32. In particular, we have the bound

$$\langle e^{t\varphi(f)}\rangle^0_\infty \leqslant e^{(t^2/2)\|f\|^2_{-1}}, \qquad (2.4)$$

which means that $\langle - \rangle^0_\infty \equiv \int - \mu_\infty(d\varphi)$ is a complete regular Gibbs measure. Let us introduce the following notation:

$$c(\varphi)(x) = :\cos \epsilon\varphi:(x), \quad s(\varphi) = :\sin \epsilon\varphi:.$$

By $\mathscr{G}_r^t(z)$ we denote the set of regular Gibbs measures corresponding to the interaction (1.8) with fixed z and ϵ. The symbol $\overset{\mu}{\underset{\eta}{\forall}}$ means for almost every η with respect to μ. A conditioned finite-volume Gibbs measure $\mu_\Lambda^\eta(d\varphi)$ is given by

$$\mu_\Lambda^\eta(d\varphi) = (Z_\Lambda^\eta)^{-1} \times \exp\left(z \int_\Lambda c(\varphi + \Psi_\eta^{\partial\Lambda})(x) d^2x\right) \mu_0^{\partial\Lambda}(d\varphi), \tag{2.5}$$

where $\Psi_\eta^{\partial\Lambda}$ is the solution of the problem (1.3) and η is randomly chosen from $S'(R^2)$.

From the *reverse martingale* theorem it follows that, for a given $\mu \in \mathscr{G}_r^t(z)$,

$$\mu_\infty^\eta(d\varphi) = \lim_{\Lambda \uparrow R^2} \mu_\Lambda^\eta(d\varphi)$$

exists for μ a.e. $\eta \in S'(R^2)$, and defines some PBC measure on $\{S'(R^2),B\}$. Moreover, the full set $\mathscr{G}'(z)$ can be obtained as convex superpositions of such limits.

For the conditioned measures μ_Λ^η the correlation inequalities (2.1) and (2.2) in general fail. However, instead of the correlation inequalities (2.1) and (2.2) one can use another set of correlation inequalities in order to analyze the content of the set $\mathscr{G}_r^t(z)$. These correlation inequalities proved below are simple adaptations of the correlation inequalities proved by Fröhlich and Pfister in Ref. 25. They all are simple applications of the Ginibre correlation inequalities.[33] In the shorthand, let us introduce the notation

$$\begin{aligned}
C_\Lambda(x_1,...,x_n) &= \int \mu_\Lambda^{\partial\Lambda}(d\varphi) \prod_{i=1}^n c(\varphi)x_i),\\
C_\Lambda^\eta(x_1,...,x_n) &= \int \mu_\Lambda^\eta(d\varphi) \prod_{i=1}^n c(\varphi)x_i),\\
S_\Lambda^\eta(x_1,...,x_n) &= \int \mu_\Lambda^\eta(d\varphi) \prod_{i=1}^n s(\varphi)(x_i),\\
S_\Lambda(x_1,...,x_n) &= \int \mu_\Lambda^{\partial\Lambda}(d\varphi) \prod_{i=1}^n s(\varphi)(x_i),
\end{aligned} \tag{2.6}$$

and similarly for the corresponding infinite volume limits. The existence of $\lim_{\Lambda\uparrow R^2} c_\Lambda$ follows from the $:\cos \epsilon\varphi:$ bound[21] and the following correlation inequalities:

$$\left\langle \prod_{i=1}^n c(\varphi)x_i; \prod_{j=1}^m c(\varphi)(y_j) \right\rangle_\Lambda^T \geqslant 0, \tag{2.7}$$

proved in Ref. 31. The existence of the limits $\lim_{\Lambda\uparrow R^2} C_\Lambda^\eta$, $\lim_{\Lambda\uparrow R^2} S_\Lambda^\eta$ (by subsequences) follows from the correlation inequalities to be proved below [see inequality (2.14)] and the compactness arguments. Moreover, from the results of Sec. IV it follows that every accumulation point of C_∞^η is equal to C_∞ (at least for regular values of z, see below).

Let us denote by $\langle - \rangle_\Lambda^{(0,\eta)}$ the expectation on $\{S'(R^2),B\}^{\otimes 2}$ with respect to the measure $\mu_\Lambda^{\partial\Lambda}(d\varphi) \otimes \mu_\Lambda^\eta(d\varphi')$.

Proposition 2.1: Let $\mu \in \mathscr{G}_r^t((z)$. Then for every $n > 0$, $f_1,...,f_n \in S(R^2)$; such that $f_i \geqslant 0$, for $i = 1,...,n$, $g \in S(R^2)$ the following correlation inequalities hold:

$$\overset{\mu}{\underset{\eta}{\forall}} \quad 0 \leqslant \left\langle \left(\prod_{i=1}^n c(\varphi)(f_i) - \prod_{i=1}^n c(\varphi'')(f_i) \right) \times \exp\left(\pm\lambda \int d^2x\, g(x) c(\varphi)(x) c(\varphi')(x) \right) \right\rangle_\Lambda^{(0,\eta)},$$
$$\lambda \in R^1. \tag{2.8}$$

Proof: It is a standard application of the duplicate variable technique. Let φ' be an identical copy of the field φ. From

$$\begin{aligned}
&\exp\left(z \int_\Lambda d^2x\, c(\varphi)x)\right) \exp\left(z \int_\Lambda dx\, c(\varphi' + \Psi_\eta^{\partial\Lambda})(x)\right)\\
&= \exp\left(z \int_\Lambda d^2x\, c\left(\frac{\varphi + \varphi' + \Psi_\eta^{\partial\Lambda}}{2}\right) \times (x) c\left(\frac{\varphi - \varphi' - \Psi_\eta^{\partial\Lambda}}{2}\right)(x)\right)
\end{aligned} \tag{2.9}$$

and

$$c(\varphi)(x)c(\varphi')(x) = \frac{1}{2} c\left(\frac{\varphi - \varphi'}{2}\right)(x) c\left(\frac{\varphi + \varphi'}{2}\right)(x), \tag{2.10}$$

we conclude, after introducing the orthogonal transformation in the space $\{\varphi,\varphi'\}$,

$$\Psi_+ = (\varphi + \varphi')/\sqrt{2}, \quad \Psi_- = (-\varphi + \varphi')/\sqrt{2}, \tag{2.11}$$

that the first exponential and the one coming from the interaction factorize after (convergent for $|\Lambda| < \infty$) expansions in z. The terms outside the exponentials factorize using the following trigonometric identities:

$$\prod_{j=1}^n \cos \varphi_j = \frac{1}{2} \sum_{\{\epsilon_j\}} \cos\left(\sum_{j=1}^n \epsilon_j \varphi_j\right), \quad \epsilon_i = \pm 1, \tag{2.12}$$

and

$$\cos\alpha - \cos\beta = 2\sin[(\alpha + \beta)/2]\sin[(\beta - \alpha)/2]. \tag{2.13}$$

Some intermediate UV regularizations are needed to justify these transformations rigorously but the removal of it is simple so that we omit the details here. Q.E.D.

These correlation inequalities lead to the following inductive statement on the independence of the moments like $C_\Lambda^\eta(x_1,...,x_n)$ of the boundary condition η.

Corollary 2.2: If for $f_1,..., f_n \geqslant 0$ and some η,

$$\lim_{\Lambda\uparrow R^2} \left\langle \prod_{i=1}^n c(\varphi)(f_i) \right\rangle_\Lambda^\eta = \left\langle \prod_{i=1}^n c(\varphi)(f_i) \right\rangle_\infty^0$$

and

$$\langle c(\varphi)(x) \rangle_\infty^0 = \langle c(\varphi)(x) \rangle_\infty^\eta > 0,$$

then for any $n + 1$ we have

$$\lim_{\Lambda\uparrow R^2} \left\langle \prod_{i=1}^{n+1} c(\varphi)(f_i) \right\rangle_\Lambda^\eta = \left\langle \prod_{i=1}^{n+1} c(\varphi)(f_i) \right\rangle_\infty^0.$$

Proof: Expanding the exponential

$$\begin{aligned}
&\exp\left(\pm\lambda \int dx\, g(x) c(\varphi)(x) c(\varphi')(x) d^2x \right)\\
&= 1 \pm \lambda \int dx\, g(x) c(\varphi)(x) c(\varphi')(x) + O(\lambda^2),
\end{aligned}$$

we have, using Proposition 2.1 and the hypothesis,

$$0 = C_\infty(f_1,\dots,f_n) - C^\eta_\infty(f_1,\dots,f_n)$$
$$\geqslant \pm\lambda \int dx\, g(x) \left[\left\langle \prod_{i=1}^n c(\varphi)(f_i) c(\varphi)(x) \right\rangle^0_\infty \times \langle c(\varphi)(x)\rangle^\eta_\infty - \left\langle \prod_{i=1}^n c(\varphi)(f_i) c(\varphi)(x) \right\rangle^\eta_\infty \times \langle c(\varphi)(x)\rangle^0_\infty \right].$$

Dividing by λ and letting λ tend to zero, we get

$$0 = \int dx\, g(x) \left[\left\langle \prod_{i=1}^n c(\varphi)(f_i) c(\varphi)(x) \right\rangle^0_\infty \langle c(\varphi)(x)\rangle^0_\infty - \left\langle \prod_{i=1}^n c(\varphi)(f_i) c(\varphi)(x) \right\rangle^\eta_\infty \langle c(\varphi)(x)\rangle^0_\infty \right].$$

By the assumed strict positivity of $\langle c(\varphi)(0)\rangle^0_\infty$ [actually it is easy to prove that $\langle c(\varphi)(0)\rangle^0_\infty > 1$], we conclude

$$\left\langle \prod_{i=1}^n c(\varphi)(f_i) c(\varphi)(g) \right\rangle^\eta_\infty = \left\langle \prod_{i=1}^n c(\varphi)(f_i) c(g) \right\rangle^0_\infty .$$

From this result we have the following corollary. Q.E.D.

Corollary 2.3: Let $\mu \in \mathscr{G}^t_r(z)$. If $c^0_\infty(0) = c^\eta_\infty(x)$ for some η, then for any $n \geqslant 1$

$$c_\infty(x_1,\dots,x_n) = c^\eta_\infty(x_1,\dots,x_n).$$

This corollary reduces the proof of independence of the moments $\{C^\eta_\infty(x_1,\dots,x_n)\}$ of the boundary condition η to the identical statement concerning the first moment only. Roughly speaking, the first moment is nothing but the derivative of the pressure. Thus, the proof of the independence of the moments $\{C^\eta_\infty(x_1,\dots,x_n)\}$ of the boundary conditions $\eta \in S'(R^2)$ has been reduced to the statement on the infinite-volume pressure independence of boundary condition η. This result will be proved in the next section.

Before ending this section let us note also other simple but remarkable correlation inequalities.

Proposition 2.4: Let $\mu \in \mathscr{G}^t_r((z)$. For any choice of $\alpha_i \in [0,2\pi)$, $i = 1,2,\dots,n$, we have

$$\overset{\mu}{\underset{\eta}{\forall}}: \quad C_\infty(x_1,\dots,x_n) \geqslant \left| \left\langle \prod_{i=1}^n c(\varphi + \alpha_i)(x_i) \right\rangle^\eta_\infty \right| . \tag{2.14}$$

Proof: This follows easily from the formula

$$C_\infty(x_1,\dots,x_n) - \left\langle \prod_{i=1}^n c(\varphi + \alpha_i)(x_i) \right\rangle^\eta_\infty = \left\langle \prod_{i=1}^n c(\varphi)(x_i) - \prod_{i=1}^n c(\varphi' + \alpha_i)(x_i) \right\rangle^{(0,\eta)}_\infty , \tag{2.15}$$

formulas (2.12) and (2.13), and the duplicate variable trick. Q.E.D.

Recalling the Fölmer, Weizsacker, and Winkler results mentioned in the Introduction we have the following corollary.

Corollary 2.5: The Gibbs measure μ_∞ is an extremal Gibbs measure in the set $\mathscr{G}^t(z)$ (for every $z > 0$).

Proof: This is a standard application of the representation of an integral over the Martin–Dynkin boundaries $\partial\mathscr{G}^t(z)$ of the set $\mathscr{G}^t(z)$. For convenience, we review some results from the Fölmer, Weizsacker, and Winkler theory in the Appendix to this paper.

Remark 2.1: For eventual future application we define also the following moments of the measure $\mu^\eta_\Lambda(d\varphi)$:

$$\hat{C}^\eta_\Lambda(x_1,\dots,x_n) \equiv E_{\mu_\Lambda}\left\{ \prod_{i=1}^n c(\varphi)(x_i) \middle| \Sigma(\Lambda^c) \right\}(\eta) = \int \mu^\eta_\Lambda(d\varphi) \prod_{i=1}^n :\cos(\epsilon\varphi(x_i) + \epsilon\Psi^{\partial\Lambda}_\eta(x_i)): \tag{2.16}$$

and

$$\hat{S}^\eta_\Lambda(x_1,\dots,x_n) \equiv E_{\mu_\Lambda}\left\{ \prod_{i=1}^n s(\varphi)(x_i) \middle| \Sigma(\Lambda^c) \right\}(\eta) = \int \mu^\eta_\Lambda(d\varphi) \prod_{i=1}^n :\sin(\epsilon\varphi(x_i) + \epsilon\Psi^{\partial\Lambda}_\eta(x_i)):. \tag{2.16'}$$

Then, it is not hard to observe the correlation inequalities listed above hold also for $\hat{C}^\eta_\Lambda$.

Proposition 2.6: Let $\mu \in \mathscr{G}^t_r(z)$. Then for every $n > 0$, $f_1,\dots,f_n \in S(R^2)$, such that $f_i \geqslant 0$, for $i = 1,\dots,n$, any $g \in S(R^2)$, the following correlation inequalities hold:

$$\overset{\eta}{\underset{\mu}{\forall}}: \underset{\lambda\in R^1}{\forall}, \quad 0 \leqslant \left\langle \left(\prod_{i=1}^n c(\varphi)(f_i) - \prod_{i=1}^n :\cos(\epsilon\varphi' + \epsilon\Psi^{\partial\Lambda}_\eta):(f_i) \right) \times \exp \pm\lambda \int d^2x\, g(x) c(\varphi) x \cos(\epsilon\varphi' + \epsilon\Psi^{\partial\Lambda}_\eta)(x): \right\rangle^{(0,\eta)}_\Lambda \tag{2.17}$$

and

$$\left| \left\langle \prod_{i=1}^n :\cos(\alpha\varphi + \epsilon\Psi^{\partial\Lambda}_\eta):(f_i) \right\rangle^\eta_\Lambda \right| \leqslant C^0_\Lambda(f_1,\dots,f_n). \tag{2.18}$$

In particular, this proposition leads to the same bootstrap principle as Corollary 2.2 for the moments

$$\lim_{\Lambda\uparrow R^2} \hat{C}^\eta_\Lambda(x_1,\dots,x_n) \equiv \hat{C}^\eta_\infty(x_1,\dots,x_n).$$

The existence of $\hat{c}^\eta_\infty(x_1,\dots,x_n)$ follows from the application of the reverse martingale theorem and the correlation inequality

(2.18). Moreover, from the correlation inequality (2.18) it follows easily that the set of the limits $\lim_\Lambda \mu_\Lambda^\eta$ forms then a weakly precompact set in the space of all probability measures on the space $\{S'(R^2),B\}$. The last remark follows easily by noting the following formula: for any $\mu\in\mathscr{G}_r^t(z)$,

$$E_\mu\{e^{i\varphi(f)}|\Sigma(\Lambda^c)\}(\eta) = e^{i\Psi_\eta^{\partial\Lambda}(f)}\int e^{i\varphi(f)}\mu_\Lambda^\eta(d\varphi) = e^{i\Psi_\eta^{\partial\Lambda}(f)}e^{-(1/2)\|f\|^2_{-1,\partial\Lambda}},$$

$$\int\mu_\Lambda^\eta(d\varphi)\exp\Big(z\int_\Lambda\int d\lambda(\alpha')dx{:}e^{i\alpha'(\varphi+\Psi_\eta^{\partial\Lambda})(x)}{:}(e^{-\alpha'(-\Delta^{\partial\Lambda}+1)^{-1}*f(x)}-1)\Big)$$

$$= e^{i\Psi_\eta^{\partial\Lambda}(f)}e^{-(1/2)\|f\|^2_{-1,\partial\Lambda}}E_{\mu_\Lambda}\Big\{\exp z\int_\Lambda\int dx\,d\lambda(\alpha'){:}e^{i\alpha'\varphi(x)}{:}(e^{-\alpha'(-\Delta^{\partial\Lambda}+1)^{-1}*f(x)}-1)|\Sigma(\Lambda^c)\Big\}(\eta),$$

$$\|f\|^2_{-1,\partial\Lambda}\equiv\int\int dx\,dy\,f(x)(-\Delta^{\partial\Lambda}+1)^{-1}f(y),$$

$$d\lambda(\alpha') = \tfrac{1}{2}(\delta(\alpha'-\alpha)+\delta(\alpha'+\alpha)).$$

III. INFINITE-VOLUME PRESSURE INDEPENDENCE OF THE BOUNDARY CONDITION

In this section we will concentrate on the proof that the infinite-volume limit pressure does not depend on the typical boundary condition "$\eta\in\operatorname{supp}\mu$" whenever $\mu\in\mathscr{G}_r^t(z)$.

Several results on the independence of the infinite-volume pressure for the so-called classical boundary conditions have been obtained.[28] However, it seems to us very likely that the class of classical boundary conditions is not of measure 1 from the point of the Gibbsian approach to this problem. Therefore, all the results[28] are incomplete for the present applications.

A. Shape independence

Let $\mu\mathscr{G}_r^t(z)$. Then, we define the finite-volume pressures

$$p_\Lambda^0(z) = -(1/|\Lambda|)\ln Z_\Lambda^0(z), \tag{3.1}$$

$$p_\Lambda^\eta(z) = -(1/|\Lambda|)\ln Z_\Lambda^\eta(z), \tag{3.2}$$

for "$\eta\in\operatorname{supp}\mu$." The corresponding infinite volume limits will be denoted by $p_\infty^0(z)$ and $p_\infty^\eta(z)$, respectively.

Lemma 3.1 (shape independence): Let $\mu\in\mathscr{G}_r^t(z)$. Whenever $\Lambda\uparrow R^2$, in the sense of van Hove, and such that $\partial\Lambda$ are piecewise C^1, then

$$\overset{\mu}{\underset{\eta}{\forall}}:\ \lim_{\Lambda\uparrow R^2}p_\Lambda^\eta(z) = p_\infty^\eta(z) \tag{3.3}$$

exists and does not depend on the chosen sequence $\Lambda\uparrow R^2$ as above.

Proof: Let us rewrite Z_Λ^η in the following way:

$$Z_\Lambda^\eta(z) = \int\mu_0^{\partial\Lambda}(d\varphi)\exp\Big(z\int_\Lambda{:}c(\varphi){:}_1(x)dx\Big)$$
$$\times\exp\Big(z\int_\Lambda{:}c(\varphi){:}_1(x)(c(\Psi_\eta^{\partial\Lambda})(x)-1)d^2x\Big)$$
$$\times\exp\Big(-z\int_\Lambda{:}s(\varphi){:}_1(x)s(\Psi_\eta^{\partial\Lambda})(x)d^2x\Big), \tag{3.4}$$

where $:\ :_1$ means the normal ordering with respect to the covariance $(-\Delta+1)^{-1}$, i.e.,

$${:}c(\varphi)(x){:}_1\text{"} = \text{"}\exp((\epsilon^2/2)S(0))\cos\epsilon\varphi(x),$$
$${:}s(\varphi)(x){:}_1\text{"} = \text{"}\exp((\epsilon^2/2)S(0))\sin\epsilon\varphi(x). \tag{3.5}$$

Using the L_2 estimate following from the (proof of) Theorem 3.4 in Ref. 22 for the half-Dirichlet state (at this point one can use conditioning inequalities), we have that there exist constants c_1,c_2 independent of η such that

$$|Z_\Lambda^\eta(z)|\leqslant c_1\exp c_2\lambda^2|\Lambda|. \tag{3.6}$$

Above we also used the following trivial bounds: $|s(\Psi_\eta^{\partial\Lambda})(x)|\leqslant 1$ and $|c(\Psi_\eta^{\partial\Lambda})(x)|<1$ pointwise on $S'(R^2)$. Recall that here the corresponding quantities are not Wick-ordered.

From the estimate (3.6) it follows that whenever $\Lambda_n\uparrow R^2$ in some well-prescribed sense there exists a subsequence $(n')\subset(n)$ such that $p_{\Lambda_{n'}}^\eta(z)$ is then convergent.

In the case when $\{\Lambda_n\}$ is such that $\partial\Lambda_n$ are C^1 piecewise and there exists $\epsilon>0$ such that

$$\lim_{n\to\infty}|\Lambda_n|/|\partial\Lambda_n|^{1+\epsilon} = \infty,$$

it follows that every accumulation point of the sequence $\{p_{\Lambda_n}^\eta(z)\}$ is equal to $p_\infty^0(z)$ and this proves the claimed convergence and shape independence. Q.E.D.

B. Estimates on $\Psi_\eta^{\partial\Lambda}$

Several local decay properties of the solutions of the stochastic Dirichlet problem (1.3) have been proved in the basic paper.[6] However, the results obtained in Ref. 6 are not sufficient for our purposes. As we will show below, some *a priori* bounds are needed for the estimation of the quantities like $\int_\Delta|\Psi_\eta^{\partial\Lambda}|^p(x)d^2x$, where Δ is a typically unit cube in R^2 and $p\geqslant 1$. Such estimates follow easily from the application of the Chebyshev inequality.

Let us denote

$$K^{\partial\Lambda}(x,y)\equiv(-\Delta+1)^{-1}(x,y)-(-\Delta^{\partial\Lambda}+1)^{-1}(x,y). \tag{3.7}$$

It is well known that $K^{\partial\Lambda}(x,x)$ is a smooth function for $x\notin\partial\Lambda$ and has exponential decay as $\operatorname{dist}(x,\partial\Lambda)\to\infty$. Moreover, as $x\to\partial\Lambda$, then $K^{\partial\Lambda}(x,x)$ behaves like $(1/2\pi)\ln|\operatorname{dist}(x,\partial\Lambda)|$ (see, i.e., Ref. 3).

Lemma 3.2: Let $\mu\in\mathcal{G}_r^t(z)$. Then for any unit cube $\Delta\subset R^2$ and any bounded $\Lambda\subset R^2$ with C^1-piecewise boundary, there exists a constant $C_3(\eta,\Lambda)$ finite for μ a.e. η such that for all $\beta<1$ the following estimate holds:

$$\int_\Delta (\Psi_\eta^{\partial\Lambda})^2(x)dx \leqslant C_3(\eta,\Lambda)\left[\int_\Delta K^{\partial\Lambda}(x,x)dx\right]^\beta. \quad (3.8)$$

Proof: Let $c(R^2)$ be the partitioning of R^2 onto unit cubes such that $\Delta\in c(R^2)$. Take $\delta>0$ arbitrary and fixed. For $\mu\in\mathcal{G}_r^t(z)$, we have

$$\mu\left\{\eta\in S'(R^2)\,\middle|\,\underset{j}{\exists}\int_{\Delta_j}(\Psi_\eta^{\partial\Lambda})^2(x)dx \geqslant \frac{1}{\delta}\left(\int_\Delta K^{\partial\Lambda}(x,x)dx\right)^\beta\right\}$$

$$\leqslant \sum_j \mu\left\{\eta\in S'(R^2)\,\middle|\,\int_{\Delta_j}(\Psi_\eta^{\partial\Lambda})^2(x)dx \geqslant \frac{1}{\delta}\left[\int_\Delta K^{\partial\Lambda}(x,x)dx\right]^\beta\right\}$$

(by the application of the Chebyshev inequality)

$$\leqslant \delta\sum_j\left[\int_{\Delta_j}K^{\partial\Lambda}(x,x)dx\right]^{-\beta}\times\int\mu(d\eta)\left(\int_{\Delta_j}(\Psi_\eta^{\partial\Lambda})^2(x)dx\right)$$

$$\leqslant \delta\ \text{const}\left(\sum_j\left(\int_{\Delta_j}K^{\partial\Lambda}(x,x)d^2x\right)^{1-\beta}\right).$$

Whenever $\beta<1$ the sum Σ_j is finite due to the exponential decay of $K^{\partial\Lambda}$. Since δ is arbitrary, the proof follows. Q.E.D.

For the case of completely regular Gibbs measure one can generalize the following.

Lemma 3.3: Let $\mu\in\mathcal{G}_{cr}^t(z)$. Then for any unit cube $\Delta\subset R^2$ and any bounded $\Lambda\subset R^2$ with a C^1-piecewise boundary there exists a constant $C_4(\eta,\Lambda)$ finite for μ almost every η and such that for all $\beta<n/2$ the following estimate holds:

$$\int_\Delta|\Psi_\eta^{\partial\Lambda}|^n(x)dx\leqslant C_4(\eta,\Lambda)\left\{\int_\Delta K^{\partial\Lambda}(x,x)\right\}^\beta. \quad (3.9)$$

Proof: The main argument is again the Chebyshev inequality applied as in the proof of Lemma 3.2. The additional argument comes from the assumed complete regularity of μ. From $\mu\in\mathcal{G}_{cr}^t(z)$ there follows from the Cauchy integral formula the following estimate:

$$\left|\int\mu(d\eta)\prod_{i=1}^n\eta(f_i)\right|\leqslant(n!)^{1/2}\ \text{const}\prod_{i=1}^n\|f_i\|_{-1}. \quad (3.10)$$

Q.E.D.

The following estimates of the quantities like $\int_\Delta(\nabla\Psi_\eta^{\partial\Lambda})^2\,dx$ for $\text{dist}(\Delta,\partial\Lambda)=\epsilon>0$ should be useful in future applications.[30,34]

Lemma 3.4: Let $\mu\in\mathcal{G}_r^t(z)$. Let $\Lambda\subset R^2$ be a bounded with C^1-piecewise boundary subset, and let Δ be a unit cube in R^2 such that $\text{dist}(\Delta,\partial\Lambda)=\delta>0$. There exists a constant $C_5(\eta,\Lambda,\delta)$ finite for μ a.e. η and such that

$$\int_\Delta|\nabla\Psi_\eta^{\partial\Lambda}(x)|^2 \leqslant C_5(\eta,\Lambda,\delta)\left(\int_\Delta\Delta K^{\partial\Lambda}(x,x)dx+\left(\int_\Delta K^{\partial\Lambda}(x,x)dx\right)^\beta\right), \quad (3.11)$$

for any $\beta<1$.

Proof: By elementary calculations we have

$$\Delta K^{\partial\Lambda}(x,x)+2K^{\partial\Lambda}(x,x) = 2\int_\Lambda\int_\Lambda(\nabla_xP^{\partial\Lambda}(x,z_1))S(z_1,z_2)(\nabla_xP^{\partial\Lambda}(x,z_2))dz_1\,dz_2, \quad (3.12)$$

for $x\notin\partial\Lambda$.

Moreover, $\Delta K^{\partial\Lambda}(x,x)$ still has exponential decay in the $\text{dist}(x,\partial\Lambda)$ argument. Therefore, we may again apply the Chebyshev inequality in the spirit of the proof of Lemma 3.1. Q.E.D.

The unpleasant feature of the obtained estimates is the *a priori* dependence of the constants C_3, C_4, and C_5 of Λ. However, this dependence is not very essential as the following estimate shows.

Estimate: Take $\mu\in\mathcal{G}_r^t(z)$ and let $\Delta\in c(R^2)$ be given. Let $\{\Lambda_n\}$ be any sequence of bounded subsets of R^2 with C^1-piecewise boundaries and such that $\Lambda_n\uparrow R^2$ monotonously and by inclusion. There exists a subsequence $(n')\subset(n)$ and a constant $D(\eta,\beta)$ finite for μ a.e. η such that, for all $\beta<1$,

$$\int_\Delta(\Psi_\eta^{\partial\Lambda_{n'}}(x))^2\,dx\leqslant D(\eta,\beta)\left[\int_\Delta K^{\partial\Lambda}(x,x)dx\right]^\beta. \quad (3.13)$$

Proof: We use again the Chebyshev inequality. Let us take $\rho>0$ to be arbitrary. Then

$$\mu\left\{\eta\in S'(R^2)\,\middle|\,\int_\Delta(\Psi_\eta^{\partial\Lambda_{n'}}(x))^2\,dx\geqslant\frac{1}{\delta}\left[\int_\Delta K^{\partial\Lambda_{r'}}(x,x)dx\right]^\beta\right\}$$

$$\leqslant\sum_{n'}\mu\left\{\eta\in S'(R^2)\,\middle|\,\int_\Delta(\Psi_\eta^{\partial\Lambda_{n'}}(x))^2\,dx\geqslant\frac{1}{\delta}\left[\int_\Delta K^{\partial\Lambda_{n'}}(x,x)dx\right]^\beta\right\}$$

$$\leqslant\delta\sum_{n'}\frac{\int d\mu(\eta)\left[\int_\Delta(\Psi_\eta^{\partial\Lambda_{n'}}(x))^2\,dx\right]}{\left[\int_\Delta K^{\partial\Lambda_{n'}}(x,x)dx\right]^\beta}\leqslant\delta\sum_{n'}\left[\int_\Delta K^{\partial\Lambda_{n'}}(x,x)dx\right]^{1-\beta}.$$

A typical contribution of the integrals to the sum $\Sigma_{n'}$ is bounded by $O(1)\exp-\text{dist}(\Delta,\partial\Lambda_{n'})$. Let us denote $\alpha_{n'}=\text{dist}(\Delta,\partial\Lambda_{n'})$. Applying the root criterion we easily conclude that the series $\Sigma_{n'}$ is convergent whenever

$$\liminf_{n'}(\alpha_{n'}/n')>0.$$

From the assumptions made on the sequence $\{\Lambda_n\}$ it follows that a subsequence of that type $(n')\subset(n)$ may always be chosen. Because δ is arbitrary the proof follows. Q.E.D.

The value of the proved estimate is the following one. In some situations we know from the very beginning that the thermodynamic limits of some quantities of interest do exist. Therefore, it is enough to control these thermodynamic limits by passing to an arbitrary subsequence. It follows from the proof that the most natural case of the applications is the case when $\Lambda_n \uparrow R^2$ in the sense of Fisher.

Remark: There exists the corresponding version of this estimate for the case of completely regular measures. In particular, they have been applied in Ref. 34 to prove convergence of the high-temperature cluster expansion in the $P(\varphi)_2$ models (however, nonuniform in the boundary data). In this paper, we will not use it, therefore we will not write them explicitly.

The following lemma also shows mild dependence on the volume $|\Lambda|$ of the constants C_3, C_4, and C_5 in the above proved lemmas.

Lemma 3.5: Let $\{\Lambda_n\}$ be any sequence of bounded subsets of R^2 with piecewise-C_1 boundaries $\{\partial\Lambda_n\}$ and such that $\Lambda_n\uparrow R^2$ monotonously and by inclusion.

(1) Let $\mu\in\mathscr{G}^t_r(z)$ and let a number $\rho>0$ be given. Then, there exists a subsequence $(n')\subset(n)$ and a function $C_6(\eta,\rho)$ finite μ-a.e. and such that

$$\int_{(\partial_1\Lambda_{n'})}(\Psi_\eta^{\partial\Lambda_{n'}}(x))^2\,dx<C_6(\eta,\rho)|\partial\Lambda_{n'}|^{1+\rho}, \tag{3.14}$$

where

$$\partial_1\Lambda=\{x\in\Lambda|\operatorname{dist}(x,\partial\Lambda)<1\}.$$

(2) Let $\mu\in\mathscr{G}_r(z)$ and let a number $\rho>0$ be given. Then, there exists a constant $C_7(\eta,\rho)$ finite μ-a.e. and a subsequence $(n')\subset(n)$ such that

$$\int|\nabla\Psi_\eta^{\partial\Lambda_{n'}}|^2(x)dx<C_7(\eta,\rho)|\partial\Lambda_{n'}|^{1+\rho}. \tag{3.15}$$

Proof: Estimates (3.14) and (3.15) are obtained rigorously again by the application of the Chebyshev inequality and the assumed regularity of μ. Instead of writing the formal proofs in detail, we explain why these estimates are true. Taking $\epsilon>0$ we have

$$\begin{aligned}&\int\mu(d\eta)\,\frac{\int_{\partial_1\Lambda_n}(\Psi_\eta^{\partial\Lambda_n}(x))^2\,dx}{|\partial\Lambda_n|^{1+\epsilon}}\\&<|\partial\Lambda_n|^{-1-\epsilon}\int_{\partial_1\Lambda_n}dx\int\mu(d\eta)(\Psi_\eta^{\partial\Lambda_n}(x))^2\\&<\text{const}|\partial\Lambda_n|^{-\epsilon}.\end{aligned}$$

The last estimate follows from the well-known fact that there exists a constant c such that for every $\Delta_j\in c(R^2)$ we have $\|K^{\partial\Lambda}\|_{L^1(\Delta)}<c$ (see Proposition 7.8.7 in Ref. 3). Using additionally formula (3.14) the evidence of the validity of (3.15) can be seen by similar arguments. Q.E.D.

For the completely regular measures we note the following estimates.

Lemma 3.6: Let $\{\Lambda_n\}$ be as in Lemma 3.5. Assume that $\mu\in\mathscr{G}^t_{cr}(z)$ and let $\rho>0$ be given. For every integer $k>1$, there exists a constant $C_8(\eta,\rho,k)$ finite μ-a.e. and a sequence $(n')\subset(n)$ such

$$\int_{\partial_1\Lambda_{n'}}|\Psi_\eta^{\partial\Lambda_{n'}}(x)|^k\,dx<C_8(\eta,\rho,k)|\partial\Lambda_{n'}|^{1+\rho}. \tag{3.16}$$

C. Shift transformation

Now we are ready to demonstrate that the effect of the conditioning is a typically bondary effect and in the case of pressure it vanishes in the thermodynamic limit.

For a given bounded $\Lambda\subset R^2$ with C^1-piecewise boundary $\partial\Lambda$, let us denote (here $0<\epsilon<1$)

$$Y=\{x\in\Lambda|\operatorname{dist}(x,\partial\Lambda)>1\},$$
$$Y_\epsilon=\{x\in\Lambda|\operatorname{dist}(x,Y)<\epsilon\},$$
$$Y^\epsilon=\Lambda-Y_\epsilon.$$

Let $\chi_\epsilon(x)$ be a function (indexed by Λ) such that

$$\chi_\epsilon\in C_0^\infty(R^2),\qquad 0<\chi_\epsilon(x)\begin{cases}=1, & x\in Y,\\ <1, & x\in Y_\epsilon,\\ =0, & x\in Y^\epsilon.\end{cases} \tag{3.17}$$

and such that

$$\begin{aligned}&\sup_{x\in\Lambda}\max\{|\partial_1\chi^\epsilon|(x),|\partial_2\chi_\epsilon|(x)\}<C_{10}(\Lambda)<\infty,\\&\sup_{x\in\Lambda}|\Delta\chi^\epsilon|=C_{11}(\Lambda)<\infty.\end{aligned} \tag{3.18}$$

In the formula defining Z^η_Λ let us perform the following shift transformation:

$$\varphi\to\varphi-\chi_\epsilon\Psi_\eta^{\partial\Lambda}.$$

Using

$$\frac{d\mu_0^{\partial\Lambda}(\varphi-\chi_\epsilon\Psi_\eta^{\partial\Lambda})}{d\mu_0^{\partial\Lambda}(\varphi)}=\exp(-\varphi(J_\eta^\epsilon))\exp\frac{1}{2}\int\chi_\epsilon\Psi_\eta^{\partial\Lambda}(x)J_\eta^\epsilon(x)dx, \tag{3.19}$$

where $J_\eta^\epsilon(x)\equiv(-\Delta+1)(\chi_\epsilon\Psi_\eta^{\partial\Lambda})(x)$ is given by

$$J_\eta^\epsilon(x)=\begin{cases}0, & x\in Y,\\ (-\Delta\chi_\epsilon)\Psi_\eta^{\partial\Lambda}(x)+2(\nabla\chi_\epsilon)(x)(\nabla\Psi_\eta^{\partial\Lambda})(x), & x\in Y_\epsilon-Y,\\ 0, & x\in Y^\epsilon.\end{cases} \tag{3.20}$$

Note that because $\overset{\mu}{\underset{\eta}{\forall}}$: $\Psi_\eta^{\partial\Lambda}$ is a C^∞ function inside Λ as it is a solution (in S') of the elliptic homogeneous equation

$(-\Delta+1)\Psi_\eta^{\partial\Lambda}(x)=0$. This is a reason why the transformation made above has a perfectly correct mathematical sense.

Using these formulas we have

$$\frac{Z_\Lambda^\eta(z)}{Z_\Lambda^0(z)}=\exp\Big(\frac{1}{2}\int\chi_\epsilon(x)\Psi_\eta^{\partial\Lambda}(x)J_\eta^\epsilon(x)dx\Big)$$
$$\times\Big\langle\exp\Big[z\int\big[c(\varphi+(1-\chi_\epsilon)\Psi_\eta^{\partial\Lambda})(x)-c(\varphi)(x)\big]\Big]dx\exp(-\varphi(J_\eta^\epsilon))\Big\rangle_\Lambda(z). \tag{3.21}$$

By the application of the Cauchy–Schwartz inequality we have

$$\frac{Z_\Lambda^\eta(z)}{Z_\Lambda^0(z)}<\Pi_\Lambda^1(\eta)(\Pi_\Lambda^2(\eta))^{1/2}(\Pi_\Lambda^3(\eta))^{1/2}, \tag{3.22}$$

where we have a defined

$$\Pi_\Lambda^1(z)=\exp\Big(\frac{1}{2}\int\chi_\epsilon(x)\Psi_\eta^{\partial\Lambda}(x)J_\eta^\epsilon(x)dx\Big), \tag{3.23}$$

$$\Pi_\Lambda^2(z)=\Big\langle\exp\Big[2z\int_{\Lambda-Y}(c(\varphi+(1-\chi_\epsilon)\Psi_\eta^{\partial\Lambda})(x)-c(\varphi))dx\Big\rangle_\Lambda(z), \tag{3.24}$$

$$\Pi_\Lambda^3(z)=\langle\exp-2\varphi(J_\eta^\epsilon)\rangle_\Lambda(z). \tag{3.25}$$

Now we prove that all these factors have a typical behavior like $\exp O(\eta)|\partial\Lambda|$.

In the next three lemmas we assume that $\{\Lambda_n\}$ is a sequence as described in the Lemma 3.5 above. Additionally for a given sequence $\{\Lambda_n\}$ we choose a sequence χ_ϵ^n such that

$$C_{10}(\epsilon)=\sup_n C_{10}(\Lambda_n)<\infty$$

and (3.26)

$$C_{11}(\epsilon)=\sup_n C_{11}(\Lambda_n)<\infty.$$

Lemma 3.7: Let $\{\Lambda_n\}$ be as above. Let $\mu\in\mathscr{G}_r^t(z)$ and $\rho>0$ be given. There exists a constant $C_{12}(\eta)$ finite on supp μ and a subsequence $(n^1)\subset(n)$ such that

$$|\Pi_{\Lambda_{n'}}^1(\eta)|<\exp C_{12}(\eta,\rho)|\partial\Lambda_{n'}|^{1+\rho}. \tag{3.27}$$

Proof: It is due to the factor J_η^ϵ in the integral over dx that this integration is made over the set $Y_\epsilon-Y\equiv\partial_\epsilon^t\Lambda$. Using the definition of J_η^ϵ given by (3.20), the properties (3.18) of χ_ϵ, and the Cauchy–Schwartz inequality, we have

$$\Pi_{\Lambda_n}^1(\eta)<\exp\frac{1}{2}C_9C_{11}\int_{\partial_\epsilon^t\Lambda_n}(\Psi_\eta^{\partial\Lambda_n}(x))^2dx$$
$$\times\exp\frac{1}{4}C_9C_{10}\Big(\int_{\partial_\epsilon^t\Lambda_n}(\nabla\Psi_\eta^{\partial\Lambda_n}(x))^2dx\Big)^{1/2}$$
$$\times\exp\frac{1}{2}C_9C_{10}\Big(\int_{\partial_\epsilon^t\Lambda_n}\Psi_\eta^{\partial\Lambda_n}(x)dx\Big)^{1/2}. \tag{3.28}$$

Given the sequence $\{\Lambda_n\}$ as in the assumptions and using then Lemma 3.5, we conclude that there exists a subsequence $(n')\subset(n)$ such that

$$\Pi_{\Lambda_{n'}}^1(\eta)<\exp C_{12}(\eta,\rho)|\partial\Lambda_{n'}|^{1+\rho}. \tag{3.29}$$

Q.E.D.

Lemma 3.8: Let $\{\Lambda_n\}$ be as above. For any $\mu\in\mathscr{G}_r^t(z)$, $\rho>0$ there exists a constant $C_{13}(\eta,\rho)$ finite μ-a.e. and a subsequence $(n')\subset(n)$ such that

$$|\Pi_{\Lambda_{n'}}^3(\eta)|<\exp C_{13}(\eta,\rho)|\partial\Lambda_{n'}|^{1+\rho}. \tag{3.30}$$

Proof: From the correlation inequality (2.1) it follows that for every $f\in H_{-1}(R^2)$ we have

$$\langle e^{\varphi(f)}\rangle_\Lambda(z)<\exp\tfrac{1}{2}\|f\|_{-1,\partial\Lambda}^2, \tag{3.31}$$

uniformly in the volume Λ. Applying this observation we have

$$|\Pi_{\Lambda_n}^3(\eta)|=\langle\exp-2\varphi(J_\eta^\epsilon)\rangle_\Lambda(z)$$
$$<\exp(2\|(-\Delta^{\partial\Lambda}+1)(\chi_\epsilon\Psi_\eta^{\partial\Lambda})\|_{-1,\partial\Lambda}^2$$
$$=\exp2\int dx\,J_\eta^\epsilon(x)(\chi_\epsilon\Psi_\eta^{\partial\Lambda})(x).$$

We see now that the integral to be estimated is an almost identical to that met in the Lemma 3.7. Q.E.D.

We proceed now to estimate the factor $\Pi_\Lambda^2(\eta)$. Here, we use the $\cos\epsilon\varphi$ bound of Fröhlich.[22] From the use of this bound there follows our technical restriction on the size of ϵ. The $\cos\epsilon\varphi$ bound says that for every regular f we have

$$\forall_{0<\theta<2\pi}\quad\langle e^{c(\varphi+\theta)(f)}\rangle_\Lambda(z)<e^{(\|f\|_1+\|f\|_{p(\epsilon)})}, \tag{3.32}$$

uniformly in the volume Λ. Here we have to assume that

$$p(\epsilon)>1/(1-\epsilon^2/4\pi). \tag{3.33}$$

Lemma 3.9: Let $\{\Lambda_n\}$ be sequence as in the Lemma 3.5. Let $\mu\in\mathscr{G}_r^t(z)$. There exists a constant $C_{14}(\eta)$ finite μ-a.e., such that

$$\Pi_{\Lambda_n}^2(\eta)<\exp C_{14}(\eta)|\partial\Lambda_{n'}|. \tag{3.34}$$

Proof: By a little algebra we have

$$\Pi_{\Lambda_n}^2(\eta)<\Big(\Big\langle\exp\Big(2z\int_{\Lambda_n}dx{:}c(\varphi){:}_{\partial\Lambda_n}(x){:}c(1-\chi_\epsilon^n\Psi_\eta^{\partial\Lambda_n}){:}(x)-1\Big)\Big\rangle_\Lambda(z)\Big)^{1/2}$$
$$\times\Big\langle\exp-2z\int_{\Lambda_n}dx{:}s(\varphi){:}_{\partial\Lambda_n}{:}s(\varphi-\chi_\epsilon^n\Psi_\eta^{\partial\Lambda_n}){:}(x)\Big\rangle_\Lambda^{1/2}(z). \tag{3.35}$$

The functions

$$:c(1-\chi_\epsilon^n\Psi_\eta^{\partial\Lambda_n})(x):-1$$
$$=\exp((\alpha^2/2)(1-\chi_\epsilon^n)^2K^{\partial\Lambda}(x,x))$$
$$\times\cos\epsilon x(1-\chi_\epsilon^n)\Psi_\eta^{\partial\Lambda_n}(x)-1 \qquad (3.36)$$

and

$$:s(1-\chi_\epsilon^n\Psi_\eta^{\partial\Lambda_n}):(x)$$
$$=\exp((\epsilon x^2/2)(1-\chi_\epsilon^n)^2K_\eta^{\partial\Lambda_n}(x,x))$$
$$\times\sin\epsilon x(1-\chi_\epsilon^n)\Psi_\eta^{\partial\Lambda_n}(x) \qquad (3.37)$$

are both supported on the set $\Lambda - Y$ and are bounded there by

$$|:c(1-\chi_\epsilon^n)\Psi_\eta^{\partial\Lambda_n}(x):-1|$$
$$\leqslant 2(1-\chi_\epsilon^n)\exp(\epsilon x^2/2)K^{\partial\Lambda_n}(x,x) \qquad (3.38)$$

and

$$|:s(1-\chi_\epsilon^n)\Psi_\eta^{\partial\Lambda_n}(x)|$$
$$\leqslant(1-\chi_\epsilon^n)(x)\exp(\epsilon x^2/2)K^{\partial\Lambda}(x,x). \qquad (3.39)$$

Note that we have changed ϵ in formula (3.35) defining interaction by α in order to exclude the possible missing of the symbols used.

The functions $K^{\partial\Lambda}(x,x)$ have locally integrable singularities on the set $\partial\Lambda$. They have the behavior like

$$K^{\partial\Lambda}(x,x)\sim-(1/2\pi)\ln|\mathrm{dist}(x,\partial\Lambda)|e^{-\mathrm{dist}(x,\partial\Lambda)} \qquad (3.40)$$

as $x\to\partial\Lambda$. In applying the $\cos\epsilon\varphi$ bound we need to have $\alpha^2<2/(1-1/2\pi)$. Assuming this holds, we can apply Lemma 3.6 to both the factors in the estimate (3.35). Q.E.D.

D. $p_\infty^\eta = p_\infty^0$

Summarizing our discussion we have the following theorem.

Theorem 3.10: Let $\epsilon^2<2/(1-1/2\pi)$ and $\mu\in\mathscr{G}_r^t(z)$. Then

$$\overset{\mu}{\underset{\eta}{\forall}}:\ \lim_{\Lambda\uparrow R^2}p_\Lambda^\eta(z)=p_\infty^\eta(z)=p_\infty^0(z). \qquad (3.41)$$

Here $\Lambda\uparrow R^2$ means any sequence $\{\Lambda_n\}$ of bounded, with C^1-piecewise boundaries subsets of R^2 such that $\Lambda_n\uparrow R^2$ monotonously and by inclusion and such that, for some $\rho>0$,

$$\lim_{n\uparrow\infty}(|\partial\Lambda_n|^{1+\rho}/|\Lambda_n|)=0.$$

Proof: From Lemma 3.1 we know that there exists for μ a.e. η a subsequence $(n')\subset(n)$ such that the limit

$$\Theta_{(n')}^\eta=\lim_{n'\uparrow\infty}\ln(Z_{\Lambda_{n'}}^\eta(z)/Z_{\Lambda_{n'}}^0(z))^{-1/|\Lambda_{n'}|}$$

exists. From formula (3.21) and the Lemmas 3.7–3.9 it follows that for μ a.e. η we have $\theta_{n'}^\eta=0$. From this we conclude that $p_\infty^0(z)$ is the only accumulation point of the sequence $\{p_{\Lambda_n}^\eta(z)\}$. Q.E.D.

IV. COMPLETING OF THE PROOF OF THEOREM 1

The sequence of moments $\{C_\infty^0(x_1,\dots,x_n)\}_{n=1,\dots}$ does not describe fully the measure μ_∞ but rather its restriction to the even part of the σ algebra $\Sigma_\mu(R^2)$ only. From the independence of the boundary conditions of the moments $\{c_\infty^\eta(x_1,\dots,x_n)\}_{n=1,\dots}$ it follows that every even function of the field φ does not depend on the boundary conditions. In particular, taking an arbitrary sequence $f_1,\dots,f_n\in S(R^2)$ we conclude that the moments

$$C_\infty^\eta(f_1,\dots,f_n|(x_1,\dots,x_n))$$
$$=\lim_{\Lambda\uparrow R^2}\int d\mu_\Lambda^\eta(\varphi)\prod_{i=1}^n:\cos\epsilon\varphi_{f_i}:(x_i), \qquad (4.1)$$

where

$$:\cos\epsilon\varphi_{f_i}:(x_i)=\exp((\epsilon^2/2)(S*f_i)(0))$$
$$\times\cos\epsilon(\varphi*f_i)(x_i) \qquad (4.2)$$

does not depend on the given η if $c_\infty^\eta(x_1,\dots,x_n)$ does not depend on η. By the arguments identical to those used in the proof of Proposition 2.4 we have the following correlation inequality.

Lemma 4.1: Let $\mu\in\mathscr{G}_r^t(z)$. Take $f_1,\dots,f_n\in\mathscr{D}(R^2)$ and $\alpha_1,\dots,\alpha_n\in[0,2\pi)$ arbitrary. Then for μ a.e. η the following correlation inequality holds:

$$\left|\left\langle\prod_{i=1}^n:\cos(\epsilon\varphi_{f_i}+\alpha_i)(x_i):\right\rangle_\infty^\eta\right|$$
$$\leqslant\left\langle\prod_{i=1}^n:\cos\epsilon\varphi_{f_i}:(x_i)\right\rangle_\Lambda^0. \qquad (4.3)$$

We note the following lemma also.

Lemma 4.2: Take $\mu\in\mathscr{G}_r^t(z)$ and $f_1,\dots,f_n$ as in Lemma 4.1. Then for μ a.e. η we have

$$\left\langle\prod_{i=1}^n:\cos\epsilon\varphi_{f_i}:(x_i):\sin\epsilon\varphi_{f_{n+1}}:(y)\right\rangle_\infty^\eta=0, \qquad (4.4)$$

assuming that $\langle c(\varphi)(x)\rangle_\infty^\eta=\langle c(\varphi(0)\rangle_\infty^0$ holds.

Proof: Let us choose an arbitrary number $\alpha\in[0,2\pi)$. Applying Lemma 4.1 and Corollary 2.3, we obtain

$$\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)s(\varphi_{f_{n+1}}\pm\alpha(y)\right\rangle_\infty^\eta$$
$$=\cos\alpha\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)c(\varphi_{f_{n+1}})(y)\right\rangle_\infty^\eta$$
$$\mp\sin\alpha\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_ic(\varphi_{f_{n+1}})(y)\right\rangle_\infty^\eta$$
$$=\cos\alpha\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)c(\varphi_{f_{n+1}})(y)\right\rangle_\infty^0$$
$$\mp\sin\alpha\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)s(\varphi_{f_{n+1}})(y)\right\rangle_\infty^\eta$$
$$\leqslant\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)c(\varphi_{f_{n+1}})(y)\right\rangle_\Lambda^0.$$

Taking $\alpha\in(0,\pi/2)$ we have

$$\pm\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)c(\varphi_{f_{n+1}})(y)\right\rangle_\infty^\eta$$
$$\leqslant\frac{1-\cos\alpha}{\sin\alpha}\left\langle\prod_{i=1}^n c(\varphi_{f_i})(x_i)c(\varphi_{f_{n+1}})(y)\right\rangle_\Lambda.$$

Letting $\alpha\downarrow0$ we get the result. Q.E.D.

From this we easily obtain the following corollary.

Corollary 4.3: Let $\mu\in\mathscr{G}_r^t(z)$ and let $f_1,\dots,f_n\in\mathscr{D}(R^2)$.

Then assuming $\langle c(\varphi)(x)\rangle_\infty^\eta = \langle c(\varphi)(0)\rangle_\infty^0$, we have

$$\left\langle \prod_{i=1}^m s(\varphi_{f_i})(x_i)\right\rangle_\infty^\eta = \left\langle \prod_{i=1}^m s(\varphi_{f_i})(x_i)\right\rangle_\infty^0 \tag{4.5}$$

and for arbitrary $g_1,\dots,g_m \in \mathscr{D}(R^2)$ we have

$$\left\langle \prod_{i=1}^n c(\varphi_{f_i})(x_i) \prod_{j=1}^m s(\varphi_{g_j})(y_j)\right\rangle_\infty^\eta$$

$$= \left\langle \prod_{i=1}^n c(\varphi_{f_i})(x_i) \prod_{j=1}^m s(\varphi_{g_j})(y_j)\right\rangle_\Lambda, \tag{4.6}$$

assuming the equality $\langle c(\varphi)(x)\rangle_\infty^\eta = \langle c(\varphi)(0)\rangle_\infty^0$.

Proof: From the repeated use of the formula

$$\sin\alpha \sin\beta = \tfrac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)]$$

and application of Lemma 4.2 and Corollary 2.3 the proof follows easily. Q.E.D.

Finally, we are ready to prove Theorem 1.

Proof of Theorem 1: It is well known that $p_\infty^0(z)$ is a concave function of the coupling constant z. From this it follows (see Refs. 35 and 36) that $p_\infty^0(z)$ is almost everywhere a differentiable function (except at most a countable set of values) and that we have the equality

$$\frac{d}{dz}\lim_{\Lambda\uparrow R^2} p_\Lambda^0(z) \equiv \lim_{\Lambda\uparrow R^2}\frac{d}{dz} p_\Lambda^0(z),$$

at the points of differentiability. The arguments of Sec. III can easily be extended to treat the following perturbed pressure:

$$p_\Lambda^\eta(z,\lambda) = -\frac{1}{|\Lambda|}\ln\int \exp\left(\lambda\int c(\varphi)(x)dx\right) \times\exp\left(z\int_\Lambda c(\varphi+\Psi_\eta^{\partial\Lambda})(x)dx\,\partial\Lambda(d\varphi)\right)\mu_0. \tag{4.7}$$

In particular, we obtain that the unique thermodynamic limit

$$p_\infty^\eta(z,\lambda) = \lim_{\Lambda\uparrow R^2} p_\Lambda^\eta(z,\lambda) \tag{4.8}$$

exists (whenever $\Lambda\uparrow R^2$ as in Sec. III) and is independent of the typical boundary condition η. Moreover, the limit is differentiable at λ and z almost everywhere. Assuming that $p_\infty^0(z)$ is differentiable at the point $z=z_0$ we obtain

$$\begin{aligned}\frac{d}{d\lambda}p_\infty^\eta(z_0,\lambda)\Big|_{\lambda=0} &= \frac{d}{d\lambda}p_\infty^0(z_0,\lambda)\Big|_{\lambda=0} = \frac{d}{dz}p_\infty^0(z)\Big|_{z=z_0}\\ &= \lim_{\Lambda\uparrow R^2}\frac{1}{|\Lambda|}\int_\Lambda\langle c(\varphi)(x)\rangle_\infty(z_0) = \langle c(\varphi)(0)\rangle_\infty(z_0)\\ &= \lim_{\Lambda\uparrow R^2}\frac{d}{d\lambda}p_\Lambda^\eta(z_0,\lambda)\Big|_{\lambda=0}\\ &= \lim_{\Lambda\uparrow R^2}\frac{1}{|\Lambda|}\int_\Lambda\langle c(\varphi)(x)\rangle_\Lambda^\eta dx,\end{aligned} \tag{4.9}$$

which shows that $p_\infty^\eta(z_0,\lambda)$ is then differentiable at the point $\lambda=0$ and that [assuming the limiting measure $\langle\ \rangle_\infty^\eta(z_0)$ has a translationally invariant first moment] the following equality holds:

$$\langle c(\varphi)(0)\rangle_\infty^0(z_0) = \langle c(\varphi)(x)\rangle_\infty^\eta(z_0). \tag{4.10}$$

Thus, the bootstrap principle of Corollary 2.3 then is applicable with the result that if μ is any regular measure then for μ a.e. $\eta\in S'(R^2)$ and $n\geqslant 1$, we have

$$\left\langle\prod_{i=1}^n c(\varphi)(f_i)\right\rangle_\infty^\eta(z_0) = \left\langle\prod_{i=1}^n c(\varphi)(f_i)\right\rangle_\infty^0(z_0). \tag{4.11}$$

For the regularized moments of the form

$$\left\langle\prod_1^n c(\varphi_{f_i}(x_i)\prod_{j=1}^m c(\varphi_{g_j})(y_j)\right\rangle_\infty^\eta$$

we then apply Corollary 4.3. The limits $f_i\downarrow\delta_j$ $g_j\downarrow\delta$ of both the sides of equality (4.6) can easily be controlled by the application of the $\cos\epsilon\varphi$-type of bounds for the measure $\mu_\infty^0(d\varphi)$. Q.E.D.

Remark 4.1: Taking into account Remark 3.1 we can eliminate the λ-perturbation argument used above.

V. CONCLUDING REMARKS

The main motive for writing this paper is the question about the global Markov property for the two-dimensional scalar fields. In the case of lattice systems some results concerning this problem have been obtained in Refs. 37 and 38. The main strategy coming back to Preston,[1] and Fölmer[39] is to introduce a certain order (the FKG order) into the set of Gibbs measures. Some simplifications have been made in the paper by Goldstein.[40] The method of this paper combined with the superstability estimates has been applied by the author to show the global Markov property also for some nonferromagnetic continuous spin systems in Ref. 41.

One of the main obstacles to applying immediately the techniques of the FKG order to the continual case is that we do not know whether such an order can be defined in space of the Gibbs measures describing the continual fields. The intriguing question is to find a suitable notion of the lattice regularization which discretizes the Dirichlet problem (1.3) in a proper sense by which we mean, first, that the discrete versions of the corresponding local specifications are convergent surely to the continual one, and second, the shift transformation exists which transform the discrete versions of local specifications to the forms considered in Refs. 37 and 38. Then, the FKG order may be induced into the set of the continual Gibbs measures on account of the assured convergence. But we have not checked any details of this intriguing program.

On the other hand, the methods of the present paper do not use any kind of ferromagnetic properties of the continual fields. Therefore, they seem to be very useful in the study of the DLR equations for continual fields that are defined by the trigonometric perturbations of Gaussian, generalized fields. Such an analysis has been performed by the author in Refs. 42–44.

ACKNOWLEDGMENTS

I would like to thank Professor R. L. Dobrushin for his kind invitation at the Sixth International Symposium on Information Theory held in September, 1984 at Tashkent, Uzbek SSR, where the result has been presented for the first

time. I would like to thank Professor S. Albeverio for his interest in this work.

APPENDIX

In this Appendix we review some results obtained by Fölmer[7] and Weizsacker and Winkler,[8,9] which are relevant for us. Related results can be found also in the second chapter of Preston.[1]

Let $\{J,\alpha\}$ be an increasing net that is countably generated. Assume that $\{\Omega,\Sigma\}$ is a standard Borel space and that for every $i \in J$ there is a sub-σ-algebra Σ_i of Σ such that $i\alpha j \Rightarrow \Sigma_j \subset \Sigma_i$. A collection V of stochastic kernels $\{p_i, i\in J\}$ from $\{\Omega,\Sigma_i\}$ to $\{\Omega,\Sigma\}$ is called *specification* iff

(s1) $\forall_{i\in J} \forall_{F\in\Sigma} P_i(\cdot,F)$ is Σ_i measurable,

(s2) $\forall_{i\in J} \forall_{F\in\Sigma_i} P_i(\cdot,F) = 1_F$,

(3) $i\alpha^j \Rightarrow P_j P_i = P_j$.

A probability measure μ on $\{\Omega,\Sigma\}$ is called the Gibbs state corresponding to the given specification V iff it satisfies the DLR equations

$$\text{(DLR)} \quad \forall_{i\in J} \mu\circ P_i = \mu.$$

We collect the fundamental results obtained in Refs. 7–9 and 44 in the following theorem.

Theorem A.1: There exist a standard Borel space $\{\Omega_\infty,\Sigma_\infty\}$ and a stochastic kernel P_∞ from Ω_∞ to Ω such that the mapping $\mu \to \mu P_\infty$ is an affine bijection from the set of probabilistic measures on $\{\Omega_\infty,\Sigma_\infty\}$ onto $G(V)$, in particular, $G(V) = \{\mu P_\infty | \mu$ runs over probabilistic measures on $\{\Omega_\infty,\Sigma_\infty\}\}$ and the extremal points of $G(V)$ (the so called Martin–Dynkin boundary)

$$\partial G(V) = \{P_\infty(\omega_\infty;—)|\omega_\infty \in \Omega_\infty\}.$$

The set $\partial G(V)$ of extremal points of $G(V)$ is measurable with respect to the evaluation σ algebra Σ and for each Gibbs state μ there is a *unique probabilistic* measure ρ on $\{\partial G(V), \partial G \cap \Sigma\}$ such that

$$\forall_{B\in\Sigma} \quad \mu(B) = \int_{\partial G(V)} \nu(B) d\rho(\nu).$$

For the application to field theory we put $\{\Omega,\Sigma\} = \{S'(R^2),B\}$, $\{J = \{\Delta_n\},C\}$ any monotone sequence of bounded regular subsets of R^2 tending to R^2 monotonously and by inclusion. Then $\Sigma_n = \Sigma(\Lambda_n^c)$.

[1]C. Preston, "Random fields," in *Lecture Notes in Mathematics*, Vol. 534 (Springer, Berlin, 1976).

[2]D. Ruelle, *Statistical Mechanics, Rigorous Results* (Benjamin, New York, 1969).

[3]J. Glimm and A. Jaffe, *Quantum Physics: A Functional Integral Point of View* (Springer, New York, 1981).

[4]R. L. Dubrushin and R. A. Minlos, in *Problems of Mechanics and Mathematical Physics* (Mir, Moscow, 1976) (in Russian).

[5]G. Benfatto, G. Gallavotti, and F. Nicollo, J. Funct. Anal. **36**, 343 (1980).

[6]S. Albeverio and R. Hoegh-Krohn, Commun. Math. Phys. **68**, 95 (1979).

[7]H. Fölmer, in "Seminaire de probabilities IX Universite de Smasbourg," *Lecture Notes in Mathematics*, Vol. 465 (Springer, Berlin, 1975), pp. 305–317.

[8]G. Winkler, Hab. Ph.D., University of Munich, 1983.

[9]H. V. Weizsacker and G. Winkler, in *Functional Analysis: Surveys and Recent Results. II, Mathematics Studies*, Vol. 38, (North-Holland, Amsterdam, 1980), pp. 115–148.

[10]B. Simon, *The $P(\varphi)_2$ Euclidean (Quantum) Field Theory* (Princeton U. P., Princeton, 1974).

[11]R. L. Dobrushin, Theor. Prob. Appl. **13**, 197 (1968) (in Russian).

[12]R. L. Dobrushin, Theory Prob. Appl. **15**, 458 (1970).

[13]M. Cassandro, E. Olivieri, A. Pellegninotti, and E. Pressuti, Z. Wahrscheinkeitsth. verw. Geb. **41**, 313 (1978).

[14]F. Guerra, L. Rosen, and B. Simon, Ann. Math. **101**, 111 (1975).

[15]J. Jmbrie, Commun. Math. Phys. **82**, 261 (1981).

[16]H. Aizenmann, Commun. Math. Phys. **73**, 83 (1980).

[17]Y. Higuchi, Colloq. Random Fields, Esztergom 1979 Colloq. Math. Soc. Janos Bolyai **27**, 517 (1981).

[18]R. Gielerak, J. Math. Phys. **24**, 247 (1983).

[19]R. Gielerak, in "Critical phenomena. Theoretical aspects," *Proceedings of the Poiana Brasov School, Progress in Physics*, Vol. 12 (Birkhauser, Boston, 1985).

[20]B. Zegarlinski, Ph.D. thesis, University of Wroclaw, 1984 (unpublished); Commun. Math. Phys. (1985).

[21]R. Gielerak and B. Zegarlinski, Fortsch. Phys. **1**, 1 (1984).

[22]J. Fröhlich, Commun. Math. Phys. **47**, 233 (1976).

[23]J. Fröhlich and Y. M. Park, Helv. Phys. Acta **50**, 315 (1978).

[24]J. Fröhlich and E. Seiler, Helv. Phys. Acta **49**, 889 (1976).

[25]J. Fröhlich and C.-E. Pfister, Commun. Math. Phys. **89**, 303 (1983).

[26]C.-E. Pfister, Commun. Math. Phys. **86**, 375 (1982).

[27]A. Messager, S. Miracle-Sole, and C.-E. Pfister, Commun. Math. Phys. **58**, 19 (1978).

[28]F. Guerra, L. Rosen, and B. Simon, Ann. Inst. H. Poincaré **15**, 223 (1975).

[29]R. Gielerak, "On the equilibrium equations for the change-symmetric continual systems," preprints P-17-85-770, 771, JINR, Dubna, 1985.

[30]R. Gielerak, to appear in Theor. Math. Phys. (in Russian).

[31]J. Fröhlich and Y. M. Park, Commun. Math. Phys. **59**, 235 (1978).

[32]K. Osterwalder and R. Schrader, Commun. Math. Phys. **31**, 281 (1973); **42**, 281 (1975).

[33]J. Ginibre, Commun. Math. Phys. **16**, 310 (1970).

[34]R. Gielerak, "Vacuum energy for $P(\varphi)_2$ theories: Independence of boundary conditions," to appear in Rep. Math. Phys.

[35]M. E. Fisher, J. Math. Phys. **6**, 1643 (1965), Lemma III.

[36]R. B. Griffiths, J. Math. Phys. **13**, 1215 (1964), Appendix A.

[37]J. Bellisard and R. Hoegh-Krohn, Commun. Math. Phys. **84**, 297 (1982).

[38]L. Picco, These 3 Marseille, 1979.

[39]H. Fölmer, in *Quantum Fields-Algebras-Processes*, edited by L. Streit (Springer, Wien, 1980), pp. 293–302.

[40]S. Goldstein, Commun. Math. Phys. **74**, 223 (1980).

[41]R. Gielerak, "Uniqueness for some (nonferromagnetic) continual spin systems," preprint JINR E-17-85-873, Dubna, 1985.

[42]R. Gielerak, "On the correlation inequalities in some models involving fermions," to appear in Fortsch. Phys.

[43]R. Gielerak, Communications JINR, E5-85-108, Dubna, 1985.

[44]R. Gielerak, "Trigonometric perturbations of the Gaussian generalized fields," to appear in Theor. Math. Phys.

[45]H. V. Weizsäcker and G. Winkler, Math. Ann. **246**, 23 (1979).

Approximate solution of Fredholm integral equations by the maximum-entropy method

Lawrence R. Mead
Department of Physics, University of Southern Mississippi, Southern Station Box 5046, Hattiesburg, Mississippi 39406

(Received 12 March 1986; accepted for publication 6 August 1986)

An approximate means of solving Fredholm integral equations by the maximum-entropy method is developed. The Fredholm integral equation is converted to a generalized moment problem whose approximate solution by maximum-entropy methods has been successfully implemented in a previous paper by Mead and Papanicolaou [L. R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984)]. Several explicit examples are given of approximate maximum-entropy solutions of Fredholm integral equations of the first and second kinds and of the Wiener–Hopf type. Both the weaknesses and strengths of the method are discussed.

I. FORMULATION OF THE MAXIMUM-ENTROPY METHOD OF SOLUTION

Consider an integral equation of the general type:

$$P(x) = \phi(x) - \int_a^b dy\, P(y)K(x,y)\,, \tag{1.1}$$

where $P(x)$ is the unknown function sought on the interval $[a,b]$, $\phi(x)$ is a given function, and $K(x,y)$ is a given kernel. When $[a,b]$ is finite in (1.1), we have a Fredholm equation of the second kind, when $[a,b] = [0,\infty]$, a Wiener–Hopf equation, and, if the left-hand side of (1.1) is replaced by zero, a Fredholm equation of the first kind.[1] As yet no restrictions are placed on the specific form of the kernel $K(x,y)$. Our strategy will be to convert (1.1) into an equivalent generalized moment problem whose maximum entropy solution has been studied in some depth by Mead and Papanicolaou.[2] To accomplish this transformation, let $M_n(x)$, $n = 0,1,2,\ldots$, be a linearly independent set of functions defined on the interval $[a,b]$ (possibly infinite). Multiply both sides of (1.1) by $M_n(x)$ and integrate both sides of the resulting equation with respect to x over the interval $[a,b]$. After some rearrangement, the new equation reads

$$\mu_n = \int_a^b dy\, P(y)G_n(y)\,, \tag{1.2}$$

where

$$\mu_n \equiv \int_a^b M_n(x)\phi(x)dx\,, \tag{1.3}$$

$$G_n(x) \equiv M_n(x) + \int_a^b dt\, K(t,x)M_n(t)\,, \tag{1.4}$$

and where, in (1.2)–(1.4), $n = 0,1,2,\ldots$. Equations (1.2) have the form of a generalized moment problem where the given generalized moments μ_n computed from (1.3) are generated by the set of known functions $G_n(x)$ defined by (1.4).

Some remarks are clearly in order at this point. First, in obtaining (1.2) from (1.1), an interchange in the order of two integrations is made. Henceforth, we assume that any singularities present in the kernel $K(x,y)$ are sufficiently weak to permit this interchange. This property of $K(x,y)$ will be explicit in the numerical examples of Sec. II (the case of singular kernels will not be investigated here). Second, it is judicious to choose a set of functions $M_n(x)$, such that integrals appearing in (1.3) and (1.4) [defining the μ_n and $G_n(x)$, respectively] may be performed analytically. Although, strictly speaking, this requirement is not necessary, the maximum-entropy algorithm for the solution of (1.2) described below is otherwise far less easily implemented. Before continuing the discussion, it is necessary to state the maximum-entropy approximation to the solution of (1.2).

Given the set of known, exact moments μ_n, $n = 0,1,2,\ldots,N$, determined from (1.3), the "optimal" function $P_N(x)$ satisfying the constraints (1.2), up to $n = N$ is found by maximizing the entropy functional[2]

$$S[P_N] = -\int_a^b dx\, P_N(x)[\ln P_N(x) - 1] + \sum_{i=0}^{N} \lambda_i \left[\mu_i - \int_a^b dx\, P_N(x)G_i(x)\right]. \tag{1.5}$$

Variation of the entropy S with respect to P_N and the λ_i yields a set of maximum-entropy (max-ent) equations

$$\mu_n = \int_a^b dx\, G_n(x)P_N(x)\,, \quad n = 0,1,2,\ldots,N\,, \tag{1.6}$$

where $P_N(x)$ is found to be

$$P_N(x) = \exp\left[-\sum_{i=0}^{N} \lambda_i G_i(x)\right]. \tag{1.7}$$

Equations (1.6) are just the moment conditions (1.2) determining the solution of the original integral equation (1.1) truncated to finite N. They are to be looked upon as a set of equations determining the unknown Lagrange multipliers λ_i, $i = 0,1,2,\ldots,N$, the knowledge of which fixes the solution $P_N(x)$ given by (1.7). Full details of the numerical algorithm used to find the λ_i may be found in Ref. 2, as well as several successful solutions of the moment problem in physical applications. Details will not be given here.

Two further important points need to be discussed. The maximum-entropy method assumes that one is seeking a probability distribution $P_N(x)$, satisfying the given constraints (1.6); that is, the max-ent method returns a $P_N(x)$

 0022-2488/86/122903-05$02.50

that is everywhere *non-negative* on the interval $[a,b]$. There is, of course, no guarantee that the original integral equation (from which the moment problem was derived) has any such solution. Nonetheless, we will seek a positive definite solution $P(x)$ to (1.1). If such a solution is not found, then it may be possible to transform the original integral equation into a new one having a positive definite solution by a change of variable. Further discussion of this point is deferred to the examples of Sec. II. Next, we must ask whether it will always be the case that the solution of the full ($N=\infty$) moment problem (1.2), also be a solution to the integral equation (1.1)? A partial answer to this question is available. For the case that $[a,b]$ is a finite interval (for example, $[0,1]$) and the $G_n(x)$ are x^n (or polynomials of degree n), Hausdorff has provided necessary and sufficient conditions on the moments μ_n such that a unique positive definite function $p(x)$ exists satisfying (1.2).[3] Thus, if Hausdorff's conditions are satisfied *and* (1.1) admits a unique positive definite solution, the two must coincide since the moment the problem is derived solely from the original integral equation. For the interval $[0,\infty]$, or for more general $G_n(x)$, numerical evidence must be relied upon at this stage.

II. NUMERICAL EXAMPLES

In this section a number of numerical examples are presented all but one of which have known, exact solutions. These equations are chosen so that the various possible difficulties pointed out in the previous section may be easily addressed. The max-ent solution $P_N(x)$, found from (1.6) and (1.7), must approximately satisfy the integral equation (1.1). The degree to which $P_N(x)$ satisfies (1.1) will be examined by evaluating for various x the left-hand side of (1.1) [just $P_N(x)$ itself], and the right-hand side of (1.1) with $P(x)$ replaced by $P_N(x)$. The latter will be denoted by $P_N^{\text{MAX}}(x)$. Notice that

$$P_N^{\text{MAX}}(x)=\phi(x)-\int_a^b dy\,K(y,x)P_N(y) \tag{2.1}$$

takes the form of [$\phi(x)$ plus] an average of $K(y,x)$ over $P_N(y)$. In Ref. 2 it is proved that averages of sufficiently well-behaved functions $F(x)$ over $P_N(x)$ converge to the exact average of $F(x)$ over the true $P(x)$; that is,

$$\langle F(x)\rangle=\int_a^b dx\,F(x)P(x)=\lim_{N\to\infty}\int_a^b dx\,F(x)P_N(x)\,. \tag{2.2}$$

Thus, assuming the moment problem (1.2) is equivalent to the original integral equation, the $P_N(x)$ and $P_N^{\text{MAX}}(x)$ ought to therefore agree for almost all x in $[a,b]$ in the limit as $N\to\infty$. Moreover, in Ref. 2 the pointwise convergence of $P_N(x)$ to $P(x)$ was examined in the context of the moment problem. It was found that while the pointwise convergence of $P_N(x)$ was slow (indeed, in some cases not even certain), averages of the form (2.2) were rapidly and smoothly convergent. It follows that for a given N, $P_N^{\text{MAX}}(x)$ defined by (2.1) is the optimal max-ent approximation to the solution of the integral equation.

The first example is the Fredholm equation

$$P(x)=-1+\tfrac{3}{2}\int_0^1 dy\,P(y)e^{|x-y|}\,, \tag{2.3}$$

with $\phi(x)=-1$, $K(x,y)=\frac{3}{2}e^{|x-y|}$. The exact solution of (2.3) is

$$\begin{aligned}P(x)&=-\tfrac{1}{4}+Ae^{2x}+Be^{-2x}\,,\\ A&=-\tfrac{1}{4}(1+e^2/3)(1-e^4/9)\,,\\ B&=e^2A\,.\end{aligned} \tag{2.4}$$

Choosing $M_n(x)=-x^n$, $n=0,1,2,\ldots$, the moments are found to be $\mu_n=1/(n+1)$. The $G_n(x)$ defined by (1.4) are nth-degree polynomials plus linear combinations of e^x and e^{-x}. These will not be displayed explicitly. The five moment ($N=4$) set of max-ent equations (1.6)–(1.7) are solved numerically by the procedure of Ref. 2. The resulting Lagrange multipliers, λ_i, $i=0,1,\ldots,4$, are given in Table I. Values of $P_4(x)$, $P_4^{\text{MAX}}(x)$, and the exact $P(x)$ are given in Table II for various x. Notice that the agreement between $P_N(x)$ and $P_N^{\text{MAX}}(x)$ (the approximate left- and right-hand sides of the integral equation, respectively) is only moderate (two significant figures). On the other hand, the agreement between $P_N^{\text{MAX}}(x)$ and the exact values of (2.4) are considerably improved (four to five significant figures) in line with the above remarks concerning averages.

The second example is a more difficult Fredholm equation studied previously by Guy *et al.*,[4]

TABLE I. The Lagrange multipliers λ_i computed for various integral equations.

	Equation number (2.3)	(2.5)	(2.10)	(2.14)
λ_0	0.524 704 574	− 0.108 510 936 + 01	− 0.424 341 962	0.363 033 198 + 01
λ_1	− 0.212 602 660 + 01	0.819 558 864 + 01	0.166 094 369 + 01	− 0.142 597 049 + 02
λ_2	0.983 172 725	− 0.745 595 374 + 02	− 0.139 898 559 + 01	0.244 451 821 + 02
λ_3	0.228 570 776 + 01	0.263 801 845 + 03	0.682 825 062	− 0.534 121 593 + 01
λ_4	− 0.114 285 388 + 01	− 0.380 322 311 + 03	− 0.147 264 389	− 0.453 974 814 + 01
λ_5		0.193 471 906 + 03	0.106 746 868 − 01	0.739 553 384 + 02
λ_6				− 0.541 204 432 + 02
λ_7				0.211 940 281 + 02
λ_8				− 0.430 448 522 + 01
λ_9				0.356 448 781

TABLE II. Five-moment max-ent solution of (2.3).[a]

x	$P_4(x)$	$P_4^{\rm MAX}(x)$	Exact
0.0	1.181 324	1.183 518	1.183 518
0.1	0.993 292	0.992 469	0.992 473
0.2	0.850 917	0.851 300	0.851 229
0.3	0.753 738	0.754 322	0.754 313
0.4	0.697 809	0.697 636	0.697 639
0.5	0.679 610	0.678 896	0.678 998

[a] Note that all functions are symmetric about the point $x = \frac{1}{2}$ on $[0,1]$.

$$P(x) = 1 - \int_0^1 dy\{2[1 + 3(x^2 + y^2)]^{1/2}\}P(y) \,. \quad (2.5)$$

Here, the kernel $K(x,y) = 2[1 + 3(x^2 + y^2)]^{1/2}$ has norm greater than 3 (see Ref. 4). Thus, the Neumann series[1] solution of (2.5) will fail to converge, whereas the Fredholm series[1] is cumbersome to compute. If we choose $M_n(x) = x^n$, then $\mu_n = 1/(n+1)$, $n = 0,1,2,\ldots$. The $G_n(x)$ may be exactly computed and are algebraic functions (n odd) or algebraic functions combined with logarithms (n even). The λ_i are again given in Table I. The approximate (six-moment) functions $P_5(x)$, $P_5^{\rm MAX}(x)$ are given for various x in Table III. In addition, values of $P(x)$ are tabulated from the numerical solution of (2.5) reported in Ref. 4. Two features are evident in Table III. The first is the property of the max-ent solution noted in the first example. The agreement between $P_5(x)$ and $P_5^{\rm MAX}(x)$ is poor, however, agreement between $P_5^{\rm MAX}(x)$ and the exact numerical values of $P(x)$ is marked (five significant figures). The second feature is that both $P_5^{\rm MAX}(x)$ and the exact $P(x)$ take on *negative* values near $x = 1$. This $P_5(x)$ is manifestly non-negative; it tries its best to become negative, however, dropping sharply to 10^{-5} at $x = 1$. That $P_5^{\rm MAX}(x)$ can become negative is due to the minus sign in (2.5).

It is appropriate at this point to discuss what one might do if the exact (and unknown) solution to an integral equation of interest happens to be negative over a significant part of the interval $[a,b]$. Consider, for example, the integral equation

TABLE III. Six-moment max-ent solution of (2.5).

x	$P_5(x)$	$P_5^{\rm MAX}(x)$	Exact[a]
0.0	0.499 835	0.410 047	0.410 043
0.1	0.369 716	0.403 856	0.403 852
0.2	0.400 899	0.385 688	0.385 685
0.3	0.384 521	0.356 634	0.356 633
0.4	0.308 136	0.318 170	0.318 169
0.5	0.241 952	0.271 848	0.271 185[b]
0.6	0.214 976	0.219 108	0.219 108
0.7	0.194 860	0.161 177	0.161 178
0.8	0.100 539	0.099 064	0.099 066
0.9	0.008 083	0.033 574	0.033 576
1.0	0.107 811 − 04	− 0.034 655	− 0.034 652

[a] See Ref. 4.
[b] This entry, as reported in Ref. 4, is probably a misprint.

$$P(x) = (1-x) + \int_0^1 (6y + x)P(y)dy \,, \quad (2.6)$$

whose exact solution is easily found to be

$$P(x) = \tfrac{1}{2} - x \,. \quad (2.7)$$

The solution (2.7) is negative over the bounded subinterval $[\frac{1}{2},1]$. Let us attempt a max-ent solution by setting $M_n(x) = 2x^n$, such that $\mu_n = 1/(n+1)(n+2)$, $n = 0,1,2,\ldots$. The first two $G_n(x)$ defined by (1.4) are

$$G_0(x) = 3 - 12x \,, \quad G_1(x) = -(\tfrac{1}{2} + 4x) \,. \quad (2.8)$$

It is immediately evident that $G_1(x)$ is negative over the entire interval $[0,1]$. Since $\mu_1 = \frac{1}{6}$ is positive, the max-ent equations (1.6) and (1.7) cannot have a solution for $n = 1$, $N \geqslant 1$. That one of the early moment-generating functions $G_n(x)$ for some n is everywhere negative is a sure sign that the exact solution of the integral equation is negative over significant regions of interest. If, as in our current example, $P(x)$ is *bounded* in this region, a simple change of variable will suffice to remedy the situation. In (2.6), let $Q(x) = P(x) + \frac{1}{2}$. The new integral equation then reads

$$Q(x) = -\frac{3x}{2} + \int_0^1 dy\, Q(y)(6y + x) \,. \quad (2.9)$$

Now, choosing $M_n(x) = -4/3x^n$, we have $\mu_n = 2/(n+2)$ and

$$G_n(x) = \frac{4}{3}\left[-x^n + \frac{6x}{n+1} + \frac{1}{n+2}\right].$$

All of these new $G_n(x)$, $n = 0,1,2,\ldots$, are positive definite functions on $[0,1]$. Hence, the max-ent algorithm will go through. Indeed, the five-moment approximation $Q_5^{\rm MAX}(x)$ to the solution of (2.9) is accurate to a remarkable 12 significant digits. If, then, the solution to the original integral equation is negative, but bounded, a simple change of variable will still allow an approximate solution by the max-ent method. In many cases even an unbounded and negative solution may be handled. Suppose $P(x) \geqslant 0$ and bounded on $[0,c]$, negative on $[c,1]$, and $P(x)$ approaches $-\infty$ as $x \to 1$. Then the change of variable $Q(x) = -P(x) + \alpha$, for some finite α, will transform the equation to one whose solution $Q(x)$ is positive definite on $[0,1]$.

The next example to be considered is the Wiener–Hopf integral equation

$$P(x) = e^{-|x|} - 4\int_0^\infty e^{-|y-x|}P(y)dy \,, \quad (2.10)$$

whose exact solution is

$$P(x) = \begin{cases} \frac{1}{2}e^{-3x}, & x \geqslant 0 \,, \\ \frac{1}{2}e^{x}, & x \leqslant 0 \,. \end{cases} \quad (2.11)$$

Since the solution is desired for all x, we first separate $P(x)$ as

$$P(x) = P^+(x) + P^-(x) \,, \quad (2.12)$$

where $P^+(x) = 0$ for $x < 0$ and $P^-(x) = 0$ for $x > 0$. With this separation (2.10) reads

$$P^+(x) = e^{-|x|} - 4\int_0^\infty e^{-|y-x|}P^+(y)dy \,, \quad x > 0 \,, \quad (2.13a)$$

$$P^-(x) = e^{-|x|} - 4\int_0^\infty e^{-|y-x|}P^+(y)dy\,, \quad x<0\,. \tag{2.13b}$$

Hence, the optimal max-ent solution (for given N), $P_N^{\rm MAX}(x)$, will be taken to be the right-hand side of (2.13a) with $P^+(x) = P_N^+(x)$, for $x>0$, and the right-hand side of (2.13b) [also with $P^+(x) = P_N^+(x)$] for $x<0$. In both cases (x positive or x negative) the associated moment problem will be derived from (2.13a). Choosing $M_n(x) = \frac{1}{2}x^n$, we find that $\mu_n = n!$, and that the $G_n(x)$ are polynomials plus a term proportional to e^{-x}. The numerical results for $N=5$ are tabulated in Table I (the λ_i) and Table IV [$P_5^{\rm MAX}(x)$ for various x]. Once again $P_5(x)$ ($x>0$) and $P_5^{\rm MAX}(x)$ show poor agreement, but $P_5^{\rm MAX}(x)$ and the exact $P(x)$ agree to three to four decimal places. That the few moment max-ent method can approximate the solution of a notoriously difficult type of integral equation (Wiener–Hopf) is encouraging, even for the moderate degree of accuracy obtained.

The last example we will consider is the equation

$$\frac{1+e^{-\pi x}}{1+x^2} = \int_0^\infty dy\, e^{-xy}P(y)\,, \tag{2.14}$$

a Fredholm integral equation of the first kind, whose exact solution is

$$P(x) = \begin{cases} \sin x\,, & 0<x<\pi\,,\\ 0\,, & x>\pi\,. \end{cases} \tag{2.15}$$

Approximate solution of (2.14) is equivalent to numerical inversion of a Laplace transform. For this equation a moment problem will be generated in a way different from the previous examples. Both the left-hand side and the kernel of (2.14) are expanded in a Taylor series about $x=0$. Comparing the two series expansions term-by-term yields the moment problem

$$\mu_n = \int_0^\infty dy\, y^n P(y)\,, \tag{2.16}$$

where the first few moments μ_n are $\mu_0 = 2$, $\mu_1 = \pi$, $\mu_2 = \pi^2 - 4$, $\mu_3 = \pi^3 - 6\pi$, $\mu_4 = \pi^4 - 12\pi^2 + 48$, and so on. The ten-moment max-ent solution of (2.16), $P_9(x)$, is given for various x (up to $x=5$) in Table V. The associated λ_i are listed in Table I. In this case, the optimal solution, $P_9^{\rm MAX}(x)$, is not available. Thus, $P_9(x)$ must be relied upon to directly generate an approximate solution. As a result, the max-ent, $P_9(x)$, approximation is accurate only to a few percent. Notice, however, that $P_9(x)$ rapidly vanishes for x just greater than π as it should. In the numerical algorithm the upper limit of integrations is taken to be *infinite* ($\simeq 50$ for practical purposes) and not π. The method itself generates the cutoff at $x \simeq \pi$.

TABLE IV. Six-moment max-ent solution of (2.10).

x	$P_5(x)$	$P_5^{\rm MAX}(x)$	Exact
−1.0	⋯	0.183 950	0.183 940
−0.8	⋯	0.224 676	0.224 664
−0.6	⋯	0.274 420	0.274 406
−0.4	⋯	0.335 178	0.335 160
−0.2	⋯	0.409 387	0.409 365
0.0	0.494 380	0.500 027	0.500 000
0.2	0.275 931	0.274 258	0.274 406
0.4	0.150 484	0.150 585	0.150 597
0.6	0.082 094	0.082 776	0.082 649
0.8	0.045 131	0.045 471	0.045 359
1.0	0.024 963	0.024 922	0.024 894

TABLE V. Ten-moment max-ent solution of (2.14).

x	$P_6(x)$	Exact
0.0	0.256 − 01	0.000
$\pi/8$	0.399	0.383
$\pi/4$	0.690	0.707
$3\pi/8$	0.946	0.924
$\pi/2$	0.971	1.000
$5\pi/8$	0.959	0.924
$3\pi/4$	0.660	0.707
$7\pi/8$	0.481	0.383
π	0.456 − 03	0.000
3.5	0.110 − 38	0.0
4.0	0.0	0.0
5.0	0.0	0.0

III. DISCUSSION

This paper will be concluded with a brief discussion of the weaknesses and strengths of the max-ent method illustrated in Sec. II. Possible weaknesses are noted first.

(1) It is not clear whether or not solutions generated by a larger number of moments than used here would significantly improve the numerical accuracy. If one needs to solve an integral equation correct to, say, six or more decimal places, other methods may be required. Ultimately, the algorithm of Ref. 2 for solving the max-ent equations (1.6) and (1.7) breaks down due to the accumulation of roundoff error (sometimes for low numbers of moments).

(2) It is clear from Sec. I that no assurance can yet be given in all cases of the convergence of the max-ent method adopted. Theoretical work is still needed.

(3) Currently the method has been worked out in some detail only for equations in one variable. In many areas of physics, such as scattering theory, the integral equations encountered are in two or more variables. Ultimately, the general usefulness of the method explored here to physics may depend upon how well it can do with multivariable problems.

(4) Often in physics, notably dispersion theory, one is faced with integral equations with singular kernels. The author has not yet attempted such problems; it may be that the maximum-entropy method cannot handle this important class of problems.

(5) Finally, it is unclear how the method can handle integral eigenvalue problems.

In spite of these difficulties, the proposed max-ent method has several advantages.

(1) The major advantage is the flexibility inherent in the method. In the previous examples, the form of $M_n(x)$ was chosen to be cx^n, c constant. This choice generated an associated moment problem. However, other choices of $M_n(x)$, such as a set of orthogonal polynomials, may be more appropriate to the problem at hand and may generate more accurate solutions. Furthermore, some thought on a given integral equation may reveal analytical information about the solution (such as its asymptotic behavior as x approaches some value), which may be incorporated into the max-ent equations (see Ref. 2 for further discussion).

(2) Even if one can obtain a max-ent solution accurate only to a few percent, the knowledge thus gained concerning the *qualitative* behavior of the solution may suggest a different method leading to accurate solution.

(3) Even if standard solution methods (such as the Newmann series) fail to converge, max-ent may provide a sufficiently accurate approximation (indeed in some cases, very accurate).

(4) The approximate max-ent function $P_N(x)$, for given moderate value of N, may be thought of as a variational starting point for an iterative scheme similar to the Newmann series itself. Indeed the function $P_N^{\mathrm{MAX}}(x)$ is nothing more than the first iterate of such a scheme. This may provide the required high accuracy solution.

ACKNOWLEDGMENTS

The author would like to thank E. T. Jaynes, A. Carlsson, D. Drabold, and N. Papanicolaou for very informative discussions.

Also, the author acknowledges the hospitality of the Department of Physics at Washington University where this work was carried out, and thanks C. M. Bender for partial financial support.

[1]H. Hochstadt, *Integral Equations* (Wiley, New York, 1973).

[2]L. R. Mead and N. Papanicolaou, J. Math. Phys. 25, 2404 (1984).

[3]The Hausdorff conditions on the moments

$$\sum_{l=0}^{k} (-)^l \binom{k}{l} \mu_{n+l} > 0, \quad n,k = 0,1,2,...,$$

are discussed in Ref. 2 and references contained therein. In Ref. 2 it is shown that these conditions are also those necessary and sufficient for the existence of a solution of the maximum-entropy equations (1.6).

[4]J. Guy, B. Mangeot, and A. Salès, J. Phys. A 17, 1403 (1984).

Classical particles with internal structure: General formalism and application to first-order internal spaces

M. V. Atre[a)]
Department of Physics, Indian Institute of Science, Bangalore 560012, India
N. Mukunda
Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560012, India

(Received 6 June 1986; accepted for publication 14 August 1986)

Group theoretic methods are used to systematically classify all possible internal structures for an elementary classical relativistic particle in terms of coset spaces of SL(2,C) with respect to its continuous subgroups. The allowed internal spaces Q are separated into first- and second-order ones, depending on whether a canonical description can be given using Q itself or if it needs the cotangent bundle T^*Q. Three of the former are found, one corresponding to the use of a Majorana spinor as the internal variable, the other two related to orbits in the Lie algebra of SO(3,1) under the adjoint action. For the latter two, a Lagrangian description of an elementary object with the corresponding internal space is set up, and the dynamics studied.

I. INTRODUCTION

Classical relativistic particles with internal structure, sometimes called relativistic rotators, have been the subject of a considerable amount of study recently. In addition to space-time position variables x^μ, such particles possess a set of internal variables q^r, which describe an internal space Q and are invariant under space-time translations. The Poincaré group $\mathscr{P}$ is the underlying symmetry and, for elementary systems, acts transitively on the total configuration space of variables (x^μ, q^r). The internal space Q admits a transitive action of the homogeneous Lorentz group SO(3,1). Such classical indecomposable objects are useful starting points for the description of, and approximation to, the concept of Regge trajectories. The action of SO(3,1) on Q gives rise to spin.

Several approaches are available and have been used to study the dynamics of such classical systems. One is to work within the Lagrangian[1] or the Hamiltonian[2] formalism; another is to directly write down manifestly covariant equations of motion after specifying a complete set of variables[3]; or finally, one may derive the equations of motion guided by the ten conservation laws and physically reasonable kinematical constraints.[4] Of these, the use of the Lagrangian formalism seems in many ways to be the most convenient. Both manifest covariance and the conservation laws are easily ensured, and possible couplings to external fields can be systematically analyzed. In addition, Dirac's theory of constrained dynamical systems provides a systematic procedure for handling all possible Lagrangians and preparing the ground for quantization.[5]

The Lagrangian approach was pioneered by Frenkel in his study of relativistic charged spinning particles in external electromagnetic fields.[6] Frenkel described spin by a second-rank antisymmetric tensor. Subsequently much important work was done by several authors, of whom we may mention Mathisson,[7] Lubanski,[8] Honl and Papapetrou,[9] and Bhabha and Corben.[10] These developments are summarized in Corben's book.[11] As representatives of more recent work in which the choice of an internal space plays a more important role, we may first mention Halbwach's use of a tetrad (vierbein) as the internal variable and as the source of spin.[12] In the work of Itzykson and Voros,[13] the internal variable was a four-vector attached to the space-time position, and the Lagrangian formalism was used. The work of Hanson and Regge,[1] which has justly attained the status of a classic in the field, again used a vierbein, equivalently the homogeneous Lorentz group itself, as the internal space. They made imaginative use of Dirac's methods for singular Lagrangian systems, and demonstrated that reparametrization invariance leads to a mass-spin Regge trajectory relationship. This shift of emphasis from describing a point particle with fixed mass and magnitude of spin to describing a family of particles for which, say, mass appears as a function of spin, occurs also in the work of Rafanelli.[14] Mukunda *et al.*[15] constructed two models within the Lagrangian formalism, in which the internal variable was, respectively, a unit spacelike vector and a Majorana spinor. In both cases a Regge relationship emerged, but the algebraic structure of the constraints was very different in the two cases.

On surveying the work in this field (in more detail than is possible here), it appears that there has so far been no systematic analysis of the problem in which all possible choices of internal space Q are exhaustively classified and examined on the basis of reasonable and uniformly valid physical requirements. It is the purpose of this paper, and following ones in this series, to provide such a treatment. We shall show that it is possible to develop a systematic classification procedure and analyze all possible internal structures for classical relativistic point particles by using group theoretic and differential geometric methods. In each case the Lagrangian will be taken as the starting point to analyze the possible dynamics and Regge relationships between mass and spin.

The material of this paper is organized as follows. In Sec. II the three basic physical requirements that will guide our analysis are explained, and some of their immediate consequences are described. These requirements are indecom-

[a)] Present address: Theoretical Physics Group, Tata Institute of Fundamental Research, Bombay 400005, India.

posability of the classical particle with internal structure, reparametrization invariance of the action, and symmetry of the theory under the Poincaré group. In particular, the first of these requirements shows that each possible internal space Q is the coset space G/H of the group $G = \mathrm{SL}(2,C)$ with respect to some subgroup H; and if one restricts H to be a continuous Lie subgroup of G, all possible Q's can be systematically classified. In Sec. III, the set of all possible Q's is divided into two types, which we call first-order spaces (FOS's) and second-order spaces (SOS's). The difference between the two is that in the former case the phase space analysis can be carried out at the level of Q itself, while with an SOS the use of T^*Q is unavoidable. From this point onward, this paper is devoted to FOS's, the SOS's being taken up for study in later papers of this series. The search for FOS's is aided by the Kostant–Kirillov–Souriau (KKS) theorem, which relates them to orbits in the Lie algebra $\mathbf{G}$ of G under the adjoint action by G. The orbits are classified and each one exhibited as a coset space G/H for a suitable H. It turns out that only two such coset spaces arise, corresponding to two possible H's, each of dimension 2. All nontrivial orbits, except for one very special orbit, correspond to one particular H, which is therefore the generic case; the special orbit corresponds to a different choice of H, and is called the exceptional orbit. The topological structures of the generic and the exceptional orbits, which are quite different, are studied via the Iwasawa decomposition theorem for G. It is finally seen that there are three possible FOS's, two realized as a generic and the exceptional orbit in $\mathbf{G}$, respectively, and the third being a twofold covering of the exceptional orbit. This last FOS turns out to be the same internal space as used in the spinor model, which has been studied elsewhere.[15] The symplectic structures on the generic and the exceptional $\mathbf{G}$-orbits are studied in Sec. IV. In both cases the symplectic two-form is shown to be exact, a necessary condition for being able to set up a global Lagrangian. The Lorentz transformation properties of the associated one-forms are also examined. Section V takes up the question of constructing the most general Lagrangian obeying the requirements of Sec. II, when the internal space is one of the two FOS's, either a generic or the exceptional orbit in $\mathbf{G}$. In both cases the internal variable is an element of $\mathbf{G}$, i.e., an antisymmetric real second-rank tensor $\xi_{\mu\nu}$, with the invariants constructed from it being assigned definite values. These values differ for the generic and the exceptional case. It is shown that while in the generic case the simplest available Lorentz covariant object that can be constructed on the internal space, capable of being coupled to the space-time coordinates, is a symmetric second-rank tensor $t_{\mu\nu}$, in the exceptional case a more elementary object, namely a four-vector V_μ, can be constructed. Therefore, the model based on a generic orbit as the internal space is referred to as the symmetric tensor model (STM). The behavior of nonspinorial quantities, such as the space-time trajectory, in the spinor model of Ref. 15, is identical with corresponding quantities in the model based on the exceptional orbit as internal space. Consequently, this latter model is not studied further here. On the other hand, for the STM, both the constraint structure and the dynamical equations are worked out in some detail. The paper concludes with some remarks in Sec. VI.

A brief resumé of these ideas may be found in Ref. 16.

II. PHYSICAL REQUIREMENTS AND GUIDELINES

For a structureless relativistic point particle, the configuration space coincides with Minkowski space-time $\mathscr{M}$ with coordinates x^μ. (The metric is $+ + + -$). The Poincaré group $\mathscr{P}$ acts transitively on $\mathscr{M}$. The canonical formalism based on $T^*\mathscr{M}$ uses the conjugate momentum p_μ and the orbital angular momentum $L_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu$ as generators for $\mathscr{P}$. Absence of internal structure is reflected in a vanishing Pauli–Lubanski vector:

$$W_\mu \equiv \tfrac{1}{2}\,\epsilon_{\mu\nu\rho\sigma}\, p^\nu L^{\rho\sigma} = 0\,. \tag{2.1}$$

To accommodate internal structure it is necessary to enlarge the configuration space to the product $\mathscr{M}\times Q$, where Q is the space of internal variables with (possibly local) coordinates q^r. By definition Q is invariant under space-time translations, and is only affected by homogeneous Lorentz transformations. In order to allow for spinorial internal variables as well as with a view towards eventual quantization, we shall use $G = \mathrm{SL}(2,C)$, the universal covering group of $\mathrm{SO}(3,1)$, in place of $\mathrm{SO}(3,1)$. Therefore G acts on Q via point transformations.

Let us now list the three basic physical requirements that will underlie our analysis, and then comment upon them. We require that (a) the action of G on Q should be transitive; (b) in each given case, the Lagrangian $\mathscr{L}$ must depend on the velocities $\dot{x}$, $\dot{q}$ (the dot representing the derivative with respect to an evolution parameter s) in such a way that the action is reparametrization invariant; and (c) the Lagrangian must be manifestly invariant or quasi-invariant under the action by $\mathscr{P}$, i.e., $\mathscr{L}$ could change by a total derivative with respect to s as a result of action by an infinitesimal element of $\mathscr{P}$.

Requirement (a) is a minimality condition expressing the idea that the particle is irreducible or indecomposable, ensuring that any two points of Q can be connected by some element of G. The transitivity of the G action on Q means that Q can be identified with the coset space G/H for some subgroup H in G. We shall take G/H to be the left coset spaces, so an element of G/H is written as gH for some $g\in G$. Since G is connected (and in fact simply connected), it follows that every coset space G/H is certainly connected, whether or not the subgroup H is connected. We shall, however, restrict H to be a closed, connected, continuous Lie subgroup of G, so that individual cosets gH will also be connected. All such nontrivial subgroups are known up to conjugation,[17] and are listed in the Appendix. The dimension of a possible internal space Q is related to that of H by $\dim Q = 6\text{-}\dim H$. Starting from the largest possible H, namely $H = G$ itself, and going down to the smallest, when H consists of just the identity element, and including both these possibilities, it turns out that there are 13 possible distinct choices for H, and in addition, there are two one-parameter families of subgroups. In the notation of Patera *et al.*, we write $F_1 = G, F_2,\ldots, F_4, F_5^\varphi, F_6,\ldots,F_{10}, F_{11}^\varphi, F_{12},\ldots, F_{14}, F_{15} = \{e\}$ for the possible continuous subgroups of G up to conjugation. The one-parameter families are F_5^φ and F_{11}^φ, with the range of the parameter φ in both cases specified in the Appendix. Thus we find that requirement (a) has led to the result that every conceivable internal space Q has to be the coset space G/H with H equal to one of the F's listed above. The choice $H=F_1=G$ leads to a trivial internal space, since then $Q=G/H$ consists of just a single point; it may be taken to correspond to the structureless mass point mentioned at the start of this section, and will hereafter be disregarded. The choice $H=F_{15}=\{e\}$ corresponds to Q being G itself, leading essentially to the Hanson–Regge model.[1] All other classical models for indecomposable objects with internal structure necessarily correspond to some intermediate choice of H. In the context of linear relativistic quantum mechanical wave equations, Finkelstein[14] made a similar classification of possible internal structures. It appears, however, that his enumeration was incomplete since some of the subgroups F listed above were missing in his work, and only 11 possibilities were listed.

Requirement (b) implies that the Lagrangian is homogeneous of degree one in the velocities $\dot{x}$ and $\dot{q}$. As a result, we have full freedom in the choice of the evolution parameter s, such as proper time, physical time, etc. Reparametrization invariance of the action is a kind of gauge degree of freedom, causing the Lagrangian to be singular and leading to a primary constraint in the Hamiltonian formulation.

Requirement (c) in any theory ensures the validity of the ten conservation laws and also ensures that in a singular theory the collection of all constraints on the phase space forms a Poincaré covariant set. From the previous paragraph we know that there will be at least one constraint due to the reparametrization invariance. In case there are no other constraints, this one is necessarily an explicitly Poincaré invariant constraint among all available and independent Poincaré scalars on phase space. For simple enough situations, these scalars are the two Casimir invariants of the Poincaré group formed out of its generators, namely $P^2 \equiv -M^2 = p^\mu p_\mu$ and $W^2 = W^\mu W_\mu$ (see Ref. 19). We thus see how under suitable conditions requirements (b) and (c) can together lead to a constraint expressible in the form $P^2=\alpha(W^2)$, which is essentially a mass-spin Regge relationship.

An interesting physical criterion worth keeping in mind, and which serves to distinguish some choices of Q from others, is whether, at the end of the constraint analysis, the Dirac brackets[5] of x^μ among themselves vanish or not. This would have implications for the possibility of introducing configuration space wave functions in a quantum theory. Of course, well before taking up this question, we must determine for each possible Q whether Lagrangians can be constructed in which there is a nontrivial coupling between the space-time variables x, $\dot{x}$ and the internal ones q, $\dot{q}$.

We now proceed to a systematic analysis keeping these physical guidelines in mind.

III. FIRST- AND SECOND-ORDER SPACES—STUDY OF FOS'S

Even with the imposition of requirement (a) of Sec. II, the number of possible distinct internal spaces Q is quite large, and one needs some physically well motivated point of view, which helps to separate all possible Q's into different types and makes the detailed analysis more manageable. We now develop such a point of view. With respect to the setting up of a canonical formalism starting from some Lagrangian, two different possibilities arise depending on the nature of Q. It may be that we are able to set up a Poisson bracket (PB) structure on Q, which is invariant under G, i.e., the point transformations realizing the action of G on Q are canonical transformations as defined by this PB. Then Q is of even dimension and provided the Lagrangian has a suitable form (described later), in the Hamiltonian treatment we need not use T^*Q at all. The physical idea is that in such a case there is no need to introduce new variables canonically conjugate to the internal variables q^r, but that G-invariant PB's can be defined among the q's themselves. A compact way of expressing all this is to say that Q is a homogeneous symplectic G-space: homogeneous since G acts transitively on Q; symplectic since there is a closed nondegenerate two-form ω on Q leading to the *PB*; and symplectic G-space since ω is invariant under G.

Internal spaces Q having the above properties will be called first-order spaces (FOS's). All other Q's will be called second-order spaces (SOS's). In the case of an FOS, the Lie algebra of G is realized by functions and PB's on Q. Provided the Lagrangian is suitably chosen (see later), for an FOS the canonical formulation for the entire problem uses $T^*\mathscr{M}\times Q$, and not $T^*(\mathscr{M}\times Q)\simeq T^*\mathscr{M}\times T^*Q$, as the full phase space. For an SOS, on the other hand, use of $T^*\mathscr{M}\times T^*Q$ is unavoidable. In either case, the generators of the physical homogeneous Lorentz group SO(3,1) have the form

$$J_{\mu\nu}=L_{\mu\nu}+S_{\mu\nu}\,, \tag{3.1}$$

with $S_{\mu\nu}$ related to the action of G on Q. The $S_{\mu\nu}$ are realized as functions on Q or on T^*Q depending on whether Q is an FOS or an SOS. Existence of the corresponding first- or second-order internal structure is confirmed by a nonvanishing Pauli–Lubanski vector:

$$W_\mu=\tfrac{1}{2}\,\epsilon_{\mu\nu\rho\sigma}\,p^\nu S^{\rho\sigma}\,. \tag{3.2}$$

The condition on a FOS Q enabling us to define a suitable Lagrangian taking advantage of the nature of Q is that the two-form ω must be exact: $\omega=d\theta$, θ a one-form on Q. The G invariance of ω means that θ is either also G invariant or else changes by a closed piece under action by an (infinitesimal) element of G. (Even if the symplectic form is not exact, a Lagrangian may be defined by working on a principal bundle on Q,[20] but such cases will not arise in the present problem.) Given the existence of the one-form θ, the Lagrangian has a leading term θ/dt linear in the internal velocities $\dot{q}$, and in the rest of the Lagrangian there must be no dependence on $\dot{q}$ at all. (Of course the question of nontrivial coupling between the space-time variables x, $\dot{x}$ and the q in the rest of the Lagrangian must be examined.) This kind of $\dot{q}$ dependence in $\mathscr{L}$ guarantees that the application of the canonical formalism to $\mathscr{L}$ leads to just the PB's on Q that are determined by ω in the first place. If Q is an SOS, there is no such natural breakup of a possible Lagrangian into a part linear in $\dot{q}$ and a part independent of $\dot{q}$. Of course, even in the case of an FOS one might decide to treat the problem as though it were an SOS, by constructing a Lagrangian with a nonlinear dependence on $\dot{q}$ and later on passing to $T^*(\mathscr{M}\times Q)$. However, in our work we shall treat every FOS in the manner detailed above, so that the final physical phase space is just $T^*\mathscr{M}\times Q$.

The dimension of an FOS is, as mentioned earlier, necessarily even and so must be 6, 4, or 2. To discover the FOS's, we make use of the Kostant–Kirillov–Souriau (KKS) theorem[21] which in the case of the group G, since it is semisimple, states the following: the only coset spaces G/H admitting a G-invariant symplectic structure are orbits in the dual $\mathbf{G}^*$ to the Lie algebra $\mathbf{G}$ of G under the coadjoint action, or covering spaces of such orbits. However, again since G is semisimple, there is no essential distinction between the coadjoint representation of G acting on $\mathbf{G}^*$ and the adjoint representation acting on $\mathbf{G}$, so we may equally well work with the orbits in $\mathbf{G}$ in searching for possible FOS's. Then if we denote by $D^{\rm adj}(g)$ the adjoint transformation representing $g\in G$ and acting on $\mathbf{G}$, and if $\mathscr{O}$ is any orbit in $\mathbf{G}$, we have $D^{\rm adj}(g)\ \mathscr{O}=\mathscr{O}$ for every $g\in G$, and moreover $\mathscr{O}$ admits a unique (up to a constant factor), closed, nondegenerate two-form, the Kirillov form, which makes $\mathscr{O}$ a symplectic manifold. We conclude that the possible FOS's are orbits (or their coverings) in $\mathbf{G}$ under $D^{\rm adj}(g)$, so all orbits must be classified and each one exhibited as a coset space G/H. This we proceed to do. We shall find that though there are several families of orbits $\mathscr{O}$ in $\mathbf{G}$, only two distinct coset spaces arise.

It is convenient to deal with the Lie algebra $\mathbf{G}$ via some faithful irreducible matrix representation of it. If the generators of such a representation are written as $\Sigma_{\mu\nu}$, they obey the commutation relations

$$[\Sigma_{\mu\nu},\Sigma_{\rho\sigma}]=g_{\mu\rho}\Sigma_{\nu\sigma}-g_{\nu\rho}\Sigma_{\mu\sigma}+g_{\mu\sigma}\Sigma_{\rho\nu}-g_{\nu\sigma}\Sigma_{\rho\mu}\,. \tag{3.3}$$

In this representation, a general element of $\mathbf{G}$ is represented by

$$J(\xi)=\tfrac{1}{2}\xi^{\mu\nu}\Sigma_{\mu\nu}\,, \tag{3.4}$$

where the real antisymmetric set $\xi^{\mu\nu}=-\xi^{\nu\mu}$ with six independent components denotes an element of $\mathbf{G}$ in the abstract. Let the above representation of $\mathbf{G}$ lead on exponentiation to the representation $D(g)$ of G. Then the adjoint action of $g\in G$ on $\xi\in\mathbf{G}$ is determined as follows:

$$\begin{aligned}&\xi'=D^{\rm adj}(g)\xi\,,\\ &J(\xi')=D(g)J(\xi)D(g)^{-1}\,,\\ &\xi'_{\mu\nu}=\Lambda(g)_\mu{}^\rho\Lambda(g)_\nu{}^\sigma\xi_{\rho\sigma}\,.\end{aligned} \tag{3.5}$$

Here $\Lambda(g)$ belongs to the defining four-dimensional "vector" representation of SO(3,1). It follows that under the change $\xi\rightarrow\xi'$, there are two independent invariants that can be expressed in terms of the three-dimensional quantities $\xi_j=\frac{1}{2}\epsilon_{jkl}\xi_{kl}$, $\eta_j=\xi_{0j}$ and can be conveniently parametrized as follows:

$$\begin{aligned}&\mathscr{C}_1=\tfrac{1}{2}\xi^{\mu\nu}\xi_{\mu\nu}=|\boldsymbol{\xi}|^2-|\boldsymbol{\eta}|^2=a^2-b^2\,,\\ &\mathscr{C}_2=-\tfrac{1}{8}\epsilon_{\mu\nu\rho\sigma}\xi^{\mu\nu}\xi^{\rho\sigma}=\boldsymbol{\xi}\cdot\boldsymbol{\eta}=ab\,.\end{aligned} \tag{3.6}$$

The single point $\xi_{\mu\nu}=0$ forms an orbit all by itself and is omitted from discussion. For specified values of the parameters a and b, the set of all $\xi_{\mu\nu}$ constitutes an orbit, which may be denoted $\mathscr{O}_{a,b}$ and over which $\mathscr{C}_1$ and $\mathscr{C}_2$ have the above constant values.

In Table I, the orbits are grouped into suitable families, with definite ranges for a and b in each family. For each individual orbit, the following additional information is given in Table I: (i) a conveniently chosen representative element on the orbit; (ii) as an aid to geometrical visualization of the nature of this representative element, an independent set of space-time vectors which are invariant under the one-parameter group of Lorentz transformations generated by this representative element; and (iii) the generators of the stability group of the representative element. The following notation has been used in making the entries in Table I: $\mathbf{J}$ and $\mathbf{K}$ are the generators of spatial rotations and pure Lorentz transformations, respectively, so that

$$J(\xi)=\boldsymbol{\xi}\cdot\mathbf{J}-\boldsymbol{\eta}\cdot\mathbf{K}\,. \tag{3.7}$$

The four unit vectors e_0, e_1, e_2, and e_3 are mutually orthogonal, e_0 is timelike, and the others spacelike.

As an example, $\mathscr{O}_{a,0}$ for $a>0$ is the orbit containing the element aJ_3 of $\mathbf{G}$ and all its transforms under $D^{(\rm adj)}(g)$. The unit timelike and unit spacelike vectors e_0 and e_3, respectively, are each invariant under the spatial rotations generated by aJ_3. For this reason, $\mathscr{O}_{a,0}$ is also specially denoted as $\mathscr{O}^{ts}_a$. Similarly, $\mathscr{O}_{0,0}$ is denoted as $\mathscr{O}^{ls}$ and $\mathscr{O}_{0,b}$ for $b>0$ as $\mathscr{O}^{ss}_b$.

To study these orbits in more detail and for practical calculations such as the determination of the stability group of the representative element, etc., it is convenient to use the defining $(\frac{1}{2},0)$ representation of $\mathbf{G}$ which in terms of the Pauli matrices $\boldsymbol{\sigma}$ is

$$\begin{aligned}&\Sigma_{jk}=-(i/2)\,\sigma_l\,,\quad (jkl)\ \text{cyclic}\,,\\ &\Sigma_{oj}=-\tfrac{1}{2}\sigma_j\,.\end{aligned} \tag{3.8}$$

TABLE I. Classification of orbits in $\mathbf{G}$.

	Orbit	Ranges of parameters	$\mathscr{C}_1$	Representative element	Invariant vectors	Generators of stability group
$\mathscr{C}_2=ab\neq0$	$\mathscr{O}_{a,b}$	$a,b>0$	a^2-b^2	aJ_3-bK_3	None	J_3,K_3
	$\mathscr{O}_{a,b}$	$a>0,\ b<0$	a^2-b^2	aJ_3-bK_3	None	J_3,K_3
$\mathscr{C}_2=ab=0$	$\mathscr{O}_{a,0}\equiv\mathscr{O}^{ts}_a$	$a>0,\ b=0$	a^2	aJ_3	e_0,e_3	J_3,K_3
	$\mathscr{O}_{0,0}\equiv\mathscr{O}^{ls}$	$a=b=0$	0	J_2+K_1	$e_0+e_3,\ e_2$	J_1-K_2, J_2+K_1
	$\mathscr{O}_{0,b}\equiv\mathscr{O}^{ss}_b$	$a=0,\ b>0$	$-b^2$	bK_3	e_1,e_2	J_3,K_3

(These differ from the conventional definitions by a factor of i.) The general Lie algebra element $J(\xi)$ of Eqs. (3.4) and (3.7) becomes the two-dimensional matrix

$$J(\xi) = \tfrac{1}{2}(\boldsymbol{\eta} - i\boldsymbol{\xi})\cdot\boldsymbol{\sigma}. \tag{3.9}$$

With this easy-to-handle representation, the results of Table I and the identification of each orbit as a coset space G/H are readily obtained. We describe the families of orbits briefly.

$\mathscr{O}_{a,b}$ for $\mathscr{C}_2 \neq 0$: These are two disjoint, simply connected, two-parameter families of orbits corresponding, respectively, to $\mathscr{C}_2 > 0$ ($a>0, b>0$) and $\mathscr{C}_2 < 0$ ($a>0, b<0$). On any such $\mathscr{O}_{a,b}$, a convenient representative element is $aJ_3 - bK_3$, which in the $(\frac{1}{2},0)$ representation is the diagonal matrix $\frac{1}{2}(b - ia)\,\sigma_3$. The stability group S of this element consists of all matrices in G commuting with σ_3, that is, of all diagonal unimodular complex matrices:

$$S = \left\{ \begin{pmatrix} re^{i\theta} & 0 \\ 0 & (1/r)e^{-i\theta} \end{pmatrix} : \; r>0, \quad 0 \leqslant \theta < 2\pi \right\}. \tag{3.10}$$

Thus S is independent of (a,b). The elements of S are products of unitary unimodular and real unimodular diagonal matrices of the forms

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}, \begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix}. \tag{3.11}$$

These constitute, respectively, the subgroup $U(1)$ generated by J_3 and the subgroup $B(3)$ of pure Lorentz transformations (boosts) generated by K_3. The stability group S is therefore F_9 in the list of subgroups of G given in the Appendix: $S = F_9 = \mathrm{U}(1)\times B(3)$. In the two-to-one homomorphism $G \to \mathrm{SO}(3,1)$, F_9 goes into the subgroup $\mathrm{SO}(2)\times\mathrm{SO}(1,1) \subset \mathrm{SO}(3,1)$. Every one of the orbits $\mathscr{O}_{a,b}$ with $\mathscr{C}_2 \neq 0$ is seen to be essentially the same coset space G/F_9, which is thus a possible FOS.

$\mathscr{O}_{a,b}$ for $\mathscr{C}_2 = 0$: Here one has the possibilities $\mathscr{C}_1 > 0$ ($a>0$, $b=0$), $\mathscr{C}_1 < 0$ ($a=0$, $b>0$) and $\mathscr{C}_1 = 0$ ($a=b=0$). In the first two cases, one has simply connected one-parameter families of orbits, with representative elements aJ_3 on $\mathscr{O}_{a,0}$ and bK_3 on $\mathscr{O}_{0,b}$. But in both cases, the stability group of the representative element is the same S of (3.10) as in the discussion of $\mathscr{O}_{a,b}$ for $\mathscr{C}_2 \neq 0$, so these orbits are topologically the same coset space G/F_9 as before. Thus consideration of these types of orbits does not lead to any new candidates for an FOS.

The singular case $\mathscr{C}_1 = 0$ with $a=b=0$ is, unlike the other cases, not a family of orbits but a single orbit all by itself, leading to a new possible FOS. For this reason we call it an exceptional orbit. As representative element on this orbit we can take the combination $J_2 + K_1$, which in the $(\frac{1}{2},0)$ representation is the nilpotent matrix

$$J_2 + K_1 = -\frac{i}{2}\sigma_2 - \frac{1}{2}\sigma_1 = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}. \tag{3.12}$$

The one-parameter subgroup generated by $J_2 + K_1$ leaves invariant a lightlike vector and a spacelike vector, which explains the notation $\mathscr{O}_{0,0} = \mathscr{O}^{ls}$. The stability group of $J_2 + K_1$ has a more intricate structure than was the case for representative elements on other orbits. Writing N_0 for this subgroup of G, it consists of all elements $n \in G$ that commute with the nilpotent matrix appearing in Eq. (3.12). We easily find that such an n must be of one of the two forms

$$\begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & \beta \\ 0 & -1 \end{pmatrix}, \tag{3.13}$$

where β is any complex number. Hence the subgroup $N_0 \subset G$ has two disjoint components, each of which is simply connected. We indicate this structure of N_0 by writing it as the union of the component N containing the identity and the component N' whose elements are the negatives of those of N:

$$N_0 = N \cup N', \quad N = \left\{ \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \right\}, \quad N' = \left\{ \begin{pmatrix} -1 & \beta \\ 0 & -1 \end{pmatrix} \right\}. \tag{3.14}$$

Here N is generated by the combinations $J_2 + K_1, J_1 - K_2$; it is a two-parameter Abelian group and is in fact the subgroup $F_{10} \subset G$. In the two-to-one homomorphism from G to $\mathrm{SO}(3,1)$, the elements $n \in N$ and $-n \in N'$ share the same image, which is contained in the image of F_{10} in $\mathrm{SO}(3,1)$. We may call the latter image F_{10} again (see also later), since the homomorphism respects distinctions among elements of F_{10} in G. We thus see that the exceptional orbit $\mathscr{O}^{ls}$ can be written as a coset space either of G or of $\mathrm{SO}(3,1)$ as follows:

$$\mathscr{O}^{ls} \simeq G/N_0 \simeq \mathrm{SO}(3,1)/F_{10}. \tag{3.15}$$

We shall deal with $\mathscr{O}^{ls}$ in this form as a possible internal space Q even though there is here a departure from the statement made in Sec. II that attention would be restricted to coset spaces G/H for H a connected Lie subgroup of G. In the present case, $H = N_0 = N \cup N'$ is not connected. However, one should also bear in mind that in dealing with orbits in G one is really tackling a problem at the level of $\mathrm{SO}(3,1)$ rather than truly at the level of G. We may at this point summarize the results of this section thus far: a detailed analysis of the structure of orbits in G and the use of the KKS theorem, shows that the only possible FOS's are the two coset spaces, G/F_9 and $G/N_0 \simeq \mathrm{SO}(3,1)/F_{10}$, or other coset spaces G/H, which may be covering spaces of one of the first two. Any orbit $\mathscr{O}_{a,b}$ with $\mathscr{C}_2 \neq 0$, $\mathscr{O}_{a,0}$ with $a>0$, or $\mathscr{O}_{0,b}$ with $b>0$ can serve as a model for the coset space G/F_9; while the exceptional orbit $\mathscr{O}_{0,0}$ is a model for the coset space G/N_0.

The topological structures of these two coset spaces are disclosed by use of the Iwasawa or KAN decomposition theorem, which states that any element g of a semisimple noncompact Lie group G can be uniquely expressed as a product of three elements: $g = kan$, where k belongs to a maximal compact subgroup K, a to an Abelian subgroup A, and n to a nilpotent subgroup N of the full group. For $G = \mathrm{SL}(2,C)$: $K = \mathrm{SU}(2)$ generated by $\mathbf{J}$; $A = B(3)$ generated by K_3; and $N = F_{10}$ generated by $J_2 + K_1, J_1 - K_2$. Topologically, $\mathrm{SU}(2)$ has the structure of the sphere S^3, A has the structure of the real line R, and N has that of the plane R^2. Thus as is well known $\mathrm{SL}(2,C)$ has the topology of $S^3 \times R^3$. Now the first coset space found above in the search for FOS's is $G/F_9 = \mathrm{SL}(2,C)/\mathrm{U}(1)\times B(3)$. Since $\mathrm{SU}(2)/\mathrm{U}(1)$ has the structure of the two-sphere S^2, we see that the coset space G/F_9 has a structure indicated by

$$G/F_9 \simeq \mathrm{SU}(2)\times B(3)\times N/\mathrm{U}(1)\times B(3) \simeq (\mathrm{SU}(2)/\mathrm{U}(1))\times N \simeq S^2\times R^2 . \qquad (3.16)$$

Each of the nonexceptional orbits has exactly this topological structure.

To deal with the exceptional orbit $\mathscr{O}_{0,0} = \mathscr{O}^{ls}$, which was shown to be the coset space $G/N_0 \simeq \mathrm{SO}(3,1)/F_{10}$, we make a few preliminary remarks regarding the Iwasawa decomposition for SO(3,1). Here the maximal compact subgroup K is SO(3), the image of SU(2) in G under the two-to-one homomorphism $G \to \mathrm{SO}(3,1)$. However, under this same homomorphism, the subgroups $A = B(3)$ and $N = F_{10}$ in G have images in SO(3,1) which are "as large as" the originals. The homomorphism maps distinct elements in A (resp. N) into distinct elements of SO(3,1). For this reason we may without fear of confusion refer to the homomorphic images of $A,N \subset G$ by the same symbols A,N but now understood as subgroups in SO(3,1). With this understanding, we have $\mathrm{SO}(3,1) = \mathrm{SO}(3)\times A\times N$. It now follows from Eq. (3.15) that the topological structure of the exceptional orbit $\mathscr{O}^{ls}$ is given thus:

$$\mathscr{O}^{ls} \simeq \mathrm{SO}(3,1)/N \simeq \mathrm{SO}(3)\times A\times N/N \simeq \mathrm{SO}(3)\times R . \qquad (3.17)$$

This is distinct from the structure of any nonexceptional orbit.

Now the KKS theorem allows a coset space G/H, which is an FOS to be either an orbit or a covering space of an orbit in **G**. One can convince oneself by simple arguments (here omitted) that in the case of any nonexceptional orbit viewed as G/F_9, there are no other coset spaces G/H that are covering spaces of G/F_9. However, the situation is different in the case of the exceptional orbit $\mathscr{O}_{0,0} = \mathscr{O}^{ls}$, which, viewed as a coset space of G, has the structure $G/(N\cup N')$. Since the subgroup involved here is made up of two disjoint components, we can see that there is another coset space G/N, using the connected subgroup N, which is "twice as big as" $\mathscr{O}_{0,0}$ and covers the latter twice, since N is "half of" $N\cup N'$. Thus we have a third candidate for an FOS, namely $Q = G/N$ with the topological structure $\mathrm{SU}(2)\times A \simeq S^3\times R$. We repeat that this is not an orbit in **G** but a twofold covering of the exceptional orbit $\mathscr{O}_{0,0}$. (One can convince oneself that no further coverings are possible.) This coset space G/N is precisely the internal space corresponding to the use of a Majorana spinor as an internal variable, which appears in the work of Ref. 15. This fact is easily established. From the Iwasawa decomposition for G, it is clear that each coset in G with respect to N has a unique representative element of the form ka where $k \in \mathrm{SU}(2)$ and $a \in A$. As a matrix, this coset representative, or "coordinate" for G/N, is

$$ka = \begin{pmatrix} \lambda & \mu \\ -\mu^* & \lambda^* \end{pmatrix}\begin{pmatrix} r & 0 \\ 0 & 1/r \end{pmatrix} = \begin{pmatrix} \lambda r & \mu/r \\ -\mu^* r & \lambda^*/r \end{pmatrix},$$
$$|\lambda|^2 + |\mu|^2 = 1, \quad r>0 . \qquad (3.18)$$

Evidently this matrix is completely determined by its first column $\binom{\lambda r}{-\mu^* r}$, which is not identically vanishing. Now the action of an element $g \in G$ on G/N causes a change in the coset representative, which amounts to this column vector transforming as an undotted spinor. This is because removal of a factor belonging to N on the right-hand side of an element of G does not alter the first column of the corresponding two-dimensional matrix, since elements of N are upper triangular matrices as shown in Eq. (3.14). To make contact with the variables of the spinor model, we write $\binom{\lambda r}{-\mu^* r}$ in the form $\binom{q_1 + iq_2}{p_1 - ip_2}$ by identifying

$$q_1 + iq_2 = 2V_0^{1/2}\lambda, \quad p_1 - ip_2 = -2V_0^{1/2}\mu^* ,$$
$$V_0 = r^2/4 = \tfrac{1}{4}(q_1^2 + q_2^2 + p_1^2 + p_2^2) . \qquad (3.19)$$

To say that $\binom{\lambda r}{-\mu^* r}$ transforms as a two-component spinor with respect to G is then to say that

$$\psi = \begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} \qquad (3.20)$$

is a real four-component Majorana spinor. Hence the space of Majorana spinors with $V_0 > 0$ is diffeomorphic to the space of nonidentically vanishing $(\frac{1}{2},0)$ spinors and this is the same as the coset space G/N. A global coordinate system on G/N is given by ψ, and as shown in Ref. 15 the fundamental canonical PB's

$$\{q_a, p_b\} = \delta_{ab}, \quad \{q_a, q_b\} = \{p_a, p_b\} = 0, \quad a,b = 1,2 \qquad (3.21)$$

are G invariant.

To summarize the results of this section, all possible internal spaces $Q = G/H$ were divided into two types namely FOS's and SOS's. With the help of the KKS theorem and a detailed study of the adjoint orbits in **G**, three possible FOS's have been found, namely the coset spaces G/F_9, G/N_0, and $G/N = G/F_{10}$. The last of these has been recognized as the internal space of the spinor model, which was earlier developed as a result of a study of the new Dirac equation.[22] In the rest of this paper we study the first two FOS's corresponding to any nonexceptional and the exceptional orbit in **G**, respectively, as possible internal space for a classical particle.

IV. SYMPLECTIC STRUCTURES ON THE FOS's

On each of the two coset spaces G/F_9 and G/N_0 we wish to define a G-invariant symplectic form whose existence is guaranteed by the KKS theorem. If ω is this form, by suitably inverting it one obtains G-invariant PB's among functions on the concerned internal space. If furthermore ω is exact, i.e., $\omega = d\theta$, where the one-form θ appears in (local) coordinates q^r for Q as

$$\theta = f_r(q)dq^r , \qquad (4.1)$$

then the total Lagrangian for a particle with internal space Q has a leading term

$$\mathscr{L}_0(q,\dot q) = f_r(q)\dot q^r , \qquad (4.2)$$

linear in $\dot q$, and in the rest of the Lagrangian there is no $\dot q$ dependence. A reparametrization invariant contribution to the action is given by $\mathscr{L}_0$, by itself.

In our problem it is physically more transparent and convenient to begin by defining manifestly G-invariant PB's, then by a process of matrix inversion arrive at ω, and then demonstrate that ω is exact. Once a suitable θ has been found, its G transformation property can be directly examined. We carry out these steps in this section, first for $Q = G/N_0$ and next for $Q = G/F_9$.

A. The case $Q=G/N_0$

The six components of $\xi_{\mu\nu}$ can be used as an overcomplete system of coordinates for the exceptional orbit $\mathscr{O}_{0,0} = \mathscr{O}^{ls}$. Among them we postulate the fundamental PB's

$$\{\xi_{\mu\nu},\xi_{\rho\sigma}\} = g_{\mu\rho}\xi_{\nu\sigma} - g_{\nu\rho}\xi_{\mu\sigma} + g_{\mu\sigma}\xi_{\rho\nu} - g_{\nu\sigma}\xi_{\rho\mu}, \tag{4.3}$$

patterned after the structure of the Lie algebra **G** itself.[19] The Jacobi identities are automatically satisfied. If we specify that the $\xi_{\mu\nu}$ themselves act as the generators of the canonical transformations realizing G on $\mathscr{O}_{0,0}$, we see that these PB's are explicitly G-invariant, and also that this is in agreement with the transformation law (3.5) for ξ under G. The constancy of the two invariants $\mathscr{C}_1$ and $\mathscr{C}_2$ is consistent with this PB structure: $\mathscr{C}_1$ and $\mathscr{C}_2$ are nontrivial neutral elements in this classical canonical realization of **G**, so the above PB (which is in fact a generalized PB) is singular.[19] The PB between any two functions f and g on $\mathscr{O}_{0,0}$ can be computed from the basic ones (4.3) by using the derivation property.

For practical calculations it is convenient to deal with the three-dimensional components $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ of $\xi_{\mu\nu}$. The orbit $\mathscr{O}_{0,0}$ is then defined as

$$\mathscr{O}_{0,0} = \{(\boldsymbol{\xi},\boldsymbol{\eta})\,|\,|\boldsymbol{\xi}| = |\boldsymbol{\eta}| > 0,\ \boldsymbol{\xi}\cdot\boldsymbol{\eta} = 0\}, \tag{4.4}$$

and the PB's (4.3) read

$$\{\xi_j,\xi_k\} = -\{\eta_j,\eta_k\} = \epsilon_{jkl}\xi_l,\quad \{\xi_j,\eta_k\} = \epsilon_{jkl}\eta_l\,. \tag{4.5}$$

If some local choice of independent variables among $\boldsymbol{\xi}$, $\boldsymbol{\eta}$ is represented by q^r, $r = 1,\ldots,4$, then the PB of any two functions $\mathscr{O}_{0,0}$ is, again locally, expressible in the form

$$\{f(q),g(q)\} = \omega^{rs}(q)\,\frac{\partial f}{\partial q^r}\,\frac{\partial g}{\partial q^s}\,. \tag{4.6}$$

The inverse $(\omega_{rs}(q))$ of the matrix $(\omega^{rs}(q))$ gives the local components of the symplectic two-form ω. The KKS theorem assures us of the existence of $\omega_{rs}(q)$ and also that it can locally be written as

$$\omega_{rs}(q) = \frac{\partial\theta_s(q)}{\partial q^r} - \frac{\partial\theta_r(q)}{\partial q^s}\,. \tag{4.7}$$

The question is whether $\theta_r(q)dq^r$ is globally defined.

As a first step towards the choice of q's, we switch over to a new but still overcomplete system of variables $\boldsymbol{\xi}$ and $\hat{\eta} = \boldsymbol{\eta}/|\boldsymbol{\eta}|$. The PB's among these have the somewhat simpler form of the $E(3)$ algebra as compared to (4.5)[23]:

$$\{\xi_j,\xi_k\} = \epsilon_{jkl}\xi_l,\quad \{\xi_j,\hat{\eta}_k\} = \epsilon_{jkl}\hat{\eta}_l,\quad \{\hat{\eta}_j,\hat{\eta}_k\} = 0\,. \tag{4.8}$$

Choosing the q^r on a suitable portion of $\mathscr{O}_{0,0}$ as

$$q^1 = \xi_1,\quad q^2 = \xi_2,\quad q^3 = \hat{\eta}_1,\quad q^4 = \hat{\eta}_2\,, \tag{4.9}$$

with ξ_3 and $\hat{\eta}_3$ determined by

$$\hat{\eta}_3 = (1 - \hat{\eta}_1^2 - \hat{\eta}_2^2)^{1/2},\quad \xi_3 = -(\xi_1\hat{\eta}_1 + \xi_2\hat{\eta}_2)/\hat{\eta}_3\,, \tag{4.10}$$

the matrix $(\omega^{rs}(q))$ is

$$(\omega^{rs}(q)) = \begin{pmatrix} 0 & \xi_3 & 0 & \hat{\eta}_3 \\ -\xi_3 & 0 & -\hat{\eta}_3 & 0 \\ 0 & \hat{\eta}_3 & 0 & 0 \\ -\hat{\eta}_3 & 0 & 0 & 0 \end{pmatrix}. \tag{4.11}$$

The local components of the symplectic two-form are obtained by inverting this matrix:

$$(\omega_{rs}(q)) = \begin{pmatrix} 0 & 0 & 0 & -1/\hat{\eta}_3 \\ 0 & 0 & 1/\hat{\eta}_3 & 0 \\ 0 & -1/\hat{\eta}_3 & 0 & \xi_3/\hat{\eta}_3^2 \\ 1/\hat{\eta}_3 & 0 & -\xi_3/\hat{\eta}_3^2 & 0 \end{pmatrix}. \tag{4.12}$$

Taking $\theta_1 = \theta_2 = 0$ as a trial in Eq. (4.7) we find we can get a solution

$$\theta_3 = \xi_2/\hat{\eta}_3,\quad \theta_4 = -\xi_1/\hat{\eta}_3\,. \tag{4.13}$$

Hence the local one-form θ can be taken to be

$$\theta = \theta_r(q)dq^r = (\xi_2\,d\hat{\eta}_1 - \xi_1\,d\hat{\eta}_2)/\hat{\eta}_3\,. \tag{4.14}$$

The local representability of ω as $d\theta$ is of course guaranteed by the closure of ω and Poincaré's lemma. But now switching back to the variables $\boldsymbol{\xi}$, $\boldsymbol{\eta}$ we see that θ can be simplified to

$$\theta = \epsilon_{jkl}\xi_j\eta_k\,d\eta_l/|\boldsymbol{\eta}|^2\,, \tag{4.15}$$

which is globally defined on $\mathscr{O}_{0,0}$ since $|\boldsymbol{\eta}| > 0$. Thus we establish that the two-form ω on $\mathscr{O}_{0,0}$ is indeed exact.

Fortunately the G-invariance of ω passes over into the G-invariance of θ as well. θ is manifestly SO(3) invariant. Under an infinitesimal pure Lorentz transformation with velocity $\mathbf{v}$, $|\mathbf{v}| \ll 1$, the changes in $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are

$$\delta\boldsymbol{\xi} = \mathbf{v}\wedge\boldsymbol{\eta},\quad \delta\boldsymbol{\eta} = -\mathbf{v}\wedge\boldsymbol{\xi}\,. \tag{4.16}$$

The change $\delta\theta$ in θ can now be easily calculated and shown to vanish since both $\mathscr{C}_1$ and $\mathscr{C}_2$ vanish. Thus we verify that in the case of the coset space $Q = G/N_0$, ω is exact, and θ is globally defined and G-invariant.

B. The case $Q=G/F_9$

Any nonexceptional orbit in **G** can be used as a model for this Q. It will soon become evident that the choice $\mathscr{O}_{0,b} \equiv \mathscr{O}_b^{ss}$ for any $b > 0$ is particularly convenient, so we make this choice. In place of Eq. (4.4) we now have

$$\mathscr{O}_{0,b} = \{(\boldsymbol{\xi},\boldsymbol{\eta})\,|\,|\boldsymbol{\eta}|^2 = |\boldsymbol{\xi}|^2 + b^2,\ \boldsymbol{\xi}\cdot\boldsymbol{\eta} = 0\}\,. \tag{4.17}$$

Thus $\boldsymbol{\eta}$ is nowhere vanishing over $\mathscr{O}_{0,b}$, while $\boldsymbol{\xi}$ may vanish. Apart from this difference in the allowed pairs $(\boldsymbol{\xi},\boldsymbol{\eta})$ in $\mathscr{O}_{0,b}$ as compared to $\mathscr{O}_{0,0}$, all of the previous equations (4.3), (4.5)–(4.15)] remain valid, and in fact the expression (4.15) for θ is globally well-defined over $\mathscr{O}_{0,b}$ since the denominator never vanishes. Thus the exactness of ω in the present case is again established. When we examine the behavior of θ under G, however, we do find a difference. While θ is again manifestly SO(3) invariant, under a pure Lorentz transformation it changes by an exact piece:

$$\delta\theta = d(-2b^2\mathbf{v}\cdot\boldsymbol{\eta}/|\boldsymbol{\eta}|^2)\,. \tag{4.18}$$

This is consistent with the G-invariance of ω.

In both cases of FOS's examined in this section, therefore, the same formal expression for the one-form θ in terms of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ is valid, and can be used in the construction of the corresponding Lagrangians. The essential topological differences between G/N_0 and G/F_9 reside in the ranges of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ in the allowed pairs $(\boldsymbol{\xi},\boldsymbol{\eta})$.

V. LAGRANGIANS AND DYNAMICS FOR PARTICLES WITH FOS

We consider now the problem of constructing the most general Lagrangian for a classical indecomposable object whose internal space Q is one of the two FOS's G/N_0 or G/F_9, consistent with the conditions spelled out in Sec. II. In either case, the internal variable is an antisymmetric tensor $\xi_{\mu\nu}$ or $(\boldsymbol{\xi},\boldsymbol{\eta})$, with suitable values for $\mathscr{C}_1$ and $\mathscr{C}_2$. Space-time translation invariance for the free system forbids the appearance of x^μ in the Lagrangian, and Q being an FOS determines also the leading term linear in $\dot{\xi}_{\mu\nu}$, namely, the expression (4.2). For both possible FOS's, therefore, we have the general form

$$\mathscr{L}(\dot{x},\xi,\dot{\xi}) = \mathscr{L}_0(\xi,\dot{\xi}) + \mathscr{L}'(\dot{x},\xi), \quad \mathscr{L}_0(\xi,\dot{\xi}) = \boldsymbol{\xi}\cdot\boldsymbol{\eta}\wedge\dot{\boldsymbol{\eta}}/|\boldsymbol{\eta}|^2, \tag{5.1}$$

for the total Lagrangian. From the previous section we know that $\mathscr{L}_0(\xi,\dot{\xi})$ is invariant or quasi-invariant under the action of an infinitesimal Lorentz transformation, depending on whether $Q=G/N_0$ or $Q=G/F_9$, respectively. We now impose on $\mathscr{L}'$ the remaining conditions from Sec. II: (a) it must be Lorentz invariant, and (b) it must be homogeneous of degree 1 in the $\dot{x}^\mu$. These conditions determine the most general form possible for $\mathscr{L}'$:

$$\mathscr{L}'(\dot{x},\xi) = (-\dot{x}^2)^{1/2} f(\cdots\zeta\cdots), \tag{5.2}$$

where $\cdots\zeta\cdots$ are a complete independent set of Lorentz scalars formed out of $\dot{x}^\mu$ and $\xi_{\mu\nu}$, with the property of being homogeneous of degree zero in the $\dot{x}^\mu$, and f is an arbitrary real function of the ζ.

We shall make the physical assumption that in all motions of interest, the velocity vector $\dot{x}^\mu$ is positive timelike, so that $\dot{x}^2<0$. The number of independent ζ's can then be immediately determined geometrically, and turns out to be just one for either choice of Q. The argument is as follows. Since $\dot{x}^\mu$ is timelike, one can always go to its rest frame where it takes the form $(\sqrt{-\dot{x}^2}, \mathbf{0})$. If in this frame one can construct a complete set of SO(3) scalars out of $\xi_{\mu\nu}$, then by suitably rewriting them in a manifestly Lorentz invariant way and also imposing the homogeneity requirement with respect to $\dot{x}^\mu$, a complete set of ζ's is obtained. Now, of the three SO(3) scalars $|\boldsymbol{\xi}|$, $|\boldsymbol{\eta}|$, and $\boldsymbol{\xi}\cdot\boldsymbol{\eta}$ that can be formed out of $\xi_{\mu\nu}$, only one is independent, since $\mathscr{C}_1$ and $\mathscr{C}_2$ have definite numerical values, and this one can be taken to be $|\boldsymbol{\eta}|^2$. The Lorentz invariant expression that reduces to this in the rest frame of $\dot{x}^\mu$ is

$$\zeta = (\xi_{\mu\nu}\dot{x}^\nu)^2/(-\dot{x}^2), \tag{5.3}$$

and is the only ζ variable for either choice of Q.

By expanding the square in Eq. (5.3) we can write ζ as

$$\zeta = t_{\mu\nu}(\xi)\dot{x}^\mu\dot{x}^\nu/(-\dot{x}^2), \quad t_{\mu\nu}(\xi) = \xi_{\mu\rho}\xi_\nu{}^\rho. \tag{5.4}$$

We thus see that for either choice of Q the nontrivial coupling between internal and space-time variables is achieved as follows: a symmetric second-rank tensor $t_{\mu\nu}(\xi)$ is formed on Q, and by contraction with $\dot{x}^\mu\dot{x}^\nu$ and imposing the homogeneity requirement the variable ζ is obtained. Then the most general Lagrangian consistent with all of the requirements is

$$\mathscr{L}(\dot{x},\xi,\dot{\xi}) = \boldsymbol{\xi}\cdot\boldsymbol{\eta}\wedge\dot{\boldsymbol{\eta}}/|\boldsymbol{\eta}|^2 + (-\dot{x}^2)^{1/2}f(\zeta), \tag{5.5}$$

and involves one arbitrary real function of one real argument. While this situation is common to both choices of Q, one is naturally led to ask whether a simpler choice of ζ could be made, i.e., whether for instance a four-vector $V_\mu(\xi)$ could be defined on Q in place of the tensor $t_{\mu\nu}(\xi)$, so that by contraction with a single factor $\dot{x}^\mu$ a more elementary ζ could be obtained. We shall show that this is in fact the case for $Q=G/N_0$, while such a possibility does not exist in the other case $Q=G/F_9$. This again points to the exceptional nature of the FOS $\mathscr{O}_{0,0}$.

On the orbit $\mathscr{O}_{0,0}$ the value of $\xi_{\mu\nu}$ at the representative point J_2+K_1, $\xi^{(0)}_{\mu\nu}$ say, is

$$\boldsymbol{\xi}^{(0)} = (0,1,0), \quad \boldsymbol{\eta}^{(0)} = (-1,0,0). \tag{5.6}$$

A generic point $\xi\in\mathscr{O}_{0,0}$ arises from $\xi^{(0)}$ by a suitable SO(3,1) transformation Λ:

$$\xi_{\mu\nu} = \Lambda_\mu{}^\rho\Lambda_\nu{}^\sigma\xi^{(0)}_{\rho\sigma}. \tag{5.7}$$

Now it is well known that the stability group of the point $\xi^{(0)}$, generated by J_1-K_2 and J_2+K_1, also leaves invariant the lightlike four-vector $V^{(0)}_\mu$ with components

$$V^{(0)}_\mu = (-1,0,0,1). \tag{5.8}$$

At $\xi^{(0)}$ the second-rank symmetric tensor $t^{(0)}_{\mu\nu}$ of Eq. (5.4) can be factorized in terms of $V^{(0)}_\mu$:

$$t_{\mu\nu}(\xi^{(0)}) = t^{(0)}_{\mu\nu} = V^{(0)}_\mu V^{(0)}_\nu. \tag{5.9}$$

Just as for ξ in Eq. (5.7), the tensor $t(\xi)$ is related to $t^{(0)}$ by two factors of the Lorentz matrix Λ. Combining this fact with the decomposition (5.9) valid at $\xi^{(0)}$ we get

$$\begin{aligned} t_{\mu\nu}(\xi) &= \Lambda_\mu{}^\rho\Lambda_\nu{}^\sigma t_{\rho\sigma}(\xi^{(0)}) \\ &= \Lambda_\mu{}^\rho\Lambda_\nu{}^\sigma V^{(0)}_\rho V^{(0)}_\sigma \\ &= V_\mu(\xi)V_\nu(\xi), \\ V_\mu(\xi) &= \Lambda_\mu{}^\rho V^{(0)}_\rho. \end{aligned} \tag{5.10}$$

It is understood here that the element $\Lambda\in$SO(3,1) is such as would carry $\xi^{(0)}$ to ξ. Now the use of the notation $V_\mu(\xi)$ is justified only if in the last line of Eq. (5.10) the use of *any* Λ carrying $\xi^{(0)}$ to ξ gives the *same* result for $V_\mu(\xi)$. But this is indeed so, since as noted above the stability group of $V^{(0)}_\mu$ is not smaller than that of the point $\xi^{(0)}$. This assures us that $V_\mu(\xi)$ is indeed a function of ξ alone, as implied by the notation. A direct definition of V_μ in terms of ξ and a check of its vector nature is also possible. Comparing the two representations (5.4) and (5.10) of $t_{\mu\nu}$, remembering $\mathscr{C}_1=\mathscr{C}_2=0$ and the values (5.8) of $V^{(0)}_\mu$, we find

$$V_0(\xi) = -|\boldsymbol{\xi}| = -|\boldsymbol{\eta}|, \quad \mathbf{V}(\xi) = \boldsymbol{\xi}\wedge\boldsymbol{\eta}/|\boldsymbol{\eta}|. \tag{5.11}$$

To verify that $V_\mu(\xi)$ behaves as a four-vector when ξ transforms as a tensor, we first extend Eq. (4.16) giving the change in ξ induced by a pure infinitesimal Lorentz transformation to expressions in which an infinitesimal spatial rotation is also included:

$$\begin{aligned} \delta\boldsymbol{\xi} &= \mathbf{u}_\wedge \boldsymbol{\xi} + \mathbf{v}_\wedge \boldsymbol{\eta} , \\ \delta\boldsymbol{\eta} &= \mathbf{u}_\wedge \boldsymbol{\eta} - \mathbf{v}_\wedge \boldsymbol{\xi} , \\ &|\mathbf{u}|,|\mathbf{v}| \ll 1 . \end{aligned} \tag{5.12}$$

It then follows from Eqs. (5.11) and (5.12) and use of $\mathscr{C}_1 = \mathscr{C}_2 = 0$ when necessary that

$$\begin{aligned} \delta V_0(\xi) &= \mathbf{v}\cdot\mathbf{V}(\xi) , \\ \delta \mathbf{V}(\xi) &= \mathbf{u}_\wedge \mathbf{V}(\xi) + \mathbf{v} V_0(\xi) , \end{aligned} \tag{5.13}$$

which are just the changes induced in the components of a four-vector by the infinitesimal Lorentz transformation $(\mathbf{u},\mathbf{v})$.

It is worth emphasizing at this point that while a four-vector $V^\mu(\xi)$ can be defined as a function of $\xi_{\mu\nu}$ on $\mathscr{O}_{0,0}$, we cannot invert their roles and express $\xi_{\mu\nu}$ as a function of V^μ. This is because the stability group of $\xi^{(0)}_{\mu\nu}$ is smaller than that of $V^{(0)}_\mu$, and similarly for $\xi_{\mu\nu}$ and $V_\mu(\xi)$. Thus it is the tensor $\xi_{\mu\nu}$ that is the primitive object while $V_\mu(\xi)$ is a derived quantity on $\mathscr{O}_{0,0}$.

An essential simplification in the choice of the variable ζ is thus possible when $Q = G/N_0$. It is possible to construct a (positive lightlike) four-vector $V^\mu(\xi)$ on Q and in place of Eqs. (5.3) and (5.4), we can define

$$\zeta = V_\mu(\xi)\dot{x}^\mu/(-\dot{x}^2)^{1/2} \tag{5.14}$$

and use this as the argument of the function f in Eq. (5.5). The Lagrangian (5.5) is then identical in form to the one developed in Refs. 15 for the spinor model, except for the replacement of $\mathscr{L}_0(\xi,\dot{\xi})$ by another $\mathscr{L}_0$ appropriate for spinor internal variables. Of course, while V_μ and the internal angular momentum $S_{\mu\nu}$ are both bilinear in the components of a Majorana spinor in the spinor model, here V_μ is a function of $\xi_{\mu\nu}$ and $S_{\mu\nu}$ coincides with $\xi_{\mu\nu}$. This extremely close relationship is due to the FOS G/N being a twofold covering of the FOS $\mathscr{O}_{0,0} = G/N_0$. Furthermore, if one restricts attention to nonspinorial quantities, the dynamics and constraint structure are also identical in the two cases, and this is in particular so for the mass-spin relation and for the space-time trajectory. These aspects will therefore not be discussed any further here.

Turning to the other FOS $Q = G/F_9$, a corresponding simplification is not possible for the following reason: there is no nonzero four-vector whose stability group is at least as large as F_9. Therefore, the tensor $t_{\mu\nu}(\xi)$ is the simplest object available for combining with $\dot{x}^\mu$ to form a Lorentz scalar ζ. We therefore call this model with the first-order internal space $Q = G/F_9$ the symmetric tensor model (STM) and go on to discuss its constraint structure and Hamiltonian dynamics.

The Lagrangian for the STM is given in Eq. (5.5), with ζ defined in Eqs. (5.3) and (5.4). In order to give the equations a neat appearance, we introduce the following two four-vectors:

$$u^\mu = \dot{x}^\mu/(-\dot{x}^2)^{1/2}, \quad U_\mu = \xi_{\mu\nu}u^\nu , \tag{5.15}$$

obeying

$$u^2 = -1, \quad U^2 = \zeta, \quad u\cdot U = 0 . \tag{5.16}$$

The 12-dimensional phase space for this problem is $T^*\mathscr{M}\times Q$, for which x^μ, $\xi_{\mu\nu}$, and p_μ can be used as an overcomplete system of coordinates. The invariants relating to $\xi_{\mu\nu}$ have values $\mathscr{C}_1 = -b^2$, $\mathscr{C}_2 = 0$. The momenta p_μ are conjugate to the x^μ and the basic nonvanishing PB's are as in Eq. (4.3) together with

$$\{x^\mu, p_\nu\} = \delta^\mu{}_\nu . \tag{5.17}$$

Consistent with the behavior of $\mathscr{L}$ under infinitesimal Poincaré transformations, the conserved generators of $\mathscr{P}$ are

$$P^\mu = p^\mu, \quad J_{\mu\nu} = x_\mu p_\nu - x_\nu p_\mu + \xi_{\mu\nu} . \tag{5.18}$$

To relate the phase space momenta p_μ to Lagrangian quantities, we calculate the derivative of $\mathscr{L}$ with respect to $\dot{x}^\mu$ and get

$$p_\mu = (2\zeta f' - f)u_\mu - 2f'\xi_{\mu\nu}U^\nu . \tag{5.19}$$

Here the prime on f denotes the derivative with respect to the argument ζ (which is omitted). The Pauli–Lubanski vector W_μ and the two Casimir invariants for the Poincaré algebra turn out to be

$$\begin{aligned} W_\mu &= \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}P^\nu J^{\rho\sigma} = (2\zeta f' - f)\xi^*_{\mu\nu}u^\nu ; \\ P^2 &= p^2 = 4\zeta(\zeta - b^2)f'^2 - f^2 ; \\ W^2 &= (\zeta - b^2)(2\zeta f' - f)^2 . \end{aligned} \tag{5.20}$$

Here, $\xi^*_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\xi^{\rho\sigma}$ is the dual to ξ, and apart from putting in the values of $\mathscr{C}_1$ and $\mathscr{C}_2$ where necessary, several identities obeyed by antisymmetric tensors have been used.[24] We see that for a given choice of the arbitrary function f in the Lagrangian, both P^2 and W^2 are functions of ζ. Thus eliminating ζ we get a functional relationship between P^2 and W^2, which is the phase space constraint resulting from the reparametrization invariance of the action. We express this relation in the form

$$\varphi = p^2 - \alpha(W^2) \approx 0 . \tag{5.21}$$

This is the sole primary constraint in the theory, hence it is first class, which is as it should be since it generates reparametrization (gauge) transformations. The Hamiltonian is an arbitrary multiple of the constraint φ, involving a Lagrange multiplier v[5]:

$$H = v\varphi . \tag{5.22}$$

It leads to the equations of motion

$$\begin{aligned} \dot{p}_\mu &= 0 , \\ \dot{x}^\mu &= 2v(p^\mu + \alpha'\xi^{*\mu\nu}W_\nu) , \\ \dot{\xi}_{\mu\nu} &= 2v\alpha'\epsilon^{\rho\sigma\alpha\beta}W_\rho p_\sigma(g_{\alpha\mu}\xi_{\beta\nu} - g_{\alpha\nu}\xi_{\beta\mu}) , \end{aligned} \tag{5.23}$$

where the prime on α denotes the derivative with respect to the argument W^2 (which is omitted). In solving these equations we can take advantage of the fact that p_μ and W_μ are both conserved, so the factor α' is also constant with respect to s. Therefore the only explicit appearance of s is through the Lagrange multiplier v. We begin by writing the equation of motion for $\xi_{\mu\nu}$ in the following form:

$$\begin{aligned} \dot{\xi}_{\mu\nu} &= 2\alpha' v(s)\,(\omega_\mu{}^\lambda\xi_{\lambda\nu} + \omega_\nu{}^\lambda\xi_{\mu\lambda}) , \\ \omega_{\mu\nu} &= \epsilon_{\mu\nu\rho\sigma}W^\rho p^\sigma . \end{aligned} \tag{5.24}$$

Since the antisymmetric tensor $\omega_{\mu\nu}$ is constant with respect to s, this way of writing the equation of motion reveals that $\xi_{\mu\nu}$ evolves with respect to s by being subjected to a continuously changing element on a one-parameter subgroup of SO(3,1), the one with $\omega_{\mu\nu}$ as generator. The rate at which this subgroup is traversed is determined by the Lagrange multiplier v. Thus we can write the solution for the equation of motion for $\xi_{\mu\nu}$ if we can write down a closed form expression for a general element on this one-parameter subgroup. For a variable parameter ψ, let

$$\Lambda(\psi) = \exp(\psi\omega) \in \mathrm{SO}(3,1)\,. \tag{5.25}$$

The two invariants associated with ω have values

$$\begin{aligned} &\tfrac{1}{2}\,\omega^{\mu\nu}\omega_{\mu\nu} = -p^2W^2 > 0\,,\\ &\epsilon_{\mu\nu\rho\sigma}\omega^{\mu\nu}\omega^{\rho\sigma} = 0\,. \end{aligned} \tag{5.26}$$

Therefore, as a result of standard identities for antisymmetric tensors,[24] ω as a generator matrix in the vector representation of SO(3,1) obeys the polynomial equation

$$\omega^3 = -\kappa^2\omega, \quad \kappa = (-p^2W^2)^{1/2} > 0\,. \tag{5.27}$$

This property, together with the result for ω^2,

$$\begin{aligned} (\omega^2)^\mu{}_\nu &= \omega^\mu{}_\rho\,\omega^\rho{}_\nu\\ &= p^2W^2\delta^\mu{}_\nu - p^2W^\mu W_\nu - W^2p^\mu p_\nu\,, \end{aligned} \tag{5.28}$$

allows evaluation of $\Lambda(\psi)$ in closed form:

$$\begin{aligned} \Lambda(\psi)^\mu{}_\nu &= \left[1 + \frac{\sin\kappa\psi}{\kappa}\,\omega + \frac{1-\cos\kappa\psi}{\kappa^2}\,\omega^2\right]^\mu{}_\nu\\ &= \cos\kappa\psi\,\delta^\mu_\nu + \frac{\sin\kappa\psi}{\kappa}\,\epsilon^\mu{}_{\nu\rho\sigma}W^\rho p^\sigma\\ &\quad + (1-\cos\kappa\psi)\left(\frac{W^\mu W_\nu}{W^2} + \frac{p^\mu p_\nu}{p^2}\right). \end{aligned} \tag{5.29}$$

The geometrical interpretation is aided by the properties

$$\Lambda(\psi)^\mu{}_\nu p^\nu = p^\mu,\ \Lambda(\psi)^\mu{}_\nu W^\nu = W^\mu\,. \tag{5.30}$$

Thus in the rest frame of p^μ, $\Lambda(\psi)$ is a purely spatial rotation around **W** as axis. The equation of motion for $\xi_{\mu\nu}$ can now be solved by making ψ a function of s obeying the appropriate differential equation:

$$\begin{aligned} &\xi^{\mu\nu}(s) = \Lambda(\psi(s))^\mu{}_\rho\Lambda(\psi(s))^\nu{}_\sigma\xi^{\rho\sigma}(0)\,,\\ &\dot\psi(s) = 2\alpha' v(s)\,,\\ &\psi(0) = 0\,. \end{aligned} \tag{5.31}$$

On using this result for $\xi_{\mu\nu}(s)$ and also the second of Eqs. (5.30), the equation of motion for x^μ can be simplified to

$$\begin{aligned} \dot x^\mu(s) &= p^\mu\frac{d}{ds}\left(\frac{\psi(s)}{\alpha'}\right)\\ &\quad + \frac{d\psi(s)}{ds}\,\Lambda(\psi(s))^\mu{}_\rho\,\xi^{*\rho\sigma}(0)\,W_\sigma\,. \end{aligned} \tag{5.32}$$

This can be easily integrated since the explicit form of $\Lambda(\psi)$ is available in Eq. (5.29):

$$\begin{aligned} x^\mu(s) &= x^\mu(0) + p^\mu\psi(s)/\alpha'\\ &\quad + \left[\psi(s)\cdot 1 + \frac{1-\cos\kappa\psi(s)}{\kappa^2}\,\omega\right.\\ &\quad \left. + \frac{\kappa\psi(s) - \sin\kappa\psi(s)}{\kappa^3}\,\omega^2\right]^\mu{}_\rho\,\xi^{*\rho\sigma}(0)\,W_\sigma\,. \end{aligned} \tag{5.33}$$

Thus we see that the space-time trajectory is influenced by the internal variables $\xi_{\mu\nu}$ even for the isolated system, and $x^\mu(s)$ follows a helical path in space-time.

Up to this point the reparametrization invariance has been maintained and the choice of $v(s)$ left free. If we now for instance choose the evolution parameter s to be physical time by imposing the gauge constraint

$$\chi = x^0 - s \approx 0\,, \tag{5.34}$$

on the one hand $v(s)$ gets determined and on the other hand the two constraints φ, χ form a second class set. This allows passage to a system of Dirac brackets (DB)[5] and explicit elimination of two phase space degrees of freedom. For instance the remaining physical phase space variables can be taken to be **x**, **p**, and $\xi_{\mu\nu}$ subject to $\mathscr{C}_1 = -b^2, \mathscr{C}_2 = 0$. It is then easily seen that the DB's among the x_j vanish,

$$\{x_j, x_k\}^* = 0\,, \tag{5.35}$$

which is thus a property of the STM shared with the spinor model.[15]

VI. CONCLUDING REMARKS

In this paper, the first in a series devoted to a systematic study of classical relativistic particles with internal structure, we have described the physical ideas guiding our study, and then dealt in detail with internal structures corresponding to first-order spaces. These are the cases where, in the Lagrangian formalism, the internal variables can be described by a first-order dynamics. After determining the possible FOS's, the symmetric tensor model based on a generic orbit in **G** as the internal space was examined in detail. It is interesting to make the following remarks concerning this model and its relation to the spinor model or the model based on the exceptional orbit as internal space. In any of these models, there is one arbitrary function of one argument which appears in the general Lagrangian, which ultimately determines the functional relationship between mass and spin characteristic of the model. Let this Regge relationship be written in the form

$$m = \beta(s)\,, \tag{6.1}$$

with m the invariant mass $\sqrt{-P^2}$, and s the magnitude of the intrinsic spin $\sqrt{-W^2/P^2}$. One can ask whether, if this relationship alone were given, one would have a clue as to which internal space was involved. It is plausible that this is not so, since in any one of the models the input arbitrary function in the Lagrangian could always be so chosen as to lead to the preassigned Regge relation. However it is interesting that by coupling the system to an external electromagnetic field and calculating, for example, the magnetic moment for particles on the Regge trajectory as a function of the spin, this "degeneracy" can be lifted. For both the spinor model and the model with $Q = G/N_0$ (the case of the exceptional orbit), the dependence of the magnetic moment on spin is determined in terms of the Regge relation as[15]

$$g(s) = \frac{d \ln \beta(s)}{d \ln s}. \tag{6.2}$$

However in the case of $Q = G/F_9$ as internal space, realized as in this paper by using the orbit $\mathscr{O}_{0,b}$ in G where $b > 0$ is an additional parameter, the magnetic moment has been calculated elsewhere[25] and turns out to be

$$g(s) = \frac{b^2 + s^2}{s^2} \frac{d \ln \beta(s)}{d \ln s}. \tag{6.3}$$

This is both reassuring and intriguing for the following reasons. On the one hand it is reassuring because it is possible to distinguish between the different internal spaces on a physical basis. It is on the other hand intriguing because the internal space $Q = G/F_9$ does not vary in any intrinsic sense as b varies, therefore the result (6.3) shows a dependence on the particular way in which the given internal space has been realized. This fact suggests the following question, which is worth pursuing: if the same internal space $Q = G/F_9$ were to be realized as one of the orbits $\mathscr{O}_{a,b}$ or $\mathscr{O}_{a,0}$, and this must certainly be possible, how is the result (6.3) altered?

In the next paper in this series we take up the study of SOS's. The number of these is quite large, and we shall have to develop special methods to handle them in a systematic manner.

APPENDIX: LIE SUBGROUPS OF G

We list here the connected Lie subgroups H of $G = \mathrm{SL}(2,C)$, up to conjugacy.[17] The extreme cases $F_1 = G$ and $F_{15} = \{e\}$ can be omitted. The rest, ranging from F_2 to F_{14}, are given in opposite order, since that is the order of increasing dimension n of the subgroup. The usual notations for generators of SU(2) and of pure Lorentz transformations are J_j and K_j respectively, while the notation J_φ will be used for the combination

$$J_\varphi = \sin \varphi J_3 + \cos \varphi K_3, \quad 0 < \varphi < \pi/2 \text{ or } \pi/2 < \varphi < \pi. \tag{A1}$$

For each subgroup listed, both the corresponding collection of matrices in the defining $(\frac{1}{2},0)$ representation of G, and the infinitesimal generators, are given.

For $n = 1$,

$$F_{14} = \text{all matrices} \begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix},$$

$$-\infty < r < \infty, \quad \text{generator} = J_2 + K_1;$$

$$F_{13} = A = \text{all matrices} \begin{pmatrix} e^v & 0 \\ 0 & e^{-v} \end{pmatrix},$$

$$-\infty < v < \infty, \quad \text{generator} = K_3;$$

$$F_{12} = U(1) = \text{all matrices} \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix},$$

$$0 \leqslant \theta < 2\pi, \quad \text{generator} = J_3;$$

$$F_{11}^\varphi = \text{all matrices} \begin{pmatrix} e^{ve^{i\varphi}} & 0 \\ 0 & e^{-ve^{i\varphi}} \end{pmatrix},$$

$$-\infty < v < \infty, \quad \varphi \text{ fixed}, \quad \text{generator} = J_\varphi.$$

For $n = 2$,

$$F_{10} = N = \text{all matrices} \begin{pmatrix} 1 & r + is \\ 0 & 1 \end{pmatrix},$$

$$-\infty < r,s < \infty, \quad \text{generators} = J_1 - K_2, J_2 + K_1;$$

$$F_9 = U(1)A = \text{all matrices} \begin{pmatrix} e^{v+i\theta} & 0 \\ 0 & e^{-v-i\theta} \end{pmatrix},$$

$$-\infty < v < \infty, \quad 0 \leqslant \theta < 2\pi; \quad \text{generators} = J_3, K_3;$$

$$F_8 = \text{all matrices} \begin{pmatrix} e^v & r \\ 0 & e^{-v} \end{pmatrix},$$

$$-\infty < v,r < \infty; \quad \text{generators} = K_3, J_2 + K_1.$$

For $n = 3$,

$$F_7 = \mathrm{AN} = \text{all matrices} \begin{pmatrix} e^v & r + is \\ 0 & e^{-v} \end{pmatrix},$$

$$-\infty < v,r,s < \infty,$$

$$\text{generators} = K_3, J_1 - K_2, J_2 + K_1;$$

$$F_6 = U(1)N = \text{all matrices} \begin{pmatrix} e^{i\theta} & r + is \\ 0 & e^{-i\theta} \end{pmatrix},$$

$$-\infty < r,s < \infty, \quad 0 \leqslant \theta < 2\pi,$$

$$\text{generators} = J_3, J_1 - K_2, J_2 + K_1;$$

$$F_5^\varphi = \text{all matrices} \begin{pmatrix} e^{ve^{i\varphi}} & r + is \\ 0 & e^{-ve^{i\varphi}} \end{pmatrix},$$

$$-\infty < v,r,s < \infty, \quad \varphi \text{ fixed},$$

$$\text{generators} = J_\varphi, J_1 - K_2, J_2 + K_1;$$

$$F_4 = \mathrm{SU}(1,1) = \text{all matrices} \begin{pmatrix} \lambda & \mu \\ \mu^* & \lambda^* \end{pmatrix},$$

$$|\lambda|^2 - |\mu|^2 = 1, \quad \text{generators} = J_3, K_1, K_2;$$

$$F_3 = \mathrm{SU}(2) = \text{all matrices} \begin{pmatrix} \lambda & \mu \\ -\mu^* & \lambda^* \end{pmatrix},$$

$$|\lambda|^2 + |\mu|^2 = 1, \quad \text{generators} = J_1, J_2, J_3.$$

For $n = 4$

$$F_2 = \mathrm{U}(1)\mathrm{AN} = \text{all matrices} \begin{pmatrix} e^{v+i\theta} & r + is \\ 0 & e^{-v-i\theta} \end{pmatrix},$$

$$-\infty < v,r,s < \infty, \quad 0 \leqslant \theta < 2\pi,$$

$$\text{generators} = J_3, K_3, J_1 - K_2, J_2 + K_1.$$

[1] A. J. Hanson and T. Regge, Ann. Phys. (NY) **87**, 498 (1974).
[2] C. Duval and P. Horvathy, Ann. Phys. (NY) **142**, 10 (1982).
[3] V. Bargmann, L. Michel, and V. L. Telegdi, Phys. Rev. Lett. **2**, 435 (1959).
[4] H. van Dam and T. W. Ruijgrok, Physica A**104**, 281 (1980).
[5] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University—Belfer Graduate School of Science, New York, 1964).
[6] J. Frenkel, Z. Phys. **37**, 243 (1926).
[7] M. Mathisson, Acta. Phys. Pol. **6**, 163, 218 (1937).
[8] L. Lubanski, Acta Phys. Pol. **6**, 356 (1937).
[9] H. Honl and A. Papapetrou, Z. Phys. **112**, 512 (1938); **114**, 478 (1939).
[10] H. J. Bhabha and H. C. Corben, Proc. R. Soc. London Ser. A **178**, 243 (1941).
[11] H. C. Corben, *Classical and Quantum Theories of Spinning Particles* (Holden–Day, San Francisco, 1968).
[12] F. Halbwachs, Prog. Theor. Phys. **24**, 291 (1960).
[13] C. Itzykson and A. Voros, Phys. Rev. D **5**, 2939 (1972).

[14]K. Rafanelli, Phys. Rev. D **9**, 2746 (1974).
[15]N. Mukunda, H. van Dam, and L. C. Biedenharn, Phys. Rev. D **22**, 1938 (1980); D **23**, 1451 (1981); "Relativistic models of extended hadrons obeying a mass-spin trajectory constraint," in *Lecture Notes in Physics* (Springer, Berlin, 1982), Vol. 165.
[16]M. V. Atre and N. Mukunda, Curr. Sci. **54**, 610 (1985).
[17]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597 (1975).
[18]D. Finkelstein, Phys. Rev. **100**, 924 (1955).
[19]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).
[20]A. P. Balachandran, G. Marmo, B. S. Skagerstam, and A. Stern, "Gauge symmetries and fiber bundles," *Lecture Notes in Physics* (Springer, Berlin, 1984), **188**.
[21]B. Y. Chu, Trans. Am. Math. Soc. **197**, 45 (1974).
[22]P. A. M. Dirac, Proc. R. Soc. London Ser A **322**, 435 (1971), A **328**, 1 (1972).
[23]P. Chand, C. L. Mehta, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **8**, 2048 (1967).
[24]See, for instance, the relevant appendix in Ref. 1.
[25]M. V. Atre, Ph. D. thesis, Indian Institute of Science, Bangalore, India, 1986.

Kepler problem with a magnetic monopole

Bruno Cordani
Dipartimento di Matematica dell'Università, via Saldini 50, 20133 Milano, Italy

(Received 2 April 1986; accepted for publication 14 August 1986)

It is shown that the usual moment map J: $T^*(\mathbb{R}^3 - \{0\}) \mapsto \mathrm{so}^*(2,4)$ of the Kepler problem can be generalized to include the magnetic term of the Dirac monopole.

I. INTRODUCTION

The maximal dynamical group of the n-dimensional Kepler problem (or the hydrogen atom) is well known to be $SO(2,n+1)$.[1-5] So we have an Ad*-equivariant moment map J: $T^*(\mathbb{R}^n - \{0\}) \mapsto \mathrm{so}^*(2,n+1)$,[6,7] whose Hamiltonians are given by

$$\begin{aligned} J_{hk} &= x_h y_k - x_k y_h, \\ L_k &= x y_k, \\ A_0 &= -\tfrac{1}{2}x(y^2-1), \\ A_k &= \tfrac{1}{2}(y^2-1)x_k - \langle x,y\rangle y_k, \\ B_0 &= -\tfrac{1}{2}x(y^2+1), \\ B_k &= \tfrac{1}{2}(y^2+1)x_k - \langle x,y\rangle y_k, \\ D &= \langle x,y\rangle, \end{aligned} \tag{1.1}$$

where $y^2 = \langle y,y\rangle$, and $x = \langle x,x\rangle^{1/2}$, $h,k = 1,...,n$. The x_h and y_k are canonical coordinates on $T^*(\mathbb{R}^n - \{0\})$, which is the phase space of the KP. The nonlinear action of $SO(2,n+1)$ on $T^*(\mathbb{R}^n - \{0\})$ can be linearized by means of the double covering $\mathrm{Spin}(2,n+1)$ of the pseudo-orthogonal group.[8] In the physical case, i.e., $n = 3$, we have the isomorphism $\mathrm{Spin}(2,4) = SU(2,2)$ and the linear symplectic action of $SU(2,2)$ on $T_0/\approx$ induces the moment map (1.1) through the Kustaanheimo–Stiefel (KS) transformation,[9,10]:

$$x_k = z^\dagger \sigma_k z, \quad y_k = \mathrm{Im}(z^\dagger \sigma_k w)/z^\dagger z. \tag{1.2}$$

Here T_0 is the space of the null twistors, i.e., the elements $\psi = \binom{z}{w}$, $z \in \mathbb{C}^2 - \{0\}$, $w \in \mathbb{C}^2$ such that

$$\psi^\dagger \mathscr{E} \psi = 0, \tag{1.3}$$

where

$$\mathscr{E} = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}, \tag{1.4}$$

and σ_μ are the Pauli matrices. The equivalence relation $\approx$ is defined by $\psi \approx \psi \exp(i\theta)$.

The first of Eqs. (1.2) shows that the KS transformation is based on the Hopf fibration: $S^3 \mapsto S^2$. Since this fibration is not trivial, on the basis of Theorem 2 below, one expects that the manifold $T^*(\mathbb{R}^3 - \{0\})$ is equipped with a symplectic form differing from the canonical one by a magnetic term. Really, Barut and Bornzin[12] showed, by a "direct, though laborious, calculation" that (1.1) can be generalized as follows:

$$\begin{aligned} \mathbf{J} &= \mathbf{x} \times \boldsymbol{\pi} - \mu \mathbf{x}/x, \quad \mathbf{L} = x\boldsymbol{\pi}, \\ A_0 &= -\tfrac{1}{2}x(\pi^2 - 1) - \tfrac{1}{2}(\mu^2/x), \\ \mathbf{A} &= \tfrac{1}{2}(\pi^2 - 1)\mathbf{x} - (\mathbf{x}\cdot\boldsymbol{\pi})\boldsymbol{\pi} + \frac{\mu}{x}\mathbf{J} + \frac{1}{2}\frac{\mu^2}{x^2}\mathbf{x}, \\ B_0 &= -\tfrac{1}{2}x(\pi^2 + 1) - \frac{1}{2}(\mu^2/x), \\ \mathbf{B} &= \frac{1}{2}(\pi^2 + 1)\mathbf{x} - (\mathbf{x}\cdot\boldsymbol{\pi})\boldsymbol{\pi} + \frac{\mu}{x}\mathbf{J} + \frac{1}{2}\frac{\mu^2}{x^2}\mathbf{x}, \\ D &= \mathbf{x}\cdot\boldsymbol{\pi} \end{aligned} \tag{1.5}$$

where now the π_h are not canonical coordinates since

$$\{\pi_h, \pi_k\} = \mu \epsilon_{hki}(x_i/x^3). \tag{1.6}$$

Here $\{\cdot,\cdot\}$ are the Poisson brackets, and μ is a parameter (magnetic charge) whose vanishing implies that (1.5) reduce to (1.1).

The aim of this paper is to recover in a natural way the moment map (1.5). In Sec. II we shortly recall two theorems on the reduction of the symplectic manifolds. In Sec. III we show how the moment map (1.5) is obtained considering the linear symplectic action of $U(2,2)$ of $T_\mu/\approx$, where T_μ is the space of the twistors of constant modulus μ.

II. REDUCTION OF A SYMPLECTIC MANIFOLD

This first theorem is due to Marsden and Weinstein.[13]

Theorem 1: Let (P,ω) by a symplectic manifold on which a Lie group G acts symplectically and let J: $P \mapsto g^*$ be and Ad*-equivariant moment map. Assume $\mu \in g^*$ is a regular value of J and that the isotropy subgroup G_μ acts freely and properly on $J^{-1}(\mu)$. Then $P_\mu := J^{-1}(\mu)/G_\mu$ is a symplectic manifold with a form ω_μ such that $\pi_\mu^* \omega_\mu = i_\mu^* \omega$, where $\pi_\mu : J^{-1}(\mu) \mapsto P_\mu$ is the canonical projection and i_μ: $J^{-1}(\mu) \mapsto P$ is the inclusion. Let H: $P \mapsto \mathbb{R}$ be G invariant: it induces a Hamiltonian flow on P_μ with Hamiltonian H_μ satisfying $H_\mu \circ \pi_\mu = H \circ i_\mu$.

In a case, the reduced symplectic manifold P_μ is identifiable more exactly. In fact we have the following theorem, due in this form to Kummer.[14]

Theorem 2: Let P be a cotangent bundle T^*M and G a one-parameter Lie group acting freely and properly on M. Let $M \mapsto N = M/G$ be the induced principal fiber bundle and α be a connection one-form on it. The reduced manifold P_μ is symplectomorphic to T^*N endowed with a symplectic form given by the canonical one plus a "magnetic term" $\mu \tau_N^* \, d\alpha$ (where τ_N is the canonical projection $T^*N \mapsto N$).

III. KEPLER MANIFOLD WITH A MAGNETIC MONOPOLE

Let $\mathscr{E}$ be a matrix representation of the U(2,2)-invariant Hermitian form. We choose a basis in $\mathbb{C}^{2,2}$ such that $\mathscr{E}$ has the form (1.4). Here $\mathbb{C}^{2,2}$ is equipped with a natural symplectic form $\omega = d\Theta$, where

$$\Theta = (i/2)(\psi^\dagger \mathscr{E}\, d\psi - d\psi^\dagger\, \mathscr{E}\psi), \tag{3.1}$$

and $\psi \in \mathbb{C}^{2,2}$. The linear action of U(2,2) on $\mathbb{C}^{2,2}$ is manifestly symplectic. The associated moment map: $\mathbb{C}^{2,2} \mapsto u^*(2,2)$ is easily found to be

$$i\psi\psi^\dagger \mathscr{E} = i\begin{pmatrix} zw^\dagger & zz^\dagger \\ ww^\dagger & wz^\dagger \end{pmatrix}, \tag{3.2}$$

where we set $\psi = \binom{z}{w}$, $z \neq 0$. The action of the center U(1) of U(2,2) is free on $(\mathbb{C}^2 - \{0\}) \oplus \mathbb{C}^2$ and induces the reduction of any submanifold T_μ of twistors of constant modulus

$$\psi^\dagger \mathscr{E} \psi = \mu \tag{3.3}$$

to $T_\mu/\approx$. We will prove that the moment map: $T_\mu/\approx \mapsto su^*(2,2)$ is given just by (1,5). To this end, choose a system of local coordinates on $T_\mu/\approx$ as follows. Let $\Xi = \frac{1}{2}(x\sigma_0 + \mathbf{x}\cdot\boldsymbol{\sigma})$, and $\Pi = (\mu/x)\sigma_0 + \boldsymbol{\pi}\cdot\boldsymbol{\sigma}$. Being $\det \Xi = 0$, we can define $\Xi^{1/2}$ as an element of $(\mathbb{C}^2 - \{0\})/\approx$ such that $\Xi^{1/2}\Xi^{\dagger\,1/2} = \Xi$. The $\mathbf{x}$ and $\boldsymbol{\pi}$ are local coordinates on $T_\mu/\approx$. Indeed, setting

$$\psi = \begin{pmatrix} \Xi^{1/2} \\ i\Pi\Xi^{1/2} \end{pmatrix}, \tag{3.4}$$

Eq. (3.3) is identically satisfied. The pullback of Θ gives

$$\Theta = \boldsymbol{\pi}\cdot d\mathbf{x} + \mu A, \tag{3.5}$$

where A is a one-form such that

$$dA = \epsilon_{hki}(x_i/x^3)dx_h\, dx_k. \tag{3.6}$$

Therefore the symplectic structure on $T_\mu/\approx$ is given by

$$\{x_h, x_k\} = 0, \quad \{x_h, \pi_k\} = \delta_{hk}, \quad \{\pi_h, \pi_k\} = \mu\epsilon_{hki}x_i/x^3. \tag{3.7}$$

It is now straightforward to verify that composing (3.2) and (3.4) we obtain (1.5).

We have accomplished the reduction process following Theorem 1. Alternatively, we can follow Theorem 2. In this way, we get a better insight as to the role of the one-form A. To this purpose, let M be $\mathbb{C}^2 - \{0\} = \mathbb{R}^4 - \{0\}$ and (z,w) coordinates on $(\mathbb{C}^2 - \{0\}) \oplus \mathbb{C}^2 = T^*M$. We have the action of U(1) on M given by

$$z \mapsto z \exp(i\beta/2), \tag{3.8}$$

and we can apply Theorem 2. Regard M as $\mathbb{R}^+ \times S^3$ so that the U(1)-action on M gives an induced action on S^3; its quotient is $S^2 = \mathbb{CP}^1$. Thus $N = \mathbb{CP}^1 \times \mathbb{R}^+$. As is well known, this principal U(1)-bundle $M \mapsto N$ has a natural connection one-form α given by

$$\alpha = \mathrm{Im}(z^\dagger dz)/z^\dagger z. \tag{3.9}$$

When restricted to S^2 (i.e., $z^\dagger z = 1$), α is the Kaehler one-form on $\mathbb{CP}^1$. The one-forms α and A represent the same connection. The action of U(1) commutes with that of SU(2,2) and thus its Hamiltonian is constant. Parametrize z in terms of the spherical coordinates (r,θ,ϕ) on $\mathbb{R}^3 - \{0\}$, getting

$$z = \begin{pmatrix} \sqrt{r}\cos(\theta/2)\exp(i[(\phi+\beta)/2]) \\ \sqrt{r}\sin(\theta/2)\exp(i[(-\phi+\beta)/2]) \end{pmatrix}. \tag{3.10}$$

The angle β is an "ignorable" coordinate for all the Hamiltonians of the SU(2,2)-action and therefore the conjugate momentum is a constant $(=\mu)$. It is easy to verify that the momentum map $T^*(\mathbb{R}^3 - \{0\}) \mapsto su^*(2,2)$, given composing the lift of (3.10) with (3.2), gives just (1.5), as required.

[1] G. Györgyi, Nuovo Cimento A **53**, 717 (1967).
[2] J. M. Souriau, Symp. Math. **14**, 343 (1974).
[3] E. Onofri and M. Pauri, J. Math. Phys. **13**, 533 (1972).
[4] E. Onofri, J. Math. Phys. **17**, 401 (1976).
[5] B. Cordani, Commun. Math. Phys. **103**, 403 (1986).
[6] R. Abraham and J. E. Marsden, *Foundation of Mechanics* (Benjamin Cummings, Reading, MA, 1978).
[7] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).
[8] B. Cordani and C. Reina, Atti Congr. AIMETA, Torino (1986).
[9] P. Kustaanheimo and E. Stiefel, J. Reine Angew. Math. **218**, 204 (1965).
[10] E. Stiefel and G. Scheifele, *Linear and Regular Celestial Mechanics* (Springer, Berlin, 1971).
[11] M. Kummer, Commun. Math. Phys. **84**, 133 (1982).
[12] A. O. Barut and G. L. Bornzin, J. Math. Phys. **5**, 841 (1971).
[13] J. E. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[14] M. Kummer, Indiana Univ. Math. J. **30**, 281 (1981).

Quantum kinematics of the harmonic oscillator

Jorge Krause
Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile

(Received 12 February 1986; accepted for publication 28 July 1986)

The formalism of non-Abelian quantum kinematics is applied to the Newtonian symmetry group of the harmonic oscillator. Within the regular ray representation of the group, the Schrödinger operator, as well as two other (new) invariant operators, are obtained as Casimir operators of the extended kinematic algebra. Superselection rules are then introduced, which permit the identification (and the explicit calculation) of the physical states of the system. Next, a complementary ray representation, attached to the space-time realization of the group, casts the Schrödinger operator into the familiar time-dependent space-time differential operator of the harmonic oscillator and thus, by means of the superselection rules, one obtains the time-dependent Schrödinger equation of the sytem. Finally, the evaluation of a Hurwitz invariant integral, over the group manifold, affords the well known Feynman space-time propagator $\langle t',x'|t,x\rangle$ of the simple harmonic oscillator. Everything comes out from the assumed symmetries of the system. The whole approach is group theoretic and "relativistic." No classical analog is used in this "quantization" scheme.

I. INTRODUCTION

This work concerns non-Abelian quantum kinematics, and arose in the context of an undertaking to interpret the quantum theory of symmetries as the cornerstone of quantum dynamics. In a previous paper[1] (hereafter referred to as paper I) a kinematic formalism has been proposed rather briefly, which one obtains from the regular representation of a Lie group. This formalism was suggested to the author by Weyl's kinematic approach to the group of space translations[2] and, somehow, constitutes a direct generalization thereof. In effect, its main feature consists in the replacement of the essential parameters q^a, $a = 1,..., r$, of a non-Abelian Lie group, by commuting Hermitian operators Q^a, which act within the carrier space of the regular representation (and may be interpreted as generalized "position" operators of the group manifold). In the current applications of Lie groups in quantum mechanics one treats the parameters as c numbers while, of course, the generators P_a (say) of the relevant unitary representations are Hermitian operators by their own right. The only exception to this standard procedure is precisely the group of space translations, which is usually *quantized* in a complete manner (i.e., one introduces Cartesian momentum *and* position operators), and whose regular representation thus plays an outstanding role as the "coordinate representation" of wave mechanics. Accordingly, one expects that the formal features of non-Abelian quantum kinematics should also arise from a *complete group-quantization* scheme. As was shown in paper I, this group quantization procedure (i.e., $q^a \to Q^a$) leads to explicit generalized commutation relations for non-Abelian dynamical variables,[3] as well as to generalized Heisenberg equations of "motion" for the parameter-dependent operators. In this way one treats all dynamical variables as q numbers that stand on the same footing, in accordance with the demands of relativity theory in general. Therefore the kinematic formalism sketched in paper I may afford an essentially new, intrinsically relativistic, approach to quantum mechanics, which would be radically different from the three current mathematical formulations of modern quantum theory.[4]

Following this trend of ideas, it seems possible to consider quantum mechanics as a theory of physical symmetries, which can be suitably formulated in the mathematical language of group theory, without recourse to the analytical models used in classical mechanics. This means that one should try to withdraw the use of classical analogs in quantum theory, as far as possible, at least as a matter of principle. In effect, this attempt cherishes the idea that *canonical quantization* is not the main point in the mathematical construction leading to the quantum model of a physical system. Rather, the symmetry structure shown by the system should be the only guide for having a self-consistent, complete, and unambiguous procedure of quantum mechanical model building. In other words, the very notion of "quantization" should be considered under a completely different perspective, i.e., as a well defined geometric procedure that describes the *observed symmetries* of a mechanical system in terms of some physically significative representations of the corresponding group. We shall refer to this particular attempt as *kinematic quantization*. It is clear that such an approach to quantum theory (if possible at all) would be an interesting achievement for elementary particle physics. There are several contributions following this idea in the recent literature.[5]

In this paper we wish to examine this matter further, working on a concrete example of quantum kinematics. Here we tackle the problem of determining the Hilbert space and deducing the Schrödinger equation of the one-dimensional harmonic oscillator. It should be understood from the beginning that we shall achieve our task by purely group theoretic considerations, since we will use exclusively the information that *the system S is invariant under a given group G*, without considering a prequantized canonical model of *S*. The example we study in this article is quite elementary indeed. However, it is far from trivial, and we deem it as suffi-

 0022-2488/86/122922-14$02.50

ciently rich for searching the huge physical possibilities of non-Abelian quantum kinematics in general. Unfortunately, we have to remark that, given the present stage of development of this issue, examples of quantum kinematics are rather lengthy and somehow annoying, even for such simple mechanical systems as the one considered in this paper.

Specifically, this paper is devoted to the kinematic model one obtains from the regular ray representations of the Euclidean group in two dimensions. In order to arrive at an adequate physical interpretation of the outcoming model, let use recall the isomorphism between the Euclidean group E_2 and the Newtonian group G of the simple harmonic oscillator. As is well known, G is a group of space-time point symmetry transformations of the equation of motion $\ddot{x} + w^2x = 0$. Indeed, the Newtonian group of this differential equation is a three-parameter Lie group that has the following rather simple realization in the space-time $\{t,x\}$ of the system:

$$t' = t + q^1,$$
$$x' = x + q^2 \cos wt + q^3 \sin wt, \tag{1.1}$$

where q^1, q^2, and q^3 are three essential parameters of the group.[6] Equation (1.1) entails a Newtonian transformation of space-time coordinates for it manages time as an absolute universal parameter.[7] Clearly, the change of variables $(t,x) \to (t',x')$ has the fundamental property of leaving invariant the equation of motion of the system; i.e., G is the *Newtonian relativity group* of the one-dimensional harmonic oscillator. Nevertheless, it can be shown quite directly that neither the Lagrangian nor the Hamiltonian are invariant under Eq. (1.1). [To be sure, the Lagrangian changes by a total time derivative, while (on the orbits) the Hamiltonian changes by an additive constant, as it indeed must.] Hence, the transformations stated in Eq. (1.1) can be visualized also as an *active* symmetry group that changes one world line of the system into another. Thus

$$x(t) = \alpha \cos wt + \beta \sin wt \to x'(t')$$
$$= \alpha' \cos wt' + \beta' \sin wt' \tag{1.2}$$

holds upon transformation (1.1), and, in consequence, a simple calculation yields the following realization of G in the *state space* $\{\alpha,\beta\}$ of the classical oscillator:

$$\begin{bmatrix} \alpha' \\ \beta' \end{bmatrix} = \begin{bmatrix} \cos wq^1 & -\sin wq^1 \\ \sin wq^1 & \cos wq^1 \end{bmatrix} \begin{bmatrix} \alpha + q^2 \\ \beta + q^3 \end{bmatrix}. \tag{1.3}$$

So we see that G is isomorphic with E_2.

It is a well known fact that the full symmetry group of the differential equation $\ddot{x} + w^2x = 0$ is an eight-parameter Lie group. In effect, Wulfman and Wybourne[8] considered the classical one-dimensional harmonic oscillator and found that its symmetry group is SL$(3,R)$. Hence, the Euclidean group E_2 acts isomorphically as the Newtonian *subgroup* of the complete space-time symmetry group of the oscillator. (This is exactly in the same way as, for instance, the Galilei group in one-dimensional space acts as the Newtonian subgroup of the eight-parameter projective group, which leaves the equation $\ddot{x} = 0$ invariant.[6]) For the sake of simplicity, in this paper we disregard five extra degrees of freedom of the symmetries of the system, and thus we assume that $G \approx E_2$ gives a sufficiently complete description of the symmetries of the harmonic oscillator. For the same reason, we do not take into account the kinematic structure of the rotational degrees of freedom of the isotropic oscillator, since these would introduce an unnecessary mathematical complication[9] which, at this stage, would not help us to grasp the main physical features of quantum kinematics.

Finally, we wish to recall the interesting discussion of the possible kinematical groups and their classification, expressing the equivalence of a large class of frames of references, which was presented some years ago by Bacry and Lévy-Leblond.[10] The three-parameter Newtonian group G we are considering in the present paper corresponds precisely to the one-dimensional version of the three-dimensional oscillator group introduced by these authors. Also, the one-dimensional Newtonian group of the simple harmonic oscillator, as well as that of the forced harmonic oscillator, has been studied (in connection with the Lagrangian gauge problem of classical mechanics) by Lévy-Leblond[11] and Houard,[12] respectively. Moreover, all the unitary continuous irreducible representations of the central extension (i.e., ray representations) of the oscillator group G can be found in a paper by Streater.[13] In summary, the group G, which we are going to *quantize* in the present work, is a familiar mathematical object indeed.[14]

The organization of this paper is as follows. Section II contains a brief review of the regular representation of the universal covering group of G, and its ray extensions, which is the starting point of the kinematic model. The extended kinematic algebra of the harmonic oscillator is studied in Sec. III; i.e., the Lie algebra and the non-Abelian canonical commutators as well as the invariant operators are discussed in this section. Section IV introduces the superselection rules attached with (and solves the eigenvalue equations of) the invariant operators; in particular, this section is devoted to the Schrödinger operator. In Sec. V we construct a complementary ray representation of the group, which takes into account its Newtonian space-time realization and thus allows us to calculate the (well known) explicit form of the time-dependent Schrödinger differential operator. The physical space-time kets $|t,x\rangle$ are obtained in Sec. VI by means of the superselection rules. Thus the time-dependent Schrödinger equation of the harmonic oscillator arises as a purely group theoretic construct. Moreover, the space-time probability amplitude $\langle t',x'|t,x\rangle$ is evaluated by means of an invariant integral over the group manifold. Section VII contains some conclusions and perspectives.

II. THE REGULAR RAY REPRESENTATIONS

We begin our work by considering the regular *ray* representations of the Newtonian group of the oscillator, since this self-contained mathematical construct offers an interesting background for quantum kinematics (cf. paper I). As we have seen, the harmonic oscillator's Newtonian group G is isomorphic to the semidirect product SO$(2) \circledS T(2)$, where $T(2)$ denotes the group of translations in the plane. Hence G is connected (but not simply connected). In this paper we are only exploring the physical possibilities of quantum kinematics, and therefore it seems advisable to start considering the universal covering group $\widetilde{G}$ of G, instead of G itself; i.e., $\widetilde{G} \simeq R_+ \otimes T(2)$, where R_+ denotes the multiplicative group of positive real numbers. One reason for this choice is the fact that $\widetilde{G}$ is connected *and* simply connected. The parametrization of the Newtonian group $\widetilde{G}$ is simply given by three (orthogonal) real axes; i.e.,

$$-\infty < q^a < +\infty, \quad a = 1,2,3. \tag{2.1}$$

These define the group manifold $M(\widetilde{G})$. The identity element of $\widetilde{G}$ corresponds to the origin, $e = (0,0,0)$, and the binary combination laws of the parameters (namely, the group law) can be written as follows:

$$\begin{aligned} q''^{\,1} &= g^1(q';q) = q'^{\,1} + q^1, \\ q''^{\,2} &= g^2(q';q) = q'^{\,2} \cos wq^1 + q'^{\,3} \sin wq^1 + q^2, \\ q''^{\,3} &= g^3(q';q) = -q'^{\,2} \sin wq^1 + q'^{\,3} \cos wq^1 + q^3. \end{aligned} \tag{2.2}$$

Hence, the group-inversion formulas for the parameters are

$$\begin{aligned} \bar{q}^1 &= -q^1, \\ \bar{q}^2 &= -q^2 \cos wq^1 + q^3 \sin wq^1, \\ \bar{q}^3 &= -q^2 \sin wq^1 - q^3 \cos wq^1. \end{aligned} \tag{2.3}$$

Next, we need to introduce the following right- and left-transport matrices for contravariant vectors in $M(\widetilde{G})$ (cf. paper I):

$$R_a^{\,b}(q) = \lim_{q' \to e} \partial'_a g^b(q';q), \tag{2.4}$$

$$L_a^{\,b}(q) = \lim_{q' \to e} \partial'_a g^b(q;q'), \tag{2.5}$$

respectively. Thus, using Eqs. (2.2), we obtain

$$R_a^{\,b}(q) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos wq^1 & -\sin wq^1 \\ 0 & \sin wq^1 & \cos wq^1 \end{bmatrix}, \tag{2.6}$$

$$L_a^{\,b}(q) = \begin{bmatrix} 1 & wq^1 & -wq^1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.7}$$

where a labels the rows and b labels the columns. Therefore, for the right infinitesimal operators on $M(\widetilde{G})$ [i.e., $X_a(q) = R_a^{\,b}(q)\partial_b$], one has

$$\begin{aligned} X_1 &= \partial_1, \\ X_2 &= \cos wq^1\, \partial_2 - \sin wq^1\, \partial_3, \\ X_3 &= \sin wq^1\, \partial_2 + \cos wq^1\, \partial_3, \end{aligned} \tag{2.8}$$

wherefrom the well known Lie algebra of E_2 follows:

$$\begin{aligned} [X_1, X_2] &= -wX_3, \\ [X_1, X_3] &= wX_2, \\ [X_2, X_3] &= 0. \end{aligned} \tag{2.9}$$

Since the determinants of (2.6) and (2.7) are $L(q) = R(q) = 1$, the Hurwitz invariant measure on $M(\widetilde{G})$ is given simply by $d\mu(q) = \mu_0\, dq^1\, dq^2\, dq^3$, where μ_0 is an arbitrary normalization constant.[15] In consequence, $\widetilde{G}$ is unimodular and the Hilbert space $\mathscr{H}(\widetilde{G})$ that carries the regular representation is defined as the set $\mathscr{L}^2(\widetilde{G})$ of square-integrable wavefunctions $\psi(q^1,q^2,q^3)$ on $M(\widetilde{G})$:

$$\langle \psi | \psi \rangle = \mu_0 \int\!\!\int\!\!\int dq^1\, dq^2\, dq^3 |\psi(q^1,q^2,q^3)|^2 < \infty. \tag{2.10}$$

It is also useful to consider the rigged Hilbert space structure attached with $\mathscr{H}(\widetilde{G})$, for this allows us to introduce an orthogonal complete continuous basis $\{|q^1q^2q^3\rangle\}$ on $\mathscr{H}(\widetilde{G})$. This basis is such that (paper I)

$$\langle q'^{\,1} q'^{\,2} q'^{\,2} | q^1 q^2 q^3 \rangle = \mu_0^{-1} \delta(q'^{\,1} - q^1)\delta(q'^{\,2} - q^2)\delta(q'^{\,3} - q^3), \tag{2.11}$$

$$\mu_0 \int\!\!\int\!\!\int dq^1\, dq^2\, dq^3 |q^1q^2q^3\rangle \langle q^1q^2q^3| = I, \tag{2.12}$$

where I denotes the identity operator in $\mathscr{H}(\widetilde{G})$.

Now we are ready to consider the unitary operators that carry a *ray* representation of $(\widetilde{G})$ within $\mathscr{H}(\widetilde{G})$. Let $\phi_k(q';q)$ be an *exponent* (or two-cocycle) of $(\widetilde{G})$; namely, ϕ_k is a continuous real-valued function of q' and q, which satisfies the well known three-point functional relation,[16]

$$\phi_k(q';q) + \phi_k[q'';g(q';q)] = \phi_k(q'';q') + \phi_k[g(q'';q');q], \tag{2.13}$$

as well as left and right homogeneous "initial" conditions at the identity

$$\phi_k(q;e) = \phi_k(e;q) \equiv 0. \tag{2.14}$$

Since $\widetilde{G}$ is connected and simply connected, we may assume without loss of generality that ϕ_k is globally well defined and differentiable on the whole space $M(\widetilde{G}) \times M(\widetilde{G})$. (See Bargmann's paper[16] for details.) Then, associated with each admissible two-cocycle ϕ_k of $(\widetilde{G})$, we introduce a set of linear operators $U_k(q^1,q^2,q^3)$ defined on $M(\widetilde{G})$, and such that

$$\begin{aligned} U_k(q'^{\,1},q'^{\,2},q'^{\,3})|q^1,q^2,q^3\rangle &= \exp[i\phi_k(q'^{\,1},q'^{\,2},q'^{\,3};q^1,q^2,q^3)] \\ &\quad \times |g^1(q';q), g^2(q';q), g^3(q';q)\rangle, \end{aligned} \tag{2.15}$$

$$U_k(q^1,q^2,q^3)|0,0,0\rangle = |q^1,q^2,q^3\rangle. \tag{2.16}$$

Since the regular basis $\{|q\rangle\}$ is complete, these equations define these operators on $\mathscr{H}(\widetilde{G})$ indeed. It is immediate that from these equations one gets

$$\begin{aligned} U_k(q'^{\,1},q'^{\,2},q'^{\,3})U_k(q^1,q^2,q^3) &= [\exp \phi_k(q'^{\,1},q'^{\,2},q'^{\,3};q^1,q^2,q^3)] \\ &\quad \times U_k[g^1(q^1;q), g^2(q';q), g^3(q';q)], \end{aligned} \tag{2.17}$$

as required for a ray representation. Moreover, it can be shown that these operators are *unitary*; in effect, one has

$$U_k^+(q) \equiv U_k^{-1}(q) = e^{-i\phi_k(q;\bar{q})} U_k(\bar{q}). \tag{2.18}$$

However, we observe that the unitarity of these operators is not "perfect" in general, unless $\phi_k(q;\bar{q}) \equiv 0$. Also, the following *consistency requirement* can be proved rather easily:

$$U_k(q) \int d\mu(q')|q'\rangle\langle q'| = \int d\mu(q')|q'\rangle\langle q'|U_k(q), \tag{2.19}$$

as a consequence of the fundamental properties of a two-cocycle. Clearly the matrix elements of the $U_k(q)$'s with respect to the regular basis $\{|q\rangle\}$ are given by

$$U_{q'q''}^{(k)}(q) = \langle q'|U_k(q)|q''\rangle$$
$$= \mu_0^{-1} e^{i\phi_k(q;q'')} \delta[q'^{\,1} - (q^1 + q''^{\,1})] \quad (2.20)$$
$$\times \delta[q'^{\,2} - g^2(q;q'')]\delta[q'^{\,3} - g^3(q;q'')],$$

and therefore one has

$$\int d\mu(q'') U_{qq''}^{(k)}(q_1) U_{q''q'}^{(k)}(q_2) = e^{i\phi_k(q_1;q_2)} U_{qq'}^{(k)}[g(q_1;q_2)], \quad (2.21)$$

for any two given points q_1 and q_2 in $M(\widetilde{G})$, as it should be.

Since the $U_k(q)$ are unitary operators in $\mathscr{H}(\widetilde{G})$, we observe that one and the same Hilbert space (which carries the regular *vector* representation of $\widetilde{G}$) also carries *all* the ray extensions of this representation. In effect, for any given vector $|\psi\rangle \in \mathscr{H}(\widetilde{G})$ one has

$$|\psi\rangle = \int d\mu(q)\psi(q)|q\rangle, \quad (2.22)$$

where $\psi(q) = \langle q|\psi\rangle$ is a wave function defined on $M(\widetilde{G})$, provided it belongs to $\mathscr{L}^2(\widetilde{G})$. Therefore, if one considers the image vector produced by one of the operators $U_k(q)$ acting on $|\psi\rangle$, say,

$$U_k(q)|\psi\rangle = |\psi_q^{(k)}\rangle = \int d\mu(q')\psi_q^{(k)}(q')|q\rangle, \quad (2.23)$$

this yields, immediately,

$$\int d\mu(q')|\psi_q^{(k)}(q')|^2 = \int d\mu(q')|\psi(q')|^2, \quad (2.24)$$

where $\psi_q^{(k)}(q')$ is given by

$$\psi_q^{(k)}(q') = e^{i\phi_k[q;g(\bar{q};q')]}\psi[g(\bar{q};q')]. \quad (2.25)$$

Thus $\psi_q^{(k)}(q') \in \mathscr{L}^2(\widetilde{G})$ if, and only if, $\psi(q') \in \mathscr{L}^2(\widetilde{G})$. Equation (2.25) states the unitary projective transformation law for the wave function belonging to $\mathscr{L}^2(\widetilde{G})$. By the way, this formula shows neatly that we are handling the ray extensions of the left regular "true" representation of $\widetilde{G}$, which reads

$$\psi_q^{(0)}(q') = \psi[g(\bar{q};q')], \quad (2.26)$$

according to the usual approach to this subject.[17] This fact is important for the purposes of quantum kinematics, because one does not need more geometric structure than $\mathscr{H}(\widetilde{G})$ itself in order to handle *all* the regular ray representations of $\widetilde{G}$.

The method for calculating an admissible (local) exponent of a given Lie group is well known.[11,12,16] This is a rather simple subject, which is becoming fashionable in several areas of theoretical physics,[18] and which belongs to the context of the cohomology theory of Lie groups.[19] In this paper we shall use the following two-cocycle of the Euclidean group in the plane:

$$\phi_k(q'^{\,1},q'^{\,2},q'^{\,3};q^1,q^2,q^3)$$
$$= (k/4)[(q'^{\,2})^2 + (q'^{\,3})^2]$$
$$\times [\tan w(q'^{\,1} + q^1) - \tan wq'^{\,1}]$$
$$+ (k/4)[(q^2)^2 + (q^3)^2]$$
$$\times [\tan w(q'^{\,1} + q^1) - \tan wq^1]$$
$$+ (k/2)\,(q'^{\,2}q^2 + q'^{\,3}q^3)$$
$$\times [\cos wq^1 \tan w(q'^{\,1} + q^1) - \sin wq^1]$$
$$+ (k/2)(q'^{\,3}q^2 - q'^{\,2}q^3)$$
$$\times [\sin wq^1 \tan w(q'^{\,1} + q^1) + \cos wq^1], \quad (2.27)$$

where k is a real *ray constant* corresponding to the two-cocycle of the extended Lie algebra. This exponent ϕ_k was calculated by means of non-Abelian analytical techniques, which are a coordinate-dependent realization of the coordinate-independent (i.e., geometric) techniques presented in Houard's paper.[12] This function ϕ_k is an admissible two-cocycle of $\widetilde{G}$, for it is nonequivalent to zero and satisfies the fundamental relation (2.13). Moreover by means of Eqs. (2.3), it can be shown that (2.27) implies

$$\phi_k(q;\bar{q}) = \phi_k(\bar{q};q) \equiv 0. \quad (2.28)$$

Therefore the unitary operators $U_k(q)$ of the ray representation associated with this two-cocycle satisfy

$$U_k^+(q) \equiv U_k^{-1}(q) \equiv U_k(\bar{q}), \quad (2.29)$$

instead of Eq. (2.18). We shall refer to this fact by saying that these operators are *perfectly unitary* and that the corresponding exponent $\phi_k(q';q)$ belongs to the *perfect unitary gauge*. Another important property of the exponent function (2.27) is that it is a *completely gauge-reduced* two-cocycle, in the sense that no term of the form $\gamma(q') + \gamma(q) - \gamma[g(q';q)]$ [with $\gamma(0,0,0) = 0$] appears in this function. Once we keep the ray constant k fixed, these two properties [namely, (a) to belong in the perfect unitary gauge, and (b) to be a completely gauge-reduced exponent] make the two-cocycle (2.27) of $\widetilde{G}$ *unique*. On the other hand, both conditions [(a) and (b)] recommend themselves, since both are physically reasonable. These mathematical minutiae are very important for quantum kinematics and will be taken up elsewhere.[20]

Finally, it can be seen that, with ϕ_k as given in Eq. (2.27), one has the following commutation rule:

$$U_k(q^1,\mathbf{q}) = \exp[-i(k/4)|\mathbf{q}|^2 \tan wq^1]\, U_k(q^1,0)\, U_k(0,\mathbf{q})$$
$$= \exp[-i(k/4)|\mathbf{q}|^2 \tan wq^1]$$
$$\times U_k[0,\mathbf{R}(wq^1)\cdot\mathbf{q}]\, U_k(q^1,0), \quad (2.30)$$

where, clearly, $\mathbf{q} = (q^2,q^3)$ is a vector in the plane, and where the rotation matrix $\mathbf{R}(wq^1)$ is given in Eq. (1.3). So much for the regular ray representations of $\widetilde{G}$.

III. KINEMATIC ALGEBRA OF THE HARMONIC OSCILLATOR

Now we come to the core of the issue, because the kinematic algebra of $\widetilde{G}$ determines the quantum rules of the system and, therefore, is of paramount importance in the kinematic approach to quantum dynamics.

First, let us consider the infinitesimal operators $P_a^{(k)}$, $a = 1,2,3$, which are the generators of the regular ray representation. Thus we get

$$U_k(\delta q) = I - (i/\hbar)\delta q^a P_a^{(k)}. \quad (3.1)$$

Hence, the infinitesimal transformation corresponding to Eqs. (2.23) and (2.25) yields the general formula

$$P_a^{(k)}|\psi\rangle = -i\hbar\int d\mu(q)\{[X_a(q) - ir_a^{(k)}(q)]\psi(q)\}|q\rangle, \tag{3.2}$$

where the functions $r_a^{(k)}(q)$ are given by

$$r_a^{(k)}(q) = \lim_{q'\to e}\partial'_a\phi_k(q';q) \tag{3.3}$$

[i.e., $r_a^{(k)}(e) = 0$]. Thus, from Eq. (2.27), in our case we get

$$\begin{aligned} r_1^{(k)} &= (k/4)w[(q^2)^2 + (q^3)^2]\sec^2 wq^1, \\ r_2^{(k)} &= -(k/2)q^3 \sec wq^1. \\ r_3^{(k)} &= (k/2)q^2 \sec wq^1. \end{aligned} \tag{3.4}$$

(We shall refer to these functions as the *exponent generators* of the ray representation.) So we have the following mapping:

$$\begin{aligned} P_1^{(k)}|\psi\rangle &\to -i\hbar\{\partial_1 - i(kw/4)q^2\sec^2 wq^1\}\psi(q^1,\mathbf{q}), \\ P_2^{(k)}|\psi\rangle &\to -i\hbar\{\cos wq^1\,\partial_2 - \sin wq^1\partial_3 \\ &\qquad + i(k/2)q^3\sec wq^1\}\psi(q^1,\mathbf{q}), \\ P_3^{(k)}|\psi\rangle &\to -i\hbar\{\sin wq^1\,\partial_2 + \cos wq^1\partial_3 \\ &\qquad - i(k/2)q^2\sec wq^1\}\psi(q^1,\mathbf{q}), \end{aligned} \tag{3.5}$$

which casts the "generalized momentum operators" (cf. paper I) into useful expressions for explicit wave-mechanical calculations on $M(\tilde{G})$. Of course, for that matter, one may also write

$$P_a^{(k)}|q\rangle = i\hbar[X_a(q) + ir_a^{(k)}(q)]|q\rangle, \tag{3.6}$$

which one obtains directly from Eq. (2.15).

The extended Lie algebras associated with ray representations of Lie groups are well known.[16] From Eq. (3.6), after some manipulations, we obtain

$$\begin{aligned} [P_1^{(k)},P_2^{(k)}] &= i\hbar w P_3^{(k)}, \\ [P_1^{(k)},P_3^{(k)}] &= -i\hbar w P_2^{(k)}, \\ [P_2^{(k)},P_3^{(k)}] &= i\hbar^2 k. \end{aligned} \tag{3.7}$$

These are the familiar formulas for the extended Lie algebra in the regular representation of the Newtonian covering group of the harmonic oscillator.[13]

Now we turn to the generalized position operators introduced in paper I; namely, we define Hermitian operators Q^a by means of their spectral integrals,

$$Q^a = \int d\mu_L(q)|q\rangle q^a\langle q|, \quad a = 1,2,3. \tag{3.8}$$

These operators are such that $Q^a|q\rangle = q^a|q\rangle$ and $[Q^a,Q^b] = 0$ hold. Taking into account the general properties of an exponent, it can be shown that these Q^a's transform according to the following law under the elements of the group:

$$U_k^+(q)Q^aU_k(q) = g^a(q;Q) = \int d\mu_L(q')|q'\rangle g^a(q;q')\langle q'|. \tag{3.9}$$

This result is valid, in general, within the left regular *ray* representations of a Lie group. Therefore, if one considers the infinitesimal version of this unitary transformation, one obtains the same *non-Abelian canonical commutators* which were already presented in paper I. Indeed, one has generalized commutation relations for non-Abelian variables; i.e.,

$$[Q^a,P_b^{(k)}] = i\hbar R_b^a(Q), \tag{3.10}$$

where the operator $R_b^a(Q)$ has the spectral representation

$$R_b^a(Q) = \int d\mu_L(q)|q\rangle R_b^a(q)\langle q|. \tag{3.11}$$

Accordingly, using Eq. (2.6), we obtain the following set of non-Abelian canonical commutators for the simple harmonic oscillator:

$$\begin{aligned} &[Q^1,P_1^{(k)}] = i\hbar, \quad [Q^2,P_1^{(l)}] = 0, \\ &[Q^1,P_2^{(k)}] = 0, \quad [Q^2,P_2^{(k)}] = i\hbar\cos wQ^1, \\ &[Q^1,P_3^{(k)}] = 0, \quad [Q^2,P_3^{(k)}] = i\hbar\sin wQ^1, \\ &\qquad [Q^3,P_1^{(k)}] = 0, \\ &\qquad [Q^3,P_2^{(k)}] = -i\hbar\sin wQ^1, \\ &\qquad [Q^3,P_3^{(k)}] = i\hbar\sin wQ^1. \end{aligned} \tag{3.12}$$

The set of commutators presented in Eqs. (3.7) and (3.12) constitute the *extended kinematic algebra* of $\tilde{G}$. We next derive some interesting consequences of these commutation relations.

First (and most importantly), from the Lie algebra (3.7) we immediately obtain the fundamental result

$$[a_k,a_k^\dagger] = I, \tag{3.13}$$

where the "ladder" operators are defined by

$$\begin{aligned} a_k &= (1/\hbar\sqrt{2k})\,(P_2^{(k)} + iP_3^{(k)}), \\ a_k^\dagger &= (1/\hbar\sqrt{2k})\,(P_2^{(k)} - iP_3^{(k)}). \end{aligned} \tag{3.14}$$

Bear in mind that these operators are *not* defined within the regular *vector* representation of $\tilde{G}$, since in that representation one sets $k = 0$. We also wish to remark that the assumption $k > 0$, which is implicitly made in Eq. (3.14), means no loss of generality. Furthermore, if in the present context one looks for a Casimir operator of the extended Lie algebra (3.7), one easily finds that the particular combination of generators

$$S_k = P_1^{(k)} + \tfrac{1}{2}\hbar w(a_k a_k^\dagger + a_k^\dagger a_k) \tag{3.15}$$

is indeed an invariant of the algebra; i.e.,

$$[S_k,P_a^{(k)}] = 0, \quad a = 1,2,3. \tag{3.16}$$

Therefore, S_k is an invariant operator of the ray representation

$$U_k^\dagger(q)S_kU_k(q) = S_k. \tag{3.17}$$

Henceforth, for obvious reasons, this invariant operator S_k will be referred to as the *Schrödinger operator* of the system.

A novel feature of this formalism is the fact that two different Hamiltonians come to the fore. In fact, the operator

$$H_k = \tfrac{1}{2}\hbar w(a_k a_k^\dagger + a_k^\dagger a_k) \tag{3.18}$$

is the familiar Hamiltonian of the harmonic oscillator (we shall call it the *dynamical Hamiltonian* of the system). On the other hand, the operator $P_1^{(k)}$ is the generator of time translation symmetry (hereafter, we shall call it the *kinematic Hamiltonian* of the system). However, it is easy to see that these two Hamiltonians are linearly independent operators. To this end, it is enough to look at the kinematic algebra obeyed by the dynamical Hamiltonian, which is as follows:

$$[P_1^{(k)},H_k]=0,$$
$$[P_2^{(k)},H_k]=i\hbar wP_3^{(k)},\tag{3.19}$$
$$[P_3^{(k)},H_k]=-i\hbar wP_2^{(k)},$$

and *also*

$$[Q^1,H_k]=0,$$
$$[Q^2,H_k]=i(w/k)\ (P_2^{(k)}\cos wQ^1+P_3^{(k)}\sin wQ^1),$$
$$[Q^3,H_k]=-i(w/k)\ (P_2^{(k)}\sin wQ^1-P_3^{(k)}\cos wQ^1).\tag{3.20}$$

[Suppose that we had considered exclusively the extended Lie algebra (3.7). Then, after comparing Eq. (3.19) with (3.7), we could think of the possibility of having a *null* Schrödinger operator, $S_k\equiv 0$, that is, $H_k=-P_1^{(k)}$. However, a glance at the canonical commutators (3.20) and (3.12) reveals immediately that such a "conclusion" is false.] Furthermore, these two Hamiltonians commute, and therefore the quantum model (one obtains by "quantizing" directly the Newtonian symmetries of the system) consists of *two noninteracting parts*. The Schrödinger operator thus appears as the *total Hamiltonian* of this composed system. (We shall return to this issue presently.)

The following commutation relations are immediate and useful:

$$[e^{\pm iwQ^1},P_1^{(k)}]=\mp\hbar we^{\pm iwQ^1},\tag{3.21}$$
$$[Q^1,H_k]=0,\tag{3.22}$$

i.e.,

$$[e^{\pm iwQ},S_k]=\mp\hbar we^{\pm iwQ^1}.\tag{3.23}$$

Since S_k is a Casimir operator,

$$[a_k,S_k]=[a_k^\dagger,S_k]=0,\tag{3.24}$$

and since for the dynamical Hamiltonian one has

$$[a_k,H_k]=\hbar wa_k,\quad [a_k^\dagger,H_k]=-\hbar wa_k^\dagger,\tag{3.25}$$

it follows that one also has "opposed" ladder effects for the kinematic Hamiltonian:

$$[a_k,P_1^{(k)}]=-\hbar wa_k,\quad [a_k^\dagger,P_1^{(k)}]=\hbar wa_k^\dagger;\tag{3.26}$$

i.e., the "annihilation" operator of H_k is a "creation" operator of $P_1^{(k)}$, and vice versa.

To end up this section, let us define two auxiliary operators

$$b_k=(k/2)^{1/2}Z^++ie^{iwQ^1}a_k^\dagger,$$
$$b_k^\dagger=(k/2)^{1/2}Z-ie^{-iwQ^1}a_k,\tag{3.27}$$

where $Z=Q^2+iQ^3$. After some manipulations, it can be shown that

$$[b_k,P_a^{(k)}]=[b_k^\dagger,P_a^{(k)}]=0,\quad a=1,2,3.\tag{3.28}$$

Hence, every Hermitian function of these operators is an observable constant of "motion" of the system and, furthermore, it also plays the role of an invariant operator of the group. In this paper we shall be particularly interested in the Hermitian combination

$$B_k=b_k^\dagger b_k,\tag{3.29}$$

because of the following striking result:

$$[b_k,b_k^\dagger]=I.\tag{3.30}$$

The set of commuting Hermitian operators $\{P_1^{(k)},H_k,B_k\}$ is complete, and may be used instead of $\{Q^1,Q^2,Q^3\}$ in order to characterize the three "degrees of freedom" of the Hilbert space $\mathscr{H}(\widetilde{G})$. Let us remark, however, that in this paper we shall not indulge on the possible physical significance of b_k and $b_k^\dagger$, which will play here a purely auxiliary role.[21]

IV. SUPERSELECTION RULES

With the aim of arriving at a reasonable physical interpretation of the model, let us examine the following postulate: *the allowable physical states of the harmonic oscillator correspond to simultaneous eigenkets of the invariant operators S_k and B_k* previously found. It is clear that the superselection rule of the Schrödinger operator S_k corresponds to the *law of conservation of total energy* of an isolated system that consists of two (noninteracting) parts. The physical meaning of the superselection rule of the operator B_k here plays an auxiliary role and remains to be discussed at another opportunity. Of course, with the aim of tackling the present endeavor, we need to solve the eigenvalue problems of S_k and B_k within the Hilbert space $\mathscr{H}(\widetilde{G})$.

Using the realization of $P_1^{(k)}$ in the Q representation, as stated in Eqs. (3.5), one easily solves the eigenvalue equation of the kinematic Hamiltonian. Indeed, one gets the eigenvectors

$$|E_1\rangle=\int d\mu(q)u_{E_1}(\mathbf{q})\exp\left\{i\left[\left(\frac{E_1}{\hbar}\right)q^1+\frac{k}{4}\mathbf{q}^2\tan wq^1\right]\right\}|q\rangle,\tag{4.1}$$

where $\mathbf{q}=(q^2,q^3)$, $\mathbf{q}^2=(q^2)^2+(q^3)^2$, and where the spectrum is given by $-\infty<E_1<+\infty$. The function $u_{E_1}(q^2,q^3)$, which figures in Eq. (4.1), corresponds to the degeneracy of these states. This function may depend on E_1, but not on q^1; otherwise it remains completely arbitrary, provided the scalar products

$$\langle E_1'|E_1\rangle=\delta(E_1'-E_1)\int\int dq^2\,dq^3\,u_{E_1}'^*(q^2,q^3)u_{E_1}(q^2,q^3)\tag{4.2}$$

are such that

$$(u_{E_1}'|u_{E_1})=\int\int dq^2\,dq^3\,u_{E_1}'^*(q^2,q^3)u_{E_1}(q^2,q^3)<\infty.\tag{4.3}$$

Here, and henceforth, we set $\mu_0=(2\pi h)^{-1}$. [It is clear that, *sensu stricto*, one should write $|E_1,[u_{E_1}]\rangle$ to denote the eigenkets defined in Eq. (4.1)].

The eigenvalue problem of H_k is well known, of course, One has

$$H_k|n_2\rangle=(n_2+\tfrac{1}{2})\hbar\omega|n_2\rangle,\quad n_2=0,1,2,...,\tag{4.4}$$

with

$$|n_2\rangle=(n_2!)^{-1/2}(a_k^\dagger)^{n_2}|0\rangle\tag{4.5}$$

(clearly, $|0\rangle=|n_2=0\rangle$), and

$$a_k|0\rangle=0.\tag{4.6}$$

Hence, all that remains to be done is to obtain these eigenkets as vectors in $\mathscr{H}(\widetilde{G})$. To this end, we cast the ladder operators in the Q representation; i.e., we set [cf. Eq. (3.5) and (3.14)]

$$a_k|\psi\rangle \to -i(2/k)^{1/2} \times \exp\{iwq^1\}[\partial_{z*}+(k/4)z(1-i\tan wq^1)]\psi(q^1,\mathbf{q}),$$

$$a_k^\dagger|\psi\rangle \to -i(2/k)^{1/2} \times \exp\{iwq^1\}[\partial_z-(k/4)z^*(1+i\tan wq^1)]\psi(q^1,\mathbf{q}), \tag{4.7}$$

where we have written $z = q^2 + iq^3$, $2\partial_z = \partial_2 - i\partial_3$, etc. So one obtains the ground state

$$|n_2=0\rangle = \int d\mu(q) v_0(q^1,z) \times\exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle, \tag{4.8}$$

wherefrom the general form of the eigenvectors follows:

$$|n_2\rangle = \left(-i\sqrt{\frac{2}{k}}\right)^{n_2}(n_2!)^{-1/2}\int d\mu(q) \times\exp(-in_2wq^1)\left[\left(\partial_z-\frac{k}{2}z^*\right)^{n_2}v_0(q^1,z)\right] \times\exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle. \tag{4.9}$$

Again, the function $v_0(q^1,z)$ describes the degeneracy of these eigenkets; it does *not* depend on n_2 (neither does it depend on $z^* = q^2 - iq^3$). Otherwise, $v_0(q^1,z)$ is a completely arbitrary analytic function of z, provided $\langle 0|0\rangle$ is finite. Quite generally, one has the following scalar product:

$$\langle n_2'|n_2\rangle = \delta_{n_2'n_2}(2\pi h)^{-1}\int_{-\infty}^{\infty}dq^1\iint d^2z \times\exp-(k/2)|z|^2[v_0'(q^1,z)]^*v_0(q^1,z). \tag{4.10}$$

Now, a glance at Eqs. (4.1) and (4.9) yields immediately the general form of the simultaneous eigenkets of the two commuting Hamiltonians, $P_1^{(k)}$ and H_k; namely, we obtain eigenkets of the form

$$|E_1n_2\rangle = \left(-i\sqrt{\frac{2}{k}}\right)^{n_2}(n_2!)^{-1/2}\int d\mu(q) \times\exp\left(\frac{i}{\hbar}(E_1-n_2\hbar w)q^1\right) \times\left[\left(\partial_z-\frac{k}{2}z^*\right)^{n_2}v_{E_1}(z)\right] \times\exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle, \tag{4.11}$$

such that

$$P_1^{(k)}|E_1n_2\rangle = (E_1-n_2\hbar w)|E_1n_2\rangle, \tag{4.12}$$

$$H_k|E_1n_2\rangle = (n_2+\tfrac{1}{2})\hbar w|E_1n_2\rangle. \tag{4.13}$$

In this manner, one has solved the eigenvalue problem of the Schrödinger operator; indeed one has

$$S_k|E_1n_2\rangle = (E_1+\tfrac{1}{2}\hbar w)|E_1n_2\rangle. \tag{4.14}$$

Moreover, the following orthogonality relation holds:

$$\langle E_1'n_2'|E_1n_2\rangle = \delta(E_1'-E_1)\delta_{n_2'n_2}\iint d^2z \times\exp(-(k/2)|z|^2)[v_{E_1}'(z)]^*v_{E_1}(z). \tag{4.15}$$

Thus, we are in a position to formulate our *first superselection rule*: for a fixed value of the ray constant k (which we take as $k>0$, without loss of generality), the allowable physical states of the system are described by kets of $\mathscr{H}(\widetilde{G})$, which are given by superpositions of the form

$$|\psi;E_1\rangle = \sum_{n_2=0}^{\infty}c_{n_2}(\psi)|E_1n_2\rangle, \tag{4.16}$$

so that

$$S_k|\psi;E_1\rangle = (E_1+\tfrac{1}{2}\hbar w)|\psi;E_1\rangle. \tag{4.17}$$

Clearly, the basic eigenkets $|E_1n_2\rangle$ defined in Eq. (4.11) depend functionally on the wave function $v_{E_1}(z)$ one uses for building a ground state vector of H_k. (More rigorously, one should write the functional kets $|E_1n_2,[v_{E_1}]\rangle$ to denote these eigenkets.) Notwithstanding this degeneracy, the following orthogonality relation holds quite generally, as a consequence of (4.15):

$$\langle\psi';E_2'|\psi;E_1\rangle = \delta(E_1'-E_1)\sum_{n_2}c_{n_2}^*(\psi')c_{n_2}(\psi)\iint d^2z \times\exp(-(k/2)|z|^2)[v_{E_1}'(z)]^*v_{E_1}(z). \tag{4.18}$$

The Hilbert subspace $\mathscr{H}_{E_1}\subset\mathscr{H}(\widetilde{G})$, defined by means of the superposition (4.16), has *two degrees of freedom* corresponding to the coefficients $c_{n_2}(\psi)$ and to the arbitrary wave function $v_{E_1}(z)$, which is implicit in $|E_1n_2\rangle$. From Eq. (4.18) we see that the probability amplitude for a transition from a state belonging to the subspace $\mathscr{H}_{E_1'}$ to a state belonging to $\mathscr{H}_{E_1}$ is zero when $E_1'\neq E_2$, as it must be. Of course, one has to manage the admissible transition between physical states by means of the well known mathematical refinements called on by a continuous spectrum. Nevertheless, it is clear that this complication does not preclude the interpretation of the superselection rule introduced by S_k as the law of conservation of total energy.

Next, let us discuss the second superselection rule. In order to obtain a new basis in $\mathscr{H}(\widetilde{G})$, to be used instead of $\{|q^1q^2q^3\rangle\}$, we shall consider the auxiliary Casimir operator B_k defined in Eqs. (3.27) and (3.29). [As we have already remarked, the set $\{P_1^{(k)},H_k,B_k\}$ is a complete set of commuting self-adjoint operators in $\mathscr{H}(\widetilde{G})$.] According to Eq. (3.30), the eigenvalue problem of B_k has the general solution

$$B_k|n_3\rangle = n_3|n_3\rangle,\quad n_3=0,1,2,\ldots. \tag{4.19}$$

The Q representation of the auxiliary ladder operators, b_k and $b_k^\dagger$, yields

$$b_k|\psi\rangle\to(k/2)^{1/2}[\partial_z+(k/4)z^*(1-i\tan wq^1)]\psi(q^1,\mathbf{q}),$$

$$b_k^\dagger|\psi\rangle\to(k/2)^{1/2}[\partial_{z*}-(k/4)z(1+i\tan wq^1)]\psi(q^1,\mathbf{q}). \tag{4.20}$$

[As the reader can appreciate, these formulas make an interesting contrast with the corresponding formulas (4.7) for a_k and $a_k^\dagger$.] Hence, the ground state $|n_3=0\rangle$ of B_k turns out to be of the form

$$|n_3=0\rangle = \int d\mu(q) w_0(q^1,z^*) \times \exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle, \quad (4.21)$$

where $w_0(q^1,z^*)$ is an arbitrary normalizable function [cf. also, Eq. (4.8)]. Thereafter, using the well known process analogous to (4.5), one obtains the eigenvectors of B_k; namely,

$$|n_3\rangle = \left(-\sqrt{\frac{2}{k}}\right)^{n_3}(n_3)^{-1/2}\int d\mu(q) \times \left[\left(\partial_{z^*}-\frac{k}{2}z\right)^{n_3} w_0(q^1,z^*)\right] \times \exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle. \quad (4.22)$$

Now, in order to obtain the desired common eigenvectors $|E_1n_2n_3\rangle$, let us equalize Eqs. (4.11) and (4.22), recalling that $v_E(z)$ does not depend on n_2, but may depend on n_3, and also that $w_0(q^1,z^*)$ does not depend on n_3, but may depend on (E_1,n_2). Thus we demand

$$\begin{aligned}(-i\sqrt{2/k})^{n_2}(n_2!)^{-1/2}\exp[(i/\hbar)(E_1-n_2hw)q^1] \\ \times[\partial_z-(k/2)z^*]^{n_2}v_{E_1n_3}(z) \\ =(-\sqrt{2/k})^{n_3}(n_3!)^{-1/2} \\ \times[\partial_{z^*}-(k/2)z]^{n_3}w_{E_1n_2}(q^1,z^*).\end{aligned} \quad (4.23)$$

The detailed analysis of this condition is rather lengthy and yields the final answer

$$\begin{aligned}|E_1n_2n_3\rangle \\ = i^{n_2}\left(\sqrt{\frac{k}{2}}\right)^{n_2+n_3+1}(\pi n_2!\, n_3!)^{-1/2}\int d\mu(q) \\ \times\exp\left(\frac{i}{\hbar}(E_1-n_2\hbar w)q^1\right)z^{*n_2}z^{n_3}F_{\{n_2,n_3\}}\left(-\frac{k}{2}|z|^2\right) \\ \times\exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle,\end{aligned} \quad (4.24)$$

where we have defined the function

$$F_{\{n_2,n_3\}}(x) = \sum_{m=0}^{\{n_2,n_3\}_<}\frac{n_2!\,n_3!}{m!(n_2-m)!(n_3-m)!}x^{-m}, \quad (4.25)$$

with $\{n_2,n_3\}_<$ denoting the smallest number in $\{n_2,n_3\}$. These eigenkets contain no degeneracy, satisfy the orthogonality relations,

$$\langle E_1'n_2'n_3'|E_1n_2n_3\rangle = \delta(E_1'-E_1)\delta_{n_2'n_2}\delta_{n_3'n_3}, \quad (4.26)$$

and the completeness condition on $\mathscr{H}(\widetilde{G})$,

$$\int_{\infty}^{\infty}dE_1\sum_{n_1=0}^{\infty}\sum_{n_3=0}^{\infty}|E_1n_2n_3\rangle\langle E_1n_2n_3| = I. \quad (4.27)$$

[Equation (4.26) can be proved by means of the following lemmas:

$$\iint d^2z\exp[-(k/2)|z|^2] \times\left(\partial_{z^*}-\frac{k}{2}z\right)f(z,z^*)=0, \quad (4.28)$$

$$\begin{aligned}z^{*n_2}z^{n_3}F_{\{n_2,n_3\}}\left(-\frac{k}{2}|z|^2\right) \\ =\left(-\frac{2}{k}\right)^{n_2}\left(\partial_z-\frac{k}{2}z^*\right)^{n_2}z^{n_3} \\ =\left(-\frac{2}{k}\right)^{n_3}\left(\partial_{z^*}-\frac{k}{2}z\right)^{n_3}z^{*n_2}.\end{aligned} \quad (4.29)$$

Of course, the proof of the completeness relation is quite involved, and requires the use of the generating function of the Hermite polynomials. We omit these details for the sake of briefness.]

In summary, according to the assumed superselection rules, the allowable physical states of the harmonic oscillator correspond to kets in $\mathscr{H}(\widetilde{G})$ satisfying the following superposition principle:

$$|\psi;E_1n_3\rangle = \sum_{n_2=0}^{\infty}c_{n_2}(\psi)|E_1n_2n_3\rangle, \quad (4.30)$$

i.e.,

$$S_k|\psi;E_1n_3\rangle = (E_1+\tfrac{1}{2}\hbar\omega)|\psi;E_1n_3\rangle, \quad (4.31)$$

$$B_k|\psi;E_1n_3\rangle = n_3|\psi;E_1n_3\rangle. \quad (4.32)$$

The Hilbert subspaces $\mathscr{H}_{E_1n_3}\subset\mathscr{H}(\widetilde{G})$ containing these physical states have just *one* degree of freedom, which corresponds to the coefficient of the superposition (4.30); namely,

$$c_{n_2'}(\psi) = \int_{-\infty}^{\infty}dE_1'\sum_{n_3'=0}^{\infty}\langle E_1'n_2'n_3'|\psi;E_1n_3\rangle. \quad (4.33)$$

Of course, one has

$$\langle\psi';E_1'n_3'|\psi;E_1n_3\rangle = \delta(E_1'-E_1)\delta_{n_3'n_3}\sum_{n_2=0}^{\infty}c_{n_2}^*(\psi')c_{n_2}(\psi), \quad (4.34)$$

and $\mathscr{H}(\widetilde{G})$ becomes an incoherent Hilbert space.

V. COMPLEMENTARY RAY REPRESENTATIONS

As it follows from the previous discussion, the quantum model stemming from the regular ray representations of the Newtonian group $\widetilde{G}$ contains all the essential mathematical features of the familiar quantum theory of the harmonic oscillator. Yet, this model still looks far from the ordinary wave mechanical model of the system. In effect, what is missing in the present approach is a space-time description, which should be concordant with the "special relativity" theory of the harmonic oscillator already sketched in the Introduction. Here we shall attain such a description by means of some purely kinematic considerations.

To this end, let us consider a new kind of kets, labeled as $|t,x\rangle$ because they are in one-to-one correspondence with the events (t,x) of Newtonian space-time. These kets belong to the rigged Hilbert space structure associated with $\mathscr{H}(\widetilde{G})$ and should be such that, by construction, one has

$$U_k(q^1,q^2,q^3)|t,x\rangle = \exp[i\rho_k(q^1,q^2,q^3;t,x)] \times|t+q^1,x+q^2\cos wt+q^3\sin wt\rangle, \tag{5.1}$$

where ρ_k is a real-valued function. Accordingly, we shall say that these kets carry a *complementary ray representation* of the space-time symmetry realization of $\widetilde{G}$ stated in Eq. (1.1). Clearly, this property would entail the following transformation law for wave functions defined on the space-time arena:

$$\langle t,x|U_k^\dagger(q)|\psi\rangle = e^{-i\rho_k(q;t,x)}\psi(t+q^1,x+q^2\cos wt+q^3\sin wt), \tag{5.2}$$

where

$$\psi(t,x) = \langle t,x|\psi\rangle, \quad |\psi\rangle\in\mathscr{H}(\widetilde{G}). \tag{5.3}$$

It is evident that in order to solve the problem set by the existence of such kets (and, by the way, to learn how to construct them) one must look for necessary and sufficient conditions for having the formal property stated in Eq. (5.1). Hence, we assume (5.1) and examine first the required properties for having an allowable *phase function* $\rho_k(q;t,x)$. An immediate consequence of Eq. (5.1) is the requirement

$$\rho_k(0,0,0;t,x)\equiv 0. \tag{5.4}$$

One also easily gets the following functional relation for the phase function ρ_k and the exponent function ϕ_k:

$$\rho_k(q;t,x)+\rho_k(q';t+q^1,x+q^2\cos wt+q^3\sin wt) - \rho_k[g(q';q);t,x] = \phi_k(q';q), \tag{5.5}$$

where the $g(q';q)$'s are given in (2.2). These two properties are enough for obtaining a constructive method for calculating an admissible phase function $\rho_k(q;t,x)$.[22] As a matter of fact, since the infinitesimal operators, corresponding to the space-time realization (1.1) of $\widetilde{G}$, are given by

$$\begin{aligned} Z_1(t,x) &= \partial_t, \\ Z_2(t,x) &= \cos wt\,\partial_x, \\ Z_3(t,x) &= \sin wt\,\partial_x, \end{aligned} \tag{5.6}$$

we obtain the following solution [which is associated with ϕ_k as given in (2.27)]:

$$\begin{aligned} \rho_k(q^1,q^2,q^3;t,x) = &-(k/4)[(q^2)^2+(q^3)^2]\tan wq^1 \\ &-(k/4)[(q^2)^2-(q^3)^2]\sin 2wt \\ &-(k/2)q^2q^3\cos 2wt \\ &-kx(q^2\sin wt-q^3\cos wt). \end{aligned} \tag{5.7}$$

The reader can check this phase function against Eqs. (2.20) and (5.5). We shall also need the *phase generators* of the complementary ray representation, i.e.,

$$\begin{aligned} \rho_1^{(k)}(t,x) &= 0, \\ \rho_2^{(k)}(t,x) &= -kx\sin wt, \\ \rho_3^{(x)}(t,x) &= kx\cos wt, \end{aligned} \tag{5.8}$$

which are defined as follows:

$$\rho_a^{(k)}(t,x) = \lim_{q\to 0}\partial_a\rho_k(q;t,x), \quad a=1,2,3. \tag{5.9}$$

Now we set, *ex hypothesi*, $|t,x\rangle\in\mathscr{H}(\widetilde{G})$, i.e.,

$$|t,x\rangle = \int d\mu(q)\psi_k^*(t,x;q)|q\rangle, \tag{5.10}$$

where, clearly, $\psi_k(t,x;q) = \langle t,x|q\rangle$. (Of course, these kets $|t,x\rangle$ must depend on k.) Then, if one considers ψ_k at the identity, say,

$$\psi_k(t,x;0,0,0) = \langle t,x|0,0,0\rangle \equiv \xi_k(t,x), \tag{5.11}$$

where ξ_k is an *arbitrary* single-valued wave function defined on the space-time manifold, one can show that a *necessary and sufficient* condition for the kets $|t,x\rangle$ to be endowed with the property (5.1) is that they have the following general form:

$$|t,x\rangle = \int d\mu(q)\xi_k^*(t+\bar{q}^1,x+\bar{q}^2\cos wt+\bar{q}^3\sin wt) \times\exp[i\rho_k(\bar{q}^1,\bar{q}^2,\bar{q}^3;t,x)]|q\rangle. \tag{5.12}$$

Thus, since ξ_k remains at our disposal, we have enormous freedom for adjusting a complementary ray representation of $\widetilde{G}$ within $\mathscr{H}(\widetilde{G})$.

Moreover, once a suitable *generating wave function* $\xi_k(t,x)$ has been adopted, it is clear that there remains some gauge freedom for fixing the phase of the kets $|t,x\rangle$ themselves locally in space-time. These σ transformations are of the form

$$|t,x\rangle \to |t,x\rangle_\sigma = e^{i\sigma(t,x)}|t,x\rangle, \tag{5.13}$$

and are completely independent of the gauge transformations that perform the equivalence between ray representations inherent to the theory of two-cocycles.[16] As a consequence of (5.13) one has the following σ transformation of the phase generators:

$$\rho_a'^{(k)}(t,x) = \rho_a^{(k)}(t,x) - Z_a(t,x)\sigma(t,x). \tag{5.14}$$

Once ξ_k is fixed, all the σ-gauge freedom of the formalism comes from this last equation. In other words, the solution presented in Eq. (5.8) is defined only within a σ transformation of the form (5.14). It must be borne in mind that the generating wave function $\xi_k(t,x)$ is not committed with these σ transformations. In effect, $\xi_k(t,x)$ must be determined independently, on some *physical* ground.

Finally, let us review some rather simple (albeit important) quantum kinematic features of the complementary ray representation carried by the space-time kets $|t,x\rangle$. First we observe that an infinitesimal transformation of these kets yields the following space-time realizations of the "generalized momentum" operators:

$$\begin{aligned} P_1^{(k)}|t,x\rangle &= i\hbar\,\partial_t|t,x\rangle, \\ P_2^{(k)}|t,x\rangle &= i\hbar(\cos wt\,\partial_x - ikx\cos wt)|t,x\rangle, \\ P_3^{(k)}|t,x\rangle &= i\hbar(\sin wt\,\partial_x + ikx\cos wt)|t,x\rangle; \end{aligned} \tag{5.15}$$

and therefore the ladder operators correspond to

$$\begin{aligned} a_k|t,x\rangle &= (i/\sqrt{2k})e^{iwt}(\partial_x - kx)|t,x\rangle, \\ a_k^\dagger|t,x\rangle &= (i/\sqrt{2k})e^{-iwt}(\partial_x + kx)|t,x\rangle. \end{aligned} \tag{5.16}$$

Hence, by means of the complementary space-time representation of $\widetilde{G}$, the dynamical Hamiltonian stated in Eq. (3.18) can be cast in the following form:

$$H_k|t,x\rangle = [-(\hbar w/2k)\partial_x^2 + \tfrac{1}{2}\hbar w k x^2]|t,x\rangle. \quad (5.17)$$

Therefore, according to the standard quantum theory of the oscillator, we recognize that the ray constant k must be interpreted as

$$k = mw/\hbar, \quad (5.18)$$

where m is the mass of the oscillator. In this fashion we arrive at the well known form of the Schrödinger operator; i.e., we get

$$S_k|t,x\rangle = [i\hbar\partial_t - (\hbar^2/2m)\partial_x^2 + \tfrac{1}{2}mw^2x^2]|t,x\rangle. \quad (5.19)$$

It must be borne in mind that one arrives at this result using exclusively the assumed Newtonian symmetries of the system.

Of course, had we used a different set of allowable phase generators $\rho_a^{(k)}(t,x)$ we would obtain a different Schrödinger operator in (5.19). This means, however, that only a local σ change of phase has been performed on $|t,x\rangle$. The solutions we have adopted in Eqs. (5.8) avoid this artificial complication from the beginning, because they correspond to a *completely gauge-reduced* set of phase generators. That is, according to Eq. (5.14), no terms of the form $Z_a(t,x)\sigma(t,x)$ appear in the solutions (5.8) and, moreover, any other conceivable set of solutions would differ from (5.8) merely by the presence of such spurious terms.

VI. THE GENERATING WAVE FUNCTION ξ_k AND THE SPACE-TIME KERNEL OF THE HARMONIC OSCILLATOR

Once a suitable phase function $\rho_k(q;t,x)$ has been found, it is rather natural to require that the space-time kets $|t,x\rangle$ themselves correspond to physically realizable states. Thus we set

$$S_k|t,x\rangle = (E_1 + \tfrac{1}{2}\hbar w)|t,x\rangle, \quad (6.1)$$

and

$$B_k|t,x\rangle = n_3|t,x\rangle; \quad (6.2)$$

i.e., in the sequel it should be understood that $|t,x\rangle$ stands for $|t,x;E_1,n_3\rangle$. Furthermore, for the sake of concreteness, since in this paper we are only interested in S_k, and not in B_k, here we consider exclusively the simplest case; namely, henceforth we set $n_3 = 0$, since this simplifies our calculations a great deal. The general situation, with $n_3 = 0,1,2,...$, will be considered elsewhere. Thus, all the physical states discussed hereafter belong to the Hilbert subspaces $\mathscr{H}_{E_1 n_3} \subset \mathscr{H}(\widetilde{G})$, with $n_3 = 0$.

According to Eq. (4.24), we are searching for a set of states $|t,x;E_1\rangle$ that satisfy simultaneously the following conditions:

$$\begin{aligned}|t,x;E_1\rangle &= \sum_{n_2=0}^{\infty} c_{n_2}^{(k)}(t,x)|E_1 n_2 0\rangle \\ &= \left(\frac{k}{2}\right)^{1/2}\int d\mu(q)\sum_{n_2}\frac{c_{n_2}^{(k)}(t,x)}{\sqrt{n_2!}} \\ &\quad\times\left(i\sqrt{\frac{k}{2}}\,z^*\exp(-iwq^1)\right)^{n_2}\exp\left(\frac{i}{\hbar}E_1q^1\right) \\ &\quad\times\exp\left(-\frac{k}{4}|z|^2(1-i\tan wq^1)\right)|q\rangle \\ &= \int d\mu(q)\xi_k^*[t-q^1;x-q^2\cos w(t-q^1) \\ &\quad - q^3\sin w(t-q^1)] \\ &\quad\times\exp\left(-i\frac{k}{4}\{[(q^2)^2-(q^3)^2]\sin 2w(t-q^1)\right. \\ &\quad\left.-2q^2q^3\cos 2w(t-q^1)\}\right)\exp\{ikx \\ &\quad\times[q^2\sin w(t-q^1)-q^3\cos w(t-q^1)]\} \\ &\quad\times\exp[i(k/4)|z|^2\tan wq^1]|q\rangle. \end{aligned} \quad (6.3)$$

Of course, the determination of $c_{n_2}^{(k)}(t,x)$ is trivial, since from Eq. (5.17) and the superselection rule (6.1), one gets

$$\begin{aligned}&[i\hbar\partial_t - (\hbar w/2k)\partial_x^2 + \tfrac{1}{2}\hbar wkx^2]c_{n_2}^{(k)}(t,x) \\ &\quad = (E_1 + \tfrac{1}{2}\hbar w)c_{n_2}^{(k)}(t,x),\end{aligned} \quad (6.4)$$

which has the well known solution

$$\begin{aligned}c_{n_2}^{(k)}(t,x) &= N_{n_2}^{(k)}\exp[-(i/\hbar)(E_1 - n_2\hbar w)t] \\ &\quad\times\exp[-(k/2)x^2]H_{n_2}(\sqrt{k}\,x),\end{aligned} \quad (6.5)$$

where H_{n_2} is the Hermite polynomial of order n_2, and $N_{n_2}^{(k)}$ is an arbitrary normalization constant. The nontrivial part of the problem consists precisely in adjusting the coefficients $N_{n_2}^{(k)}$ in such a manner as to have

$$\begin{aligned}\xi_k^*(\alpha;x-a) &= \exp\left(-\frac{i}{h}E_1\alpha\right)\left(\frac{k}{2}\right)^{1/2}\sum_{n_2=0}^{\infty}\frac{i^{n_2}N_{n_2}^{(k)}}{\sqrt{n_2!}} \\ &\quad\times\left(\sqrt{\frac{k}{2}}\right)^{n_2}(a+ib)^{n_2}H_{n_2}(\sqrt{k}\,x)\} \\ &\quad\times\exp\left(-\frac{k}{4}(a^2+b^2)\right) \\ &\quad\times\exp\left(i\frac{k}{2}(ab-2bx)\right),\end{aligned} \quad (6.6)$$

where we have written $\alpha = t - q^1$ and $a + ib = z^* e^{i\alpha w}$. Therefore, by means of the generating function of the Hermite polynomials,

$$\begin{aligned}&\sum_{n_2=0}^{\infty}(n_2!)^{-1}\left[\frac{\sqrt{k}}{2}(a+ib)\right]^{n_2}H_{n_2}(\sqrt{k}\,x) \\ &\quad = \exp\left(-\frac{k}{4}(a+ib)^2 + kx(a+ib)\right),\end{aligned} \quad (6.7)$$

one sees that the unique choice of $N_{n_2}^{(k)}$, which gives a function that depends on α and $x - a$, but *not* on b, in (6.6), is given by

$$N_{n_2}^{(k)} = (2\pi/k)^{1/2}(2^{n_2}n_2!)^{-1/2}(-i)^{n_2}N(k), \quad (6.8)$$

where the normalization constant $N(k)$ does not depend on n_2. This yields the desired answer, i.e.,

$$\xi_k^*(t;x) = N(k)e^{-(i/\hbar)E_1 t}e^{-(k/2)x^2}, \quad (6.9)$$

and

$$|t,x;E_1\rangle = (2\pi/k)^{1/2}N(k)e^{-(i/\hbar)E_1 t}e^{-(k/2)x^2} \times \sum_{n_2=0}^{\infty}(-i)^{n_2}(2^{n_2}n_2!)^{-1/2} \times e^{in_2 wt}H_{n_2}(\sqrt{k}x)|E_1 n_2 0\rangle. \quad (6.10)$$

Thus we obtain the physical space-time states $|t,x;E_1\rangle$ which carry the complementary ray representation of $\widetilde{G}$ within $\mathscr{H}_{E_1 0}\subset\mathscr{H}(\widetilde{G})$ in the following equivalent forms:

$$|t,x;E_1\rangle = N(k)e^{-(i/\hbar)E_1 t}e^{-(k/2)x^2}\int d\mu(q)\exp\left(\frac{i}{\hbar E_1 q^1}\right) \times \sum_{n_2=0}^{\infty}\frac{(\sqrt{k}/2)^{n_2}}{n_2!}\exp(in_2 w(t-q^1))z^{*n_2}H_{n_2}(\sqrt{k}x) \times\exp(k/4|Z|^2(1-i\tan wq^1))|q\rangle, \quad (6.11)$$

or

$$|t,x;E_1\rangle = N(k)\int d\mu(q)\exp(i/\hbar E_1(t-q^1)) \times\exp(-k/2[x-q^2\cos w(t-q^1) - q^3\sin w(t-q^1)]^2) \times\exp(-i(k/4)\{[(q^2)^2-(q^3)^2] \times\sin 2w(t-q^1)-2q^2q^3\cos 2w(t-q^1)\}) \times\exp\{ikx[q^2\sin w(t-q^1)-q^3\cos w(t-q^1)]\} \times\exp(i(k/4)[(q^2)^2+(q^3)^2]\tan wq^1)|q\rangle. \quad (6.12)$$

Interestingly enough, the fundamental wave function ξ_k that generates these states is the familiar ground state wave function (6.9) of the harmonic oscillator. Using the fact that

$$\langle E_1' n_2' 0|E_1 n_2 0\rangle = \delta(E_1'-E_1)\delta_{n_2' n_2}, \quad (6.13)$$

then, for any given physical state $|\psi;E_1\rangle\in\mathscr{H}_{E_1 0}$, i.e.,

$$|\psi;E_1\rangle = \sum_{n_2} c_{n_2}(\psi)|E_1 n_2 0\rangle, \quad (6.14)$$

one has

$$\langle t,x;E_1'|\psi;E_1\rangle = \delta(E_1'-E_1)\psi_{E_1}(t,x), \quad (6.15)$$

where the wave function $\psi_{E_1}(t,x)$ is given by

$$\psi_{E_1}(t,x) = \left(\frac{k}{2}\right)^{1/2}N^*(k)e^{(i/\hbar)E_1 t}e^{-(k/2)x^2} \times\sum_{n_2=0}^{\infty}\frac{i^{n_2}c_{n_2}(\psi)}{\sqrt{2^{n_2}}n_2!}e^{-in_2 wt}H_{n_2}(\sqrt{k}x). \quad (6.16)$$

Therefore, for the time-dependent Schrödinger equation, one gets

$$\langle t,x;E_1'|S_k|\psi;E_1\rangle = \delta(E_1'-E_1)[-i\hbar\partial_t-(\hbar^2/2m)\partial_x^2 + \tfrac{1}{2}mw^2x^2]\psi_{E_1}(t,x) = \delta(E_1'-E_1)(E_1+\tfrac{1}{2}\hbar w)\psi_{E_1}(t,x), \quad (6.17)$$

as it should be. In this way, one obtains the physical description of actual interest modulated by a delta function which comes from the superselection rule. Plainly, integrating over E_1' yields the final answer

$$[-i\hbar\partial_t-(\hbar^2/2m)\partial_x^2+\tfrac{1}{2}mw^2x^2]\psi_{E_1}(t,x) = (E_1+\tfrac{1}{2}\hbar w)\psi_{E_1}(t,x), \quad (6.18)$$

which, of course, is consistent with Eq. (6.16).

Finally, let us briefly discuss the Feynman propagator of the harmonic oscillator from the standpoint of quantum kinematics. From Eqs. (6.10) and (6.13) one obtains

$$\langle t,x';E_1'|t,x;E_1\rangle = \delta(E_1'-E_1)\delta(x'-x), \quad (6.19)$$

if one uses the following normalization:

$$N(k) = 2^{-1/2}\pi^{-3/4}k^{3/4}. \quad (6.20)$$

Hence, since

$$\langle t',x';E_1'|S_k|t,x;E_1\rangle = [i\hbar\partial_t-(\hbar w/2k)\partial_x^2+(\hbar wk/2)x^2] \times\langle t',x';E_1'|t,x;E_1\rangle = (E_1+\tfrac{1}{2}\hbar w)\langle t',x';E_1'|t,x;E_1\rangle, \quad (6.21)$$

it follows that the infinite series

$$\langle t',x';E_1'|t,x;E_1\rangle = \sqrt{\pi k}\,\delta(E_1'-E_1)e^{(i/\hbar)E_1(t'-t)}e^{-(k/2)(x'^2+x^2)} \times\sum_{n_2=0}^{\infty}(2^{n_2}n_2!)^{-1}e^{-in_2 w(t'-t)}H_{n_2}(\sqrt{k}x')H_{n_2}(\sqrt{k}x) \quad (6.22)$$

must be the quantum mechanical amplitude (i.e., the *kernel*) to get from the event (t,x) (in a "medium" with energy E_1) to the event (t',x') (in a "medium" with energy E_1'). Clearly, during this process the harmonic oscillator evolves freely, without interacting with the medium [the delta function $\delta(E_1'-E_1)$ explicitly describes this fact]. Indeed, it is well known that (6.22) corresponds to the full Green's function of Eq. (6.17), satisfying the initial condition (when $t'=t$) stated in Eq. (6.19). However, the important point we wish to remark is the following. If one uses Eq. (6.12), then a rather lengthy, albeit direct, process of integration casts this kernel in a closed form. Thus one obtains

$$\langle t',x';E_1'|t,x,E_1\rangle = \frac{1}{2}\left(\frac{k^3}{\pi}\right)^{1/2}\delta(E_1'-E_1)\exp\left(\frac{i}{\hbar}E_1(t'-t)\right)\iint dq^2\,dq^3 \times\exp\left(\frac{k}{2}[(x'-q^2\cos wt'-q^3\sin wt')^2+(x-q^2\cos wt-q^3\sin wt)^2]\right) \times\exp i(k/2)(q^2\cos wt'+q^3\sin wt')(q^2\sin wt'-q^3\cos wt') \times\exp(-i(k/2)(q^2\cos wt+q^3\sin wt)(q^2\sin wt-q^3\cos wt))$$

$$\times \exp\{-ik[x'(q^2 \sin wt' - q^3 \cos wt') - x(q^2 \sin wt - q^3 \cos wt)]\}$$
$$= \delta(E_1' - E_1)\exp\left(\frac{i}{\hbar}\left[E_1 + \frac{1}{2}\hbar w\right](t'-t)\right)\left[\frac{mw}{2\pi i h \sin w(t'-t)}\right]^{1/2}$$
$$\times \exp\left(i\frac{mw}{2\hbar}\frac{(x'^2+x^2)\cos w(t'-t) - 2x'x}{\sin w(t'-t)}\right), \qquad (6.23)$$

which corresponds exactly with the kernel of the simple harmonic oscillator, as obtained in the *path integral* approach.[23] Nevertheless, the integral one evaluates in Eq. (6.23) is a *Hurwitz invariant integral over the group manifold*.

We deem this fundamental group theoretic result (which here appears as a humble exercise or consistency control) as something deeply rooted in the quantum formalism, which deserves further research. For this reason, let us briefly discuss a possible physical significance of Eq. (6.23) in a rough (and tentative) manner. Let us recall some well known elementary ideas. Given an event $E_1 = (t_1, x_1)$ (see Fig. 1) it is clear from Eq. (1.2) that all the sinusoidal world lines (like W_1) with parameters (α, β) on the locus $\alpha = \beta \tan wt_1 + x_1 \sec wt_1$ (of the classical state space) go through the given event, and *vice versa*. Hence, given *two* events, $E_1 = (t_1, x_1)$ and $E_2 = (t_2, x_2)$, there is in general *one*, and only one, sinusoid (not shown in Fig. 1) that contains these two space-time points (or else, under very particular circumstances, there is none, but we may disregard these exceptional cases). Therefore, given a world line W_1 through E_1, the *classical* probability of finding the oscillator particle at E_2 is *zero* or *one*. (In the configuration shown in Fig. 1, it would be zero, since W_1 does not go through E_2.) In quantum mechanics, however, because of the quantum fluctuations there is a continuous (nonzeroth) probability amplitude for the oscillator to go from event E_1 to any other given event E_2, whatever its state may be at E_1. Now, as we show in Fig. 1 (though in a rather sketchy way), the main conceptual difference between the Feynman and the kinematic "pictorial" interpretation lies in the fact that, according to quantum kinematics, the oscillator evolves permanently under the Hooke law of force and, therefore, the quantum fluctuations (schematically denoted as the events A,B,C,D,..., in Fig. 1) derail the system, which jumps from one allowable world line to another. In other words, because of the simultaneous presence of both mechanical entities (i.e, the elastic force *and* the quantum fluctuations) the system can go from E_1 to E_2 following any piecewise continuous curve (like E_1ABCDE_2) whose arcs belong to the admissible (i.e., sinusoidal) world lines. Hence, in order to determine the probability amplitude for the desired space-time transition ($E_1 \to E_2$) one has to sum the "contributions" coming from *all the allowable world lines*, whether they pass through the given events or not. But this is precisely the job performed by a Hurwitz-invariant integral pertaining to the group of transformations that interconvert one world line into another. On the other hand, as is indeed well known, in the path integral approach to this issue[23] one visualizes the system as going from E_1 to E_2 along *all conceivable world lines* connecting these two events, whether they are consistent with the force law or not. Curiously enough, for the harmonic oscillator, both descriptions give the same answer.

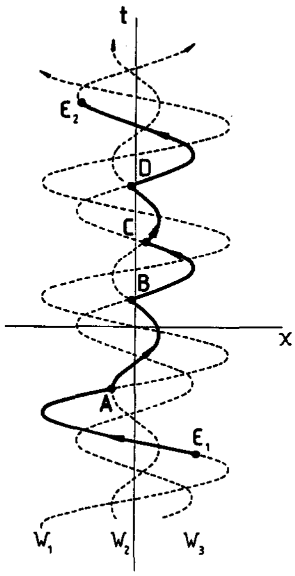


FIG. 1. Because of the quantum fluctuations (A,B,C,D,...) there is a nonzeroth probability amplitude for the particle to go from the event E_1 to the event E_2, under the permanent action of the elastic force.

VII. CONCLUDING REMARKS AND PERSPECTIVES

In this paper we have "quantized" the harmonic oscillator through its Newtonian relativity group. It is interesting to observe that, in the sense of Eq. (1.1), a harmonic oscillator S describes a simple harmonic motion only with respect to a preferred set of Newtonian observers (i.e., frames of reference) who also perform simple harmonic motions (of the same frequency) relative to the "laboratory" frame. Furthermore, the laboratory system itself belongs to this set (though, of course, it performs a simple harmonic motion of frequency w with zero amplitude relative to itself). It is clear that for any other Newtonian observer of S, who does *not* belong to the set defined in Eq. (1.1), S will *not* appear as a simple harmonic oscillator. Thus, Eq. (1.1) entails the *special relativity theory of the one-dimensional Newtonian harmonic oscillator*, and what we have done in the present work is to quantize this relativity group.

Elementary as they are, these relativistic features are very important from the standpoint of quantum kinematics. Hence, let us briefly underline some points in connection with this issue. If one manages an *isolated system S* by means of its *special relativity group G* [10] one is handling, once and for all, the whole set of preferred frames of reference relative to which S appears in essentially the same fashion and it behaves in a well defined standard manner. Of course, the elements of G induce a change in the states of S (the states of S are not invariant under G), but G does not change the very nature of S. It is S itself that is left invariant by the elements of G and, whence, it seems possible to use G as a *definition* of S [7]. Therefore, if one "sweeps" the group manifold of G one goes over all the allowable classical states of S; that is, one may consider all the possible histories of the system, over its configuration space-time, by means of its special relativity group. This is precisely what quantum kinematics requires for the description of S, since from the point of view of this formulation of the subject, as we have seen in our particular example, quantum mechanics arises as a peculiar theory in which the space-time propagator of S (i.e., the Green's function, or kernel, of the time-dependent Schrödinger equation) can be calculated by means of a Hurwitz-invariant integration over the whole group manifold, thus sweeping all the possible classical histories of S. This makes an intricate contrast with Feynman's path integral approach to quantization.[22] Furthermore, the Schrödinger equation itself can be obtained in a rather natural way, within the context of quantum kinematics, by means of a complementary space-time representation of the relativity group. In this sense, it must be borne in mind that the algebraic form of the familiar Hamiltonian operator H_k (including the energy scale factor $\hbar w/2$) has been *calculated* in the present endeavor, once we demand an invariant operator of the extended Lie algebra. No matter how easy this calculation has been in this particular instance, what we wish to underline is the fact that the classical Hamiltonian of the system has been totally ignored from the beginning (at least, explicitly) in order to achieve the correct form of the dynamical Hamiltonian operator, as well as to obtain the fundamental commutation relation of the ladder operators. Thus, everything seems to come out from the symmetries of the system. This result may have far reaching mechanical consequences if one thinks of an isolated system S for which a Lie group G affords a sufficiently complete description of its physically meaningful symmetries. No mechanical symmetry of S (whether external or internal) is lacking in G, by the hypothesis, and all its elements transform (any state of) S into (another state of) S, thus *leaving the system invariant*. In consequence, it seems reasonable to postulate the existence of an invariant operator (like the Schrödinger operator, for instance) which will contain all the information regarding the mechanical structure and the dynamical behavior of the system. In effect, like the system S itself, such an operator is invariant under G, and therefore it is a reasonable candidate for having a "mathematical representative" of the desired quantum model of S. Interestingly enough, these considerations (if valid generally) would throw new light into the fundamental role played by the Schrödinger equation in quantum mechanics, putting the whole quantum formalism into a new perspective, essentially as a *mechanics of symmetries*, quite independent of the classical formalism (which, then, would appear as the derived construct). On the other hand, the invariant operators of G, when interpreted physically, should correspond to some permanent mechanical properties characterizing the system S, and therefore they offer the possibility of introducing *superselection rules* in the Hilbert space $\mathscr{H}(\tilde{G})$, which carries the regular ray representations of the group. This is not just an *ad hoc* postulate since in any given physical state the permanent properties of S must have well defined permanent values. Hence, this postulate is a rather natural logical necessity of the kinematic approach to quantum theory. This means that, from a mechanical standpoint, we must look at $\mathscr{H}(G)$ as an incoherent Hilbert space, within which the superselection rules will be able to identify automatically the physical Hilbert subspaces of the system.

It is clear that in the present kinematic formalism one can interpret the kinematic Hamiltonian $P_1^{(k)}$ as the Hamiltonian of the surrounding medium of the harmonic oscillator (i.e., of the whole collection of material bodies by means of which the Newtonian frames of reference of the system become realized). [Let us recall, for instance, the "opposed ladder effects" stated in Eqs. (3.25) and (3.26).] Hence the continuous eigenvalue E_1 fixes the arbitrary zero level against which the energy of the oscillator is being measured. In quantum mechanics one usually sets $E_1 = -\frac{1}{2}\hbar w$, and then Eq. (6.18) yields the familiar time-dependent Schrödinger equation of the system.

Another important point we wish to discuss, which seems to correspond to a general fact of the relativity theory of motion, is that there are two equivalent quantum kinematic theories of the simple harmonic oscillator. One is directly related with the space-time realization (1.1) of its Newtonian kinematic group G, while the other is based on the state-space realization (1.3) of the group E_2, isomorphic with G. Both theories must be physically interesting since both embrace the mechanical system, though from quite different geometric points of view. This paper was devoted to the space-time kinematic model exclusively. The state-space kinematic model of the harmonic oscillator is already under study, and will be considered in a forthcoming publication. The way one handles the time coordinate in the construct considered in this paper should also be discussed. Indeed, in quantum kinematics one considers t on the same footing as x, as demanded in general by the spirit of relativity theory. This demand is not merely formal (and, thus, devoid of physical importance) for it is at the very root of every modern theory of motion. In the last analysis, Eq. (5.2) gives us the relativistic recipe by which the equivalent (or preferred) Newtonian observers of the harmonic oscillator transform the wave function describing the state of the system. For the Galileo and the Poincaré groups (that is, for free particle systems) this is a well known subject, of course, and what we have done in the present paper amounts to a direct generalization thereof. Let us recall that in the current approach to quantum mechanics one sets $\psi(t,x) = \langle x|\psi;t\rangle$, instead of Eq. (5.3), with $|\psi;t\rangle = \exp[-(i/h)tH]|\psi;0\rangle$, instead of Eq. (5.2). Hence, the current formalism is completely different from quantum kinematics, for it singles out t as a c number, while quantizes x, and therefore contradicts the relativistic requirement, even at the Newtonian level examined in this paper. We will come back to the issue of the *time operator* from the standpoint of quantum kinematics (which has not been touched in this work).

As it stands, quantum kinematics is a general framework rather than a specific theory (as it also happens with supersymmetry,[24] for instance). It seems to be an interesting subject worth further research.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Louis Michel for useful discussions and encouragement in connection with the work described here, and for his friendly hospitality at the Institute des Haute Etudes Scientifiques (I.H.E.S.) where much of this work was done. The author thanks Henry Bacry, whose wise criticism prompted some of the thoughts behind this paper. (Of course, the authors accepts full responsibility for the outcome of their prompting.)

A travel grant from PNUD (UNESCO) is hereby acknowledged. This work was supported in part by FNDCYT, under Contract 1106 (85), in part by DIUC, Grant 21/85, and in part by DGI (UCV), Grant 123.746/85.

[1] J. Krause, J. Phys. A **18**, 1309 (1985).

[2] H. Weyl, *The theory of Groups and Quantum Mechanics* (Dover, New York, 1931).

[3] The problem of quantizing a system that is primarily described by non-Abelian dynamical variables seems to be hitherto an unsolved problem of quantum kinematics [cf. K. Yamada, Phys. Rev. D **25**, 3256 (1982)], because the conventional commutators of Heisenberg and Jordan, i.e., $[Q^a, P_b] = i\hbar\delta^a_b$, apply only to Cartesian variables, and also because Weyl's quantization contains the non-Abelian canonical commutators only implicitly. Thus it turns out to be not flexible enough for the purposes of physics [cf. I. Daubechies, J. Math. Phys. **24**, 1453 (1983)]. The same problem arises in connection with some recent attempts to generalize the theory of minimum-uncertainty states to variables other then Cartesian [cf. M. M. Nieto, L. M. Simmons, Jr., and V. P. Gutschick, Phys. Rev. D **23**, 927 (1981)]. The quantum kinematics of non-Abelian variables is also an indispensable device in the formulation of lattice gauge theory; cf., for instance, J. Kogut, Phys. Rep. **67**(I), 67 (1980).

[4] See, i.e., G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York, 1969).

[5] V. Aldaya and J. A. Azcárraga, J. Math. Phys. **23**, 1297 (1982); R. T. Prosser, *ibid.* **24**, 548 (1983), and references quoted therein.

[6] See, for instance, M. Aguirre and J. Krause, J. Math. Phys. **25**, 210 (1984).

[7] K. Mariwalla, Phys. Rep. **20**, 287 (1975).

[8] C. E. Wulfman and B. G. Wybourne, J. Phys. A **9**, 597 (1976).

[9] J. Krause, "SO(3) commutators for angular momentum and rotation observables," in *Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 201, edited by G. Denardo *et al.* (Springer, Berlin, 1984), p. 502.

[10] H. Bacry and J. M. Lévy-Leblond, J. Math. Phys. **9**, 1605 (1968).

[11] J. M. Lévy-Leblond, Commun. Math. Phys. **12**, 64 (1969).

[12] J. C. Houard, J. Math. Phys. **18**, 502 (1977).

[13] R. F. Streater, Commun. Math. Phys. **4**, 217 (1967).

[14] Applications of the theory of continuous groups in quantum mechanics are legion, and not even a listing of recent work can be given here. However, we would call the reader's attention to several articles on the quantum mechanical harmonic oscillator; i.e., G. A. Baker, Phys. Rev. **103**, 1119 (1956); V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960), **23**, 177 (1961); also, A. O. Barut, Phys. Rev. **139**, B1433 (1965).

[15] E. P. Wigner, *Group Theory* (Academic, New York, 1959).

[16] V. Bargmann, Ann. Math. **59**, 1 (1954).

[17] Cf. for instance, M. A. Naimark and A. I. Stern, *Theory of Group Representations* (Springer, New York, 1982).

[18] For example, two-cocycles have been investigated recently, in connection with the Wess–Zumino–Witten anomaly, as nontrivial extensions of current algebra, and also in the cohomology of Wess–Zumino Lagrangians of gauge fields; cf. T. Fujiwara, Phys. Lett. **152**, 103 (1985), and Y. S. Wu, *ibid.* B **153**, 70 (1985), respectively. Even three-cocycles are becoming fashionable in the quantum mechanical description of a magnetic monopole; see Y. S. Wu and A. Zee, Phys. Lett. **152**, 98 (1985); B. Grossman, *ibid.* **152**, 93 (1985). Compare also R. Jackiw, "Quantum mechanical symmetry breaking," in *Recent Developments in Quantum Field Theory*, edited by J. Ambjorn *et al.* (Elsevier, New York, 1985).

[19] S. Eilenberg and S. Maclane, Ann. Math. **48**, 57, 326 (1947); S. Eilenberg, Bull. Am. Math. Soc. **55**, 3 (1949). These are the "classical" papers on the subject. See, also, L. Michel, "Invariance in quantum mechanics and group extension," in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).

[20] J. Krause, "Pedestrian approach to two-cocycles for unitary ray representations of Lie groups," preprint, 1986, submitted for publication.

[21] The invariant operators presented in Eqs. (3.15) and (3.27) have been obtained by means of the *associated adjoint representation* (cf., for instance, G. Rosen, Ref. 4) of the extended kinematic algebra of $\widetilde{G}$. By the same token, it can be shown that the system has no more invariants than S_k, b_k, $b_k^\dagger$, and functions thereof, as it must be. See, also, J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).

[22] An interesting discussion of this subject (though in a different physical context) can be found in the papers by J. M. Lévy-Leblond, Ref. 11, and J. C. Houard, Ref. 12. Also, J. Krause, "Regular and complementary ray representations of Lie groups," preprint, 1986.

[23] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).

[24] B. Zumino, CERN Courrier January-February, 18 (1983).

Geometric quantization: Modular reduction theory and coherent states

S. Twareque Ali
Department of Mathematics, Concordia University, Montréal, Canada H4B 1R6
Gérard G. Emch[a)]
Institut fur Theoretische Physik, Universität Göttingen, D-3400 Göttingen, West Germany

(Received 30 October 1985; accepted for publication 6 August 1986)

The natural role played by coherent states in the geometric quantization program is brought out by studying the mathematical equivalence between two physical interpretations that have recently been proposed for this program. These interpretations are based, respectively, on the modular algebra structure of prequantization, and the reproducing kernel structure of phase space quantization. The arguments are presented in this paper for the particular case where the phase space of the system considered is the cotangent bundle T^*M of a homogeneous manifold M, and for didactic reasons, the latter is taken to be a real vector space.

I. INTRODUCTION

The purpose of this paper is to propose a synthesis between two approaches,[1,2] developed separately by the present authors, to Schrödinger quantization. In the process of this synthesis, a new element will be brought to light, namely the central role played by *coherent states* in justifying the choice of polarizations that are made in both the modular algebra approach of Ref. 1, and the reproducing kernel approach of Ref. 2. We postpone to Sec. IV the discussion of the physical interpretation of these polarizations in terms of measuring processes.

In Ref. 1, one starts by associating to the prequantization map (see Refs. 3–5, and references cited there to the development issuing from the work of Souriau and Kostant) a representation $\mathscr{W}_\Gamma$ of the canonical commutation relations (CCR) algebra, acting on the Koopman Hilbert space $\mathscr{H}_\Gamma = L^2(T^*M)$, where M is the configuration space of the system, and T^*M is the cotangent bundle of M. This representation will be written out explicitly at the beginning of Sec. II (where, for the sake of simplicity, we will work with one degree of freedom only). The transition from prequantization to the usual Schrödinger quantum theory is then made by giving a prescription for isolating irreducible subrepresentations of $\mathscr{W}_\Gamma$. This is achieved by exploiting the existence of atomic maximal Abelian von Neumann subalgebras $\mathscr{A}$ of $\mathscr{W}_\Gamma$. To each such subalgebra $\mathscr{A}$ is associated a faithful normal state φ on $\mathscr{W}_\Gamma$, which in fact is a vector state; the corresponding vector Φ is both cyclic and separating for $\mathscr{W}_\Gamma$. A KMS structure[6] is thus associated canonically to the pair $\{\mathscr{W}_\Gamma, \mathscr{A}\}$ and $\mathscr{A}$ turns out to be the centralizer of $\mathscr{W}_\Gamma$ with respect to φ. Conversely, each normal pure state on $\mathscr{A}$ is shown to lead, in a canonical manner, to an irreducible subrepresentation of $\mathscr{W}_\Gamma$.

In Ref. 2 one studies a representation U_0 of the extended Galilei group $\widetilde{G}$ on $L^2(T^*\mathbb{R}^3)$. This representation is obtained by starting with a representation of the CCR algebra $\mathscr{W}$ on $L^2(T^*\mathbb{R}^3)$, and then by extending it to a representation of $\widetilde{G}$. A complete decomposition theory of U_0 has been worked out in Ref. 7 (see, in particular Theorems 2.1–3.4), using the technique of reproducing kernel Hilbert spaces. Picking out from this decomposition, specific irreducible subrepresentations of U_0, corresponds again to the passage from prequantization to the usual Schrödinger representation.

II. REDUCTION THEORY AND COHERENT STATES

We consider two Hilbert spaces $\mathscr{H}_\Gamma \equiv L^2(T^*\mathbf{R}, dp\, dq)$ and $\mathscr{H}_s \equiv L^2(\mathbf{R}, dk)$, each supporting a specific representation of the CCR algebra $\mathscr{W}$. Let $\mathscr{U}(\mathscr{H}_\Gamma)$ [resp. $\mathscr{U}(\mathscr{H}_s)$] be the group of all unitary operators on $\mathscr{H}_\Gamma$ (resp. $\mathscr{H}_s$).

(i) On $\mathscr{H}_\Gamma$, let

$$\mathscr{W}_\Gamma: (p,q)\in T^*\mathbf{R} \to \mathscr{W}_\Gamma(p,q)\in \mathscr{U}(\mathscr{H}_\Gamma)$$

be defined by

$$(\mathscr{W}_\Gamma(p,q)\Psi)(p',q') \equiv \exp\{(i/\hbar)p(q'-q)\}\Psi(p'-p, q'-q), \tag{2.1}$$

for all $\Psi\in\mathscr{H}_\Gamma$. Then

$$\mathscr{W}_\Gamma(p_1,q_1)\mathscr{W}_\Gamma(p_2,q_2) = \exp\{(i/\hbar)p_1 q_2\}\mathscr{W}_\Gamma(p_1+p_2, q_1+q_2), \tag{2.2}$$

$$\mathscr{W}_\Gamma(p,q)^* = \exp\{(i/\hbar)pq\}\mathscr{W}_\Gamma(-p,-q). \tag{2.3}$$

Upon differentiating (2.1), we define two self-adjoint operators P_Γ and Q_Γ, namely

$$P_\Gamma = -i\hbar\frac{\partial}{\partial q} \quad \text{and} \quad Q_\Gamma = q + i\hbar\frac{\partial}{\partial p}, \tag{2.4}$$

with common domain of essential self-adjointness in $\mathscr{H}_\Gamma$, and such that

$$\mathscr{W}_\Gamma(p,q) = \exp\{-(i/\hbar)qP_\Gamma\}\exp\{(i/\hbar)pQ_\Gamma\}, \quad \forall (p,q)\in T^*\mathbf{R}, \tag{2.5}$$

and

$$[P_\Gamma, Q_\Gamma] = -\hbar I_\Gamma. \tag{2.6}$$

The representation $\mathscr{W}_\Gamma$ of the CCR algebra thus defined on $\mathscr{H}_\Gamma$ is reducible, and we shall soon exhibit special operators on $\mathscr{H}_\Gamma$ that commute with all $\mathscr{W}_\Gamma(p,q)$ [see (2.18) and (3.1) below].

[a)] Permanent address: Department of Mathematics, University of Florida, Gainesville, Florida 32611.

 0022-2488/86/122936-08$02.50

(ii) On $\mathscr{H}_s$, let

$$\mathscr{W}_s:\ (p,q)\in T^*\mathbf{R}\to \mathscr{W}_s(p,q)\in\mathscr{U}(\mathscr{H}_s)$$

be defined by

$$(\mathscr{W}_s(p,q)\widetilde{\Psi})(k)\equiv\exp\{-(i/\hbar)kq\}\widetilde{\Psi}(k-p), \quad (2.7)$$

for all $\Psi\in\mathscr{H}_s$. Then $\mathscr{W}_s$ also satisfies the algebraic relations written as (2.2) and (2.3) for $\mathscr{W}_\Gamma$; upon differentiating (2.7), we define two self-adjoint operators P_s and Q_s, namely

$$P_s = k \quad \text{and} \quad Q_s = i\hbar\frac{\partial}{\partial k}, \quad (2.8)$$

with common domain of essential self-adjointness in $\mathscr{H}_s$, and such that

$$\mathscr{W}_s(p,q) = \exp\{-(i/\hbar)qP_s\}\exp\{(i/\hbar)pQ_s\},\quad \forall(p,q)\in T^*\mathbf{R}, \quad (2.9)$$

and

$$[P_s, Q_s] = -i\hbar I_s. \quad (2.10)$$

Note that $\mathscr{W}_s$ is irreducible.

(iii) We next associate to every vector $\widetilde{\xi}\in\mathscr{H}_s$ with

$$\|\widetilde{\xi}\|^2 = (2\pi\hbar)^{-1}, \quad (2.11)$$

the partial isometry $\widehat{V}_\xi$: $\widetilde{\Psi}\in\mathscr{H}_s\mapsto\Psi_\xi\in\mathscr{H}_\Gamma$ defined by

$$\begin{aligned}(\widehat{V}_\xi\widetilde{\Psi})(p,q) &\equiv \int_{\mathbf{R}} dk\,\exp\left\{\frac{i}{\hbar}kq\right\}\widetilde{\xi}(k-p)^*\widetilde{\Psi}(k)\\ &= \langle\widetilde{\xi}^{p,q},\widetilde{\Psi}\rangle, \quad \text{with} \quad \widetilde{\xi}^{p,q}\equiv\mathscr{W}_s(p,q)\widetilde{\xi}. \end{aligned}\quad (2.12)$$

Note that

$$\widehat{V}_\xi^*\widehat{V}_\xi = I_s \quad \text{and} \quad \widehat{V}_\xi\widehat{V}_\xi^* = \mathbf{P}_\xi, \quad (2.13)$$

where the explicit action of the projector $\mathbf{P}_\xi$ is given by

$$(\mathbf{P}_\xi\Psi)(p,q) = \iint_{T^*\mathbf{R}} dp'\,dq'\,K_\xi(p,q;p',q')\Psi(p',q'), \quad (2.14)$$

with

$$K_\xi(p,q;p',q')\equiv\int_{\mathbf{R}} dk\,\exp\left\{\frac{i}{\hbar}k(q-q')\right\}\times\widetilde{\xi}(k-p)^*\widetilde{\xi}(k-p'). \quad (2.15)$$

Furthermore, the partial isometry $\widehat{V}_\xi$ intertwines the representations $\mathscr{W}_s$ and $\mathscr{W}_\Gamma$, i.e.,

$$\mathscr{W}_\Gamma(p,q)\widehat{V}_\xi = \widehat{V}_\xi\mathscr{W}_s(p,q), \quad \forall(p,q)\in T^*\mathbf{R}, \quad (2.16)$$

or equivalently

$$\widehat{V}_\xi^*\mathscr{W}_\Gamma(p,q)\widehat{V}_\xi = \mathscr{W}_s(p,q), \quad \forall(p,q)\in T^*\mathbf{R}. \quad (2.17)$$

Together with (2.13), (2.16) implies

$$[\mathbf{P}_\xi, \mathscr{W}_\Gamma(p,q)] = 0, \quad \forall(p,q)\in T^*\mathbf{R}, \quad (2.18)$$

so that $\mathbf{P}_\xi$ reduces $\mathscr{W}_\Gamma$. We denote by $\mathscr{H}_\xi$ the subspace $\mathbf{P}_\xi\mathscr{H}_\Gamma$; by V_ξ: $\mathscr{H}_s\to\mathscr{H}_\xi$ the unitary operator obtained from $\widehat{V}_\xi$; by $\mathscr{W}_\xi$ the representation of the CCR algebra obtained as the restriction of $\mathscr{W}_\Gamma$ to the stable subspace $\mathscr{H}_\xi$; by P_ξ and Q_ξ the generators of $\mathscr{W}_\xi$, i.e.,

$$\mathscr{W}_\xi(p,q) = \exp\{-(i/\hbar)qP_\xi\}\exp\{(i/\hbar)p\,Q_\xi\},\quad \forall(p,q)\in T^*\mathbf{R}. \quad (2.19)$$

We thus obtain

$$\mathscr{W}_\xi(p,q) = V_\xi W_s(p,q)\widehat{V}_\xi^*, \quad \forall(p,q)\in T^*\mathbf{R}, \quad (2.20)$$

so that $\mathscr{W}_\xi$ is unitarily equivalent to $\mathscr{W}_s$ and is thus irreducible. Upon differentiating (2.17) and (2.20) we obtain

$$\widehat{V}_\xi^*P_\Gamma\widehat{V}_\xi = P_s \quad \text{and} \quad \widehat{V}_\xi^*Q_\Gamma\widehat{V}_\xi = Q_s, \quad (2.21)$$

$$P_\xi = V_\xi P_s\widehat{V}_\xi^* \quad \text{and} \quad Q_\xi = V_\xi Q_s\widehat{V}_\xi^*. \quad (2.22)$$

The operators

$$P_\xi = -i\hbar\frac{\partial}{\partial q} \quad \text{and} \quad Q_\xi = q + i\hbar\frac{\partial}{\partial p} \quad (2.23)$$

have a common domain of essential self-adjointness in $\mathscr{H}_\xi$, on which they act irreducibly and satisfy the CCR, i.e.,

$$[P_\xi, Q_\xi] = -i\hbar I_\xi. \quad (2.24)$$

A choice of $\widetilde{\xi}\in\mathscr{H}_s$ leads thus to a sort of polarization—a point to which we shall return in Sec. IV.

(iv) In line with the Koopman Hilbert space formalism for classical mechanics,[8] we also consider the two classical (commuting) operators P^{cl} and Q^{cl}, of momentum and position, respectively, defined by

$$\begin{aligned}(P^{\mathrm{cl}}\Psi)(p,q) &= p\Psi(p,q),\\ (Q^{\mathrm{cl}}\Psi)(p,q) &= q\Psi(p,q).\end{aligned} \quad (2.25)$$

Note that there exists a dense domain $\mathscr{D}\subset\mathscr{H}_\Gamma$ that can serve as a common domain of essential self-adjointness for P^{cl}, Q^{cl}, P_ξ, and Q_ξ.

For every $\widetilde{\xi}\in\mathscr{H}_s$ such that

$$\begin{aligned}\langle P\rangle_\xi &\equiv \|\widetilde{\xi}\|^{-2}\int dk\,k\,|\widetilde{\xi}(k)|^2,\\ \langle Q\rangle_\xi &\equiv \|\widetilde{\xi}\|^{-2}\,i\hbar\int dk\,\xi(k)^*\frac{\partial\xi}{\partial k}(k)\end{aligned} \quad (2.26)$$

are well-defined and finite, we obtain

$$\begin{aligned}\widehat{V}_\xi^*\overline{P}^{\mathrm{cl}}\widehat{V}_\xi &= P_s,\\ \widehat{V}_\xi^*\overline{Q}^{\mathrm{cl}}\widehat{V}_\xi &= Q_s,\end{aligned} \quad (2.27)$$

where

$$\begin{aligned}\overline{P}^{\mathrm{cl}} &\equiv P^{\mathrm{cl}} + \langle P\rangle_\xi = U_\Gamma(\xi)^*P^{\mathrm{cl}}U_\Gamma(\xi),\\ \overline{Q}^{\mathrm{cl}} &\equiv Q^{\mathrm{cl}} + \langle Q\rangle_\xi = U_\Gamma(\xi)^*Q^{\mathrm{cl}}U_\Gamma(\xi),\end{aligned} \quad (2.28)$$

with

$$U_\Gamma(\xi)\equiv\mathscr{W}_\Gamma(\langle P\rangle_\xi, \langle Q\rangle_\xi). \quad (2.29)$$

Equivalently

$$\mathbf{P}_\xi\overline{P}^{\mathrm{cl}}\mathbf{P}_\xi = P_\xi \quad \text{and} \quad \mathbf{P}_\xi\overline{Q}^{\mathrm{cl}}\mathbf{P}_\xi = Q_\xi, \quad (2.30)$$

or (see already Refs. 2 and 7)

$$\begin{aligned}\mathbf{P}_\xi P^{\mathrm{cl}}\mathbf{P}_\xi &= P_\xi - \langle P\rangle_\xi I_\xi = U_\xi P_\xi U_\xi^*,\\ \mathbf{P}_\xi Q^{\mathrm{cl}}\mathbf{P}_\xi &= Q_\xi - \langle Q\rangle_\xi I_\xi = U_\xi Q_\xi U_\xi^*,\end{aligned} \quad (2.31)$$

with

$$U_\xi = \mathscr{W}_\xi(\langle P\rangle_\xi, \langle Q\rangle_\xi). \quad (2.32)$$

The operators (2.31) satisfy the CCR, and indeed when $\langle P\rangle_\xi = 0 = \langle Q\rangle_\xi$ (cf. the discussion at the end of this section), they coincide with the reduced operators (2.23).

Let us analyze next some special features of the subspace $\mathscr{H}_\xi$, which carries the irreducible representation $\mathscr{W}_\xi$ of the CCR algebra.

(i) For each $\tilde{\xi}\in\mathscr{H}_s$, the set

$$\tilde{\mathscr{S}}_\xi \equiv \{\tilde{\xi}^{p,q} | (p,q)\in T^*\mathbf{R}\} \tag{2.33}$$

of vectors $\tilde{\xi}^{p,q}\in\mathscr{H}_s$ appearing in (2.12) is overcomplete, and constitutes a family of generalized coherent states in the sense of Klauder[9] and Perelomov.[10] For each Borel set $\Delta \subset T^*\mathbf{R}$, let us define a positive operator

$$\tilde{a}_\xi(\Delta) \equiv \int_\Delta dp\, dq |\tilde{\xi}^{p,q}\rangle\langle\tilde{\xi}^{p,q}|. \tag{2.34}$$

The set of operators $\tilde{a}_\xi(\Delta)$, for all Borel sets $\Delta\subset T^*\mathbf{R}$, then constitutes a positive-operator-valued (POV) measure, in the sense that the following measure-theoretic properties hold:

$$\tilde{a}_\xi(\varnothing) = 0 \quad \text{(where } \varnothing \text{ denotes the null set)}, \tag{2.35}$$

$$\tilde{a}_\xi\Big(\bigcup_{j\in J} \Delta_j\Big) = \sum_{j\in J} \tilde{a}_\xi(\Delta_j). \tag{2.36}$$

Moreover

$$\tilde{a}_\xi(T^*\mathbf{R}) = I_s. \tag{2.37}$$

In (2.36), J is a discrete index set, $\Delta_i \cap \Delta_j = \varnothing$, for $i\neq j$, and the convergence of the sum holds in the weak operator topology. In view of (2.37), we call $\tilde{a}_\xi$ a normalized POV measure.

(ii) for each $(p,q)\in T^*\mathbf{R}$, consider the vector $\xi^{p,q}\in\mathscr{H}_\xi\subset\mathscr{H}_\Gamma$ obtained as the image of $\tilde{\xi}^{p,q}$ under V_ξ. From

$$\begin{aligned} &\xi = V_\xi\tilde{\xi}, \\ &\xi(p,q) = \langle\tilde{\xi}^{p,q}|\tilde{\xi}\rangle, \\ &\xi^{p,q} = \mathscr{W}_\xi(p,q)\xi, \end{aligned} \tag{2.38}$$

we conclude that the set

$$\mathscr{S}_\xi \equiv \{\xi^{p,q} | (p,q)\in T^*\mathbf{R}\} \tag{2.39}$$

is overcomplete in $\mathscr{H}_\xi$. Moreover, on $\mathscr{H}_\Gamma$ we have the POV measure

$$\begin{aligned} a_\xi(\Delta) &\equiv V_\xi\tilde{a}_\xi(\Delta)V_\xi^* \\ &= \int_\Delta dp\, dq |\xi^{p,q}\rangle\langle\xi^{p,q}\rangle|, \end{aligned} \tag{2.40}$$

with

$$a_\xi(T^*\mathbf{R}) = \mathbf{P}_\xi. \tag{2.41}$$

(iii) The action of $\mathbf{P}_\xi$ on a vector $\Psi\in\mathscr{H}_\Gamma$ is given via a kernel K_ξ: $T^*\mathbf{R}\times T^*\mathbf{R}\to\mathbf{C}$:

$$(\mathbf{P}_\xi\Psi)(p,q) = \int_{T^*\mathbf{R}} dp'dq'\, K_\xi(p,q;p',q')\Psi(p',q'), \tag{2.42}$$

where

$$K_\xi(p,q;p',q') = \langle\xi^{p,q}|\xi^{p',q'}\rangle. \tag{2.43}$$

In particular, K_ξ is a reproducing kernel, since it enjoys the properties[11,12]

$$K_\xi(p,q;p,q) > 0, \tag{2.44}$$

$$K_\xi(p,q;p',q') = K_\xi(p',q';p,q)^*, \tag{2.45}$$

$$K_\xi(p,q;p',q') = \int_{T^*\mathbf{R}} dp''\, dq''\, K_\xi(p,q;p'',q'') \times K_\xi(p'',q'';p',q'). \tag{2.46}$$

Specifically, (2.46) is called the reproducing property, since it leads to the consequence that K_ξ, acting on any vector $\Psi_\xi\in\mathscr{H}_\xi$, reproduces the same vector, i.e., for all $(p,q)\in T^*\mathbf{R}$:

$$\int_{T^*\mathbf{R}} dp'\, dq' K_\xi(p,q;p',q')\Psi_\xi(p',q') = \Psi_\xi(p,q). \tag{2.47}$$

(iv) The existence of the reproducing kernel K_ξ on $\mathscr{H}_\xi$ is, in fact, an intrinsic property of the latter. This is made more transparent if we use (2.47) to define, for each $(p,q)\in T^*\mathbf{R}$, a bounded, linear, evaluation map $E_\xi^{p,q}$: $\mathscr{H}_\xi\to\mathbf{C}$ such that, for all $\Psi_\xi\in\mathscr{H}_\xi$,

$$E_\xi^{p,q}\Psi_\xi \equiv \int_{T^*\mathbf{R}} dp'\, dq' K_\xi(p,q;p',q')\Psi_\xi(p',q'), \tag{2.48}$$

i.e., see (2.47),

$$E_\xi^{p,q}\Psi_\xi = \Psi_\xi(p,q). \tag{2.49}$$

In fact the vectors $\Psi_\xi\in\mathscr{H}_\xi$ can be chosen to be *continuous* functions of (p,q) such that

$$|\Psi_\xi(p,q)| \leqslant (2\pi\hbar)^{-1}\|\Psi_\xi\|, \quad \forall (p,q)\in T^*\mathbf{R}. \tag{2.50}$$

In terms of $E_\xi^{p,q}$ and its adjoint map $(E_\xi^{p,q})^*$: $\mathbf{C}\to\mathscr{H}_\xi$, one may write[12]

$$K_\xi(p,q;p',q') = E_\xi^{p,q}(E_\xi^{p',q'})^*, \tag{2.51}$$

$$a_\xi(\Delta) = \int_\Delta dp\, dq (E_\xi^{p,q})^* E_\xi^{p,q}. \tag{2.52}$$

Thus, it is the existence of the bounded, linear evaluation map $E_\xi^{p,q}$ on $\mathscr{H}_\xi$, for each $(p,q)\in T^*\mathbf{R}$, which leads to the existence of the reproducing kernel K_ξ, the associated POV measure a_ξ, as well as the generalized coherent states $\xi^{p,q}$; indeed, by (2.12), (2.38), and (2.43),

$$K_\xi(p,q;p',q') = \xi^{p'q'}(p,q). \tag{2.53}$$

In its turn, the existence of $E_\xi^{p,q}$ is a consequence of a simple measure-theoretic property of $\mathscr{H}_\xi$. The space $\mathscr{H}_\Gamma$ consists of vectors that are really equivalence classes (with respect to the measure $dp\, dq$) of functions $[\Psi]$, i.e., two functions Ψ_1 and Ψ_2 belong to the same equivalence class $[\Psi]$ if and only if $\Psi_1(p,q) = \Psi_2(p,q)$ almost everywhere. In each equivalence class $[\Psi]$ we may choose a representative function Ψ to denote the class. However the choice $[\Psi]\mapsto\Psi$ cannot be made linear for the whole space $\mathscr{H}_\Gamma$. *Yet* on the subspace $\mathscr{H}_\xi$, the restricted map $[\Psi_\xi]\mapsto\Psi_\xi$ *is* linear. It is this association that then defines the evaluation map $E_\xi^{p,q}$ and hence the kernel K_ξ. Moreover, the vector ξ, which in view of (2.38)–(2.41) we call the *resolution generator* for $\mathscr{H}_\xi$, is unique. In other words, a bounded, linear, association $[\Psi]\mapsto\Psi$ on a subspace of $\mathscr{H}_\Gamma$ determines[7] a unique vector ξ in that subspace and a resulting reproducing kernel K_ξ. Thus with $\tilde{\xi}\in\mathscr{H}_s$ as in (2.11) and ξ being given by (2.38), the association $\tilde{\xi}\mapsto K_\xi$ is one-to-one.

We next see how $\mathscr{W}_\Gamma$ can be completely decomposed into irreducible subrepresentations of the type $\mathscr{W}_\xi$ just analyzed. For this purpose, let $\{\tilde{\xi}_n | n\in\mathbf{Z}^+\}$ be a basis in $\mathscr{H}_s$, normalized so that

$$\langle\tilde{\xi}_n|\tilde{\xi}_m\rangle = (2\pi\hbar)^{-1}\delta_{nm}. \tag{2.54}$$

To each $\tilde{\xi}_n$ we associate the reproducing kernel Hilbert space $\mathscr{H}_n \equiv \mathscr{H}_{\xi_n} \subset \mathscr{H}_\Gamma$, the image of $\mathscr{H}_s$ in $\mathscr{H}_\Gamma$ through the partial isometry $V_n \equiv \hat{V}_{\xi_n}$ defined in (2.12). Then [for instance as a consequence of Theorems (2.1) and (2.2) in Ref. 7], $\mathscr{H}_\Gamma$ decomposes as the infinite direct sum

$$\mathscr{H}_\Gamma = \bigoplus_{n\in\mathbf{Z}^+} \mathscr{H}_n, \tag{2.55}$$

i.e., $\mathscr{H}_\Gamma$ decomposes into mutually orthogonal subspaces of continuous functions. Denoting by $\mathscr{W}_n$ the restriction of $\mathscr{W}_\Gamma$ to the stable subspace $\mathscr{H}_n$, we also have, for all $(p,q)\in T^*\mathbf{R}$,

$$\mathscr{W}_\Gamma(p,q) = \bigoplus_{n\in\mathbf{Z}^+} \mathscr{W}_n(p,q). \tag{2.56}$$

Each pair $\{\mathscr{H}_n, \mathscr{W}_n\}$ constitutes the GNS representation of the CCR algebra $\mathscr{W}$, built from the pure state φ_n on $\mathscr{W}$ defined by

$$\langle \varphi_n; \mathscr{W}(p,q)\rangle = 2\pi\hbar\langle \xi_n | \mathscr{W}_\Gamma(p,q)\xi_n\rangle. \tag{2.57}$$

The arbitrariness in the choice of the basis $\{\tilde{\xi}_n | n\in\mathbf{Z}^+\}$ in $\mathscr{H}_s$ implies that the decomposition (2.55) and (2.56) is not unique. In particular, it is always possible to choose this basis (e.g., take the harmonic oscillator wave functions) in such a way that the relations (2.31) hold with

$$\langle P\rangle_{\xi_n} = 0 = \langle Q\rangle_{\xi_n}, \quad \forall n\in\mathbf{Z}^+, \tag{2.58}$$

i.e.,

$$\begin{aligned} \mathbf{P}_{\xi_n} P^{\mathrm{cl}} \mathbf{P}_{\xi_n} &= P_{\xi_n} = -i\hbar\frac{\partial}{\partial q}, \\ \mathbf{P}_{\xi_n} Q^{\mathrm{cl}} \mathbf{P}_{\xi_n} &= Q_{\xi_n} = q + i\hbar\frac{\partial}{\partial p}, \end{aligned} \quad \forall\, n\in\mathbf{Z}^+ \tag{2.59}$$

each of these pairs acting irreducibly [cf. (2.20) or (2.22)] in the corresponding Hilbert space $\mathscr{H}_n \equiv \mathscr{H}_{\xi_n}$; moreover, see again (2.55) and (2.56). Note also that if we take $\mathbf{R}^3$ as our configuration space, then the rotational symmetry exploited in Ref. 7 automatically ensures (2.58), and hence (2.59).

III. MODULAR ALGEBRAIC STRUCTURES

The analysis of $\mathscr{W}_\Gamma$ in Ref. 1 is based upon the observation that $\mathscr{H}_\Gamma$ supports a second representation of the CCR, generated by the operators

$$\hat{P} = -i\hbar\frac{\partial}{\partial q} - p \quad \text{and} \quad \hat{Q} = i\hbar\frac{\partial}{\partial p}. \tag{3.1}$$

Since both $\hat{P}$ and $\hat{Q}$ commute with the generators P_Γ and Q_Γ [cf. (2.4)] of $\mathscr{W}_\Gamma$, they generate an algebra that lies in $\mathscr{W}'_\Gamma$, the commutant of $\mathscr{W}_\Gamma$. An application of the Tomita–Takesaki theory of modular algebras[6] to an analysis of $\mathscr{W}_\Gamma$ then shows that $\hat{P}$ and $\hat{Q}$ do, in fact, generate $\mathscr{W}'_\Gamma$. We briefly review the aspects of the Tomita–Takesaki theory that are relevant to our analysis.

(i) If Φ is a cyclic (i.e., $[\mathscr{W}_\Gamma\Phi] = \mathscr{H}_\Gamma$) and separating (i.e., $A\Phi = 0$ and $A\in\mathscr{W}_\Gamma$ imply $A = 0$) vector for $\mathscr{W}_\Gamma$, then the antilinear map

$$S_\Phi:\ W\Phi\in\mathscr{W}_\Gamma\Phi\subset\mathscr{H}_\Gamma \mapsto W^*\Phi\in\mathscr{W}_\Gamma\Phi\subset\mathscr{H}_\Gamma \tag{3.2}$$

is closable. Its closure, which we again denote by S_Φ, has the polar decomposition

$$S_\Phi = J_\Phi \Delta_\Phi^{1/2}, \tag{3.3}$$

where J_Φ is an antiunitary operator satisfying

$$J_\Phi^2 = I, \quad J_\Phi\Phi = \Phi, \tag{3.4}$$

and Δ_Φ is a self-adjoint operator. Moreover, the map $W \mapsto J_\Phi W J_\Phi$ is an antilinear isomorphism from $\mathscr{W}_\Gamma$ onto $\mathscr{W}'_\Gamma$. It is this property that leads to the presence of $\hat{P}$ and $\hat{Q}$.

(ii) The faithful normal state φ on $\mathscr{W}_\Gamma$ defined, with Φ as in (i), by

$$\langle\varphi; W\rangle = \langle\Phi | W\Phi\rangle, \quad \forall W\in\mathscr{W}_\Gamma, \tag{3.5}$$

is a KMS state for the one-parameter group of evolution $t\mapsto\alpha_\varphi(t)$ on $\mathscr{W}_\Gamma$,

$$W \mapsto \alpha_\varphi(t)[W] = \Delta_\Phi^{-it/\beta} W \Delta_\Phi^{it/\beta} \tag{3.6}$$

(for further details and the analyticity properties of α_φ, refer to Ref. 1 for the particular case studied here, and to Ref. 6 for the general theory).

(iii) Defining the centralizer $\mathscr{M}_\varphi$ of $\mathscr{W}_\Gamma$, with respect to φ, as the von Neumann algebra

$$\mathscr{M}_\varphi \equiv \{A\in\mathscr{W}_\Gamma \,|\, \langle\varphi; [A, W]\rangle = 0, \quad \forall W\in\mathscr{W}_\Gamma\}, \tag{3.7}$$

one has, in addition,

$$\mathscr{M}_\varphi = \{A\in\mathscr{W}_\Gamma \,|\, \alpha_\varphi(t)[A] = A, \quad \forall t\in\mathbf{R}\}. \tag{3.8}$$

The analysis in Ref. 1 then establishes that every faithful normal state φ on $\mathscr{W}_\Gamma$ is a vector state, in the sense of (3.5), with the corresponding vector Φ being cyclic and separating for $\mathscr{W}_\Gamma$. Additionally, whenever φ is nondegenerate the centralizer $\mathscr{M}_\varphi$ is a maximal Abelian, atomic[13] von Neumann subalgebra of $\mathscr{W}_\Gamma$. Conversely, every maximal Abelian, atomic von Neumann subalgebra of $\mathscr{W}_\Gamma$ is the centralizer of a faithful, normal, nondegenerate state, and thus is obtained from a cyclic and separating vector.

Irreducible subrepresentations of $\mathscr{W}_\Gamma$ are then obtained in Ref. 1 by using normal pure states on such maximal Abelian, atomic subalgebras. The decomposition theory of $\mathscr{W}_\Gamma$ hence reduces to the problem of isolating the maximal Abelian, atomic von Neumann subalgebras of $\mathscr{W}_\Gamma$ or, what amounts to the same thing, the faithful, normal, nondegenerate states on $\mathscr{W}_\Gamma$. The existence of several such states leads, in this scheme, to the nonuniqueness of the decomposition of $\mathscr{W}_\Gamma$ into irreducible representations. Equivalently, the nonuniqueness of the decomposition reflects the existence of different evolutions (3.6) with their attendant KMS structures.

We now make explicit the link between this modular algebraic analysis of $\mathscr{W}_\Gamma$ and the analysis made in Sec. II, where reproducing kernels played a central role.

Let us fix a decomposition of $\mathscr{W}_\Gamma$ according to (2.55) and (2.56), i.e., let us choose a basis

$$\{\tilde{\xi}_n | n\in\mathbf{Z}^+\}\subset\mathscr{H}_s, \tag{3.9}$$

satisfying (2.54); we consider the corresponding orthogonal set

$$\{\xi_n \equiv \hat{V}_n\tilde{\xi}_n | n\in\mathbf{Z}^+\}\subset\mathscr{H}_\Gamma. \tag{3.10}$$

For each $n\in\mathbf{Z}^+$, ξ_n is a resolution generator for the subspace $\mathscr{H}_n$, and

$$\xi_n(p,q) = \langle\tilde{\xi}^{p,q} | \tilde{\xi}\rangle. \tag{3.11}$$

We fix, in addition, a sequence

$$\{\lambda_n | n \in \mathbf{Z}^+\} \subset \mathbf{R}$$

such that $\lambda_n > 0, \quad \forall n \in \mathbf{Z}^+$,

$\lambda_n \neq \lambda_m$, whenever $n \neq m$,

$$\sum_{n \in \mathbf{Z}^+} \lambda_n = 1. \tag{3.12}$$

We finally construct in $\mathscr{H}_\Gamma$ the vectors

$$\Phi = (2\pi\hbar)^{1/2} \sum_{n \in \mathbf{Z}^+} \lambda_n^{1/2} \xi_n, \tag{3.13}$$

$$\Phi^{p,q} = \mathscr{W}_\Gamma(p,q)\Phi$$

$$= (2\pi\hbar)^{1/2} \sum_{n \in \mathbf{Z}^+} \lambda_n^{1/2} \xi_n^{p,q}. \tag{3.14}$$

The following picture now emerges.

(i) Φ is a cyclic and separating vector for $\mathscr{W}_\Gamma$, and

$$\langle \varphi; \mathscr{W}_\Gamma(p,q) \rangle = \langle \Phi | \mathscr{W}_\Gamma(p,q) \Phi \rangle \tag{3.15}$$

is a faithful, normal, nondegenerate state on $\mathscr{W}_\Gamma$.

(ii) From the discussion of the modular algebraic structure of $\mathscr{W}_\Gamma$ in Ref. 1, as outlined above, it follows that every faithful normal, nondegenerate state φ on $\mathscr{W}_\Gamma$ can be obtained from the following prescription: find a basis (3.9) satisfying (2.54); construct the resolution generators $\xi_n \in \mathscr{H}_n$ using (3.10) and (2.38); and define Φ and φ as in (3.13) and (3.15). In this connection, we note also that if ρ_φ is the density matrix on $\mathscr{H}_s$, for which

$$\langle \varphi; \mathscr{W}_\Gamma(p,q) \rangle = \mathrm{Tr}\{\rho_\varphi \mathscr{W}_s(p,q)\}, \tag{3.16}$$

then

$$\rho_\varphi = (2\pi\hbar) \sum_{n \in \mathbf{Z}^+} \lambda_n |\tilde{\xi}_n\rangle\langle\tilde{\xi}_n\rangle|. \tag{3.17}$$

(iii) Corresponding to a given Φ, the von Neumann algebra $\mathscr{M}_\varphi \subset \mathscr{W}_\Gamma$ generated by the projectors

$$P_n = (2\pi\hbar) |\xi_n\rangle\langle\xi_n| \tag{3.18}$$

is atomic and maximal Abelian, and it is the centralizer of $\mathscr{W}_\Gamma$ with respect to φ. Thus, the nonuniqueness of the decomposition (2.55) and (2.56) (and hence of the choice of polarization) is due to the availability of different atomic, maximal Abelian von Neumann subalgebras of $\mathscr{W}_\Gamma$. Conversely, in view of Proposition 3 in Ref. 1, each atomic, maximal Abelian von Neumann subalgebra of $\mathscr{W}_\Gamma$ determines a decomposition of the type (2.55) and (2.56). It also follows from this proposition that every irreducible subrepresentation of $\mathscr{W}_\Gamma$ is of the type $\mathscr{W}_\xi$, i.e., is obtained [cf. (2.18)–(2.20)] as the restriction of $\mathscr{W}_\Gamma$ to a stable subspace of the type $\mathscr{H}_\xi$, which is thus a reproducing kernel Hilbert space in $\mathscr{H}_\Gamma$.

(iv) The mapping S_Φ in (3.2) and (3.3) can be explicitly computed in the context of the decomposition of $\mathscr{W}_\Gamma$ given by $\{\tilde{\xi}_n | n \in \mathbf{Z}^+\}$ and Φ. Indeed, the set $\{\Phi^{p,q} | (p,q) \in T^*\mathbf{R}\}$, with $\Phi^{p,q}$ defined in (3.14), is total in $\mathscr{H}_\Gamma$, and since

$$S_\Phi \mathscr{W}_\Gamma(p,q)\Phi = \mathscr{W}_\Gamma(p,q)^*\Phi, \tag{3.19}$$

we obtain, for every $(p,q) \in T^*\mathbf{R}$,

$$S_\Phi \Phi^{p,q} = \exp\{(i/\hbar)pq\} \, \Phi^{-p,-q}, \tag{3.20}$$

and, for every pair $\{(p,q),(p',q')\} \in T^*\mathbf{R} \times T^*\mathbf{R}$

$$\langle \Phi^{p',q'} | S_\Phi^* \Phi^{p,q} \rangle = \exp\{(i/\hbar)p'q'\} \langle \Phi^{-p',-q'} | \Phi^{p,q} \rangle, \tag{3.21}$$

so that the closability of S_Φ is transparent in the particular case at hand. We note, next, that for each $m \in \mathbf{Z}^+$ the set of vectors

$$\{\xi_{mn} \equiv V_m \tilde{\xi}_n | n \in \mathbf{Z}^+\} \tag{3.22}$$

forms a basis in $\mathscr{H}_m$, with

$$\langle \xi_{mk} | \xi_{ml} \rangle = (2\pi\hbar)^{-1} \delta_{kl}. \tag{3.23}$$

Also, as vectors in $\mathscr{H}_m$

$$\begin{aligned} \xi_{mn}(p,q) &= \langle \tilde{\xi}_m^{p,q} | \tilde{\xi}_n \rangle \\ &= \exp\{(i/\hbar)pq\} \xi_{nm}(-p,-q)^*. \end{aligned} \tag{3.24}$$

Using these preliminary results, one computes

$$J_\Phi \xi_{mn} = \xi_{nm}, \quad \forall n, m \in \mathbf{Z}^+, \tag{3.25}$$

and for all $\Psi \in \mathscr{H}_\Gamma$,

$$(J_\Phi \Psi)(p,q) = \exp\{(i/\hbar)pq\} \Psi(-p,-q)^*. \tag{3.26}$$

Furthermore

$$\Delta_\Phi \xi_{nm} = \lambda_n \lambda_m^{-1} \xi_{nm}, \tag{3.27}$$

and J_Φ transforms the operators P_Γ and Q_Γ in (2.4) into the operators $\widehat{P}$ and $\widehat{Q}$ in (3.1); specifically

$$\begin{aligned} J_\Phi P_\Gamma J_\Phi &= -i\hbar \frac{\partial}{\partial q} - p = \widehat{P}, \\ J_\Phi Q_\Gamma J_\Phi &= i\hbar \frac{\partial}{\partial p} = \widehat{Q}, \end{aligned} \tag{3.28}$$

to be compared with

$$J_\Phi \mathscr{W}_\Gamma J_\Phi = \mathscr{W}'_\Gamma. \tag{3.29}$$

Note that the commutation relations satisfied by $\widehat{P}$ and $\widehat{Q}$, namely

$$[\widehat{P}, \widehat{Q}] = +i\hbar I_\Gamma \tag{3.30}$$

differ by a sign from the CCR in (2.6), as is to be expected from the fact that J_Φ is an (involutive) *antiunitary* operator. This makes explicit the remark opening this section.

IV. GEOMETRIC QUANTIZATION REVISITED

In view of the analysis carried out in the previous two sections, the geometric quantization program can be described as follows.

(i) The primary object of the theory is the configuration space manifold M. In this paper, M was taken to be a vector space; this circumstance does simplify the presentation of the arguments; these, however, extend to the case where M is a homogeneous manifold, see, e.g., Ref. 14 for the setting of the problem in this more general framework. Furthermore, the quantization procedure outlined in (iv) below generalizes as well to situations where Γ is not necessarily a cotangent bundle, see, e.g., Ref. 2. The cotangent bundle $\Gamma = T^*M$ equipped with its canonical symplectic form ω, and the associated Poisson bracket $\{\cdots,\cdots\}$, supports the Hamiltonian formulation of classical mechanics. In particular, the elements of the C^*-algebra $C^\infty(\Gamma)$, of complex, bounded, infinitely differentiable functions on Γ, are interpreted as the classical observables of the theory.

(ii) The space T^*M is naturally equipped with a measure ν, so that $H_\Gamma = L^2(T^*M, d\nu)$ is a canonical object of the theory, which Koopman[8] already took as the starting point of his Hilbert space formulation of classical Hamiltonian mechanics. This space is also taken as the starting point for the quantization procedures discussed in this paper.

(iii) In addition to the *Abelian* involutive algebra structure defined on $C^\infty(\Gamma)$ by pointwise addition, multiplication, and complex conjugation, the Poisson bracket equips $C^\infty(\Gamma)$ with a Lie structure. This structure is captured in the composition laws of transformations of H_Γ, defining the prequantization map,[3–5] that associates to every $f\in C^\infty(\Gamma)$, the first-order differential operator

$$P(f) = -i\hbar\nabla_{X_f} + f, \tag{4.1}$$

where ∇ denotes the covariant derivative, and X_f is the vector field canonically associated to f by the symplectic form ω on T^*M:

$$X_f \lrcorner\, \omega = -df. \tag{4.2}$$

The remarkable property of the prequantization map is that it satisfies the following requirements of the Dirac problem:

$$\mathscr{P}(\alpha f + \beta g) = \alpha\mathscr{P}(f) + \beta\mathscr{P}(g), \tag{4.3}$$

$$\mathscr{P}(1) = I_\Gamma, \tag{4.4}$$

$$\mathscr{P}(\{f,g\}) = [\mathscr{P}(f), \mathscr{P}(g)]/i\hbar, \tag{4.5}$$

for all $f, g\in C^\infty(\Gamma)$, and all $\alpha, \beta\in\mathbb{C}$. In (4.4), 1 denotes the constant function with value $1\in\mathbb{R}$, and I_Γ denotes the identity operator on $\mathscr{H}_\Gamma$; in (4.5), $\{f,g\}$ is the Poisson bracket between f and g, and $[A,B]$ is the commutator $AB-BA$ of the operators A, B on $\mathscr{H}_\Gamma$. However, in contradiction with the operators P_s and Q_s [cf. (2.8)] of the usual Schrödinger formulation of quantum mechanics the operators $\mathscr{P}(p)$ and $\mathscr{P}(q)$ [obtained by natural extension of (4.1)] do *not* generate an irreducible algebra; it is in fact impossible (see, for example, Ref. 14 for a streamlined proof) to find a map that would satisfy this last requirement, together with (4.3)–(4.5). Moreover, the usual (Jordan) algebra structure of $C^\infty(\Gamma)$, provided by the composition law $f\cdot g$, given by pointwise multiplication, does *not* carry over to the ordinary composition of transformations of $\mathscr{H}_\Gamma$: the composition of two first-order differential operators is not, usually, a first-order differential operator! This is precisely where the Hilbert space formulations of classical and quantum mechanics differ, and it is also where our quantization procedure differs from the usual geometric quantization procedure: we focus on the operators $P_\Gamma = \mathscr{P}(p)$ and $Q_\Gamma = \mathscr{P}(q)$, which satisfy (2.4) and (2.6)/(4.5), and we define from them, using the usual composition laws of transformations on $\mathscr{H}_\Gamma$, the Weyl algebra

$$\mathscr{W}_\Gamma \equiv \{\mathscr{W}_\Gamma(p,q)\,|\,(p,q)\in T^*\mathbb{R}\}'',$$

with $\mathscr{W}_\Gamma(p,q)$ defined in (2.1). Note that $\mathscr{W}_\Gamma$ is isomorphic, as a von Neumann algebra, to

$$\mathscr{W}_s = \{\mathscr{W}_s(p,q)\,|\,(p,q)\in T^*\mathbb{R}\}''$$

with $\mathscr{W}_s(p,q)$ defined in (2.7); we denote the abstract W^*-algebra $\mathscr{W}_\Gamma \simeq \mathscr{W}_s$ by $\mathscr{W}$. Consequently, all predictions made from a formalism that takes $\mathscr{W}_\Gamma$ as its basic object will be the same as those made from the usual Schrödinger representation $\mathscr{W}_s$ of the CCR. The question therefore is to determine whether the choice of $\mathscr{W}_s$ rather than $\mathscr{W}_\Gamma$ is only a historical accident, or whether $\mathscr{W}_s$ can be selected out on physical grounds. Mathematically, $\mathscr{W}_\Gamma$ and $\mathscr{W}_s$ are both obtained as GNS representations of the CCR; $\mathscr{W}_\Gamma$, however, arises from faithful normal states on $\mathscr{W}$, while $\mathscr{W}_s$ arises from pure normal states on this W^*-algebra of quantum observables. Physically, rather than giving a prescription to extract $\mathscr{W}_s$ from $\mathscr{P}(C^\infty(\Gamma))$, we choose to give a prescription to extract $\mathscr{W}_s$ from $\mathscr{W}_\Gamma$, the latter being obtained from $\mathscr{P}(C^\infty(\Gamma))$ as just indicated. We further show [cf. (2.27)–(2.32)] how the resulting operators P_s and Q_s are linked to the Koopman operators $P^{\rm cl}$ and $Q^{\rm cl}$.

(iv) The quantization procedure that emerges, and that we consider to be a reinterpretation of the geometric quantization program, will be shown to provide a way to discriminate between various "correspondence principles"—or "ordering rules"—on a physical basis, namely quantum measurement processes. It is first to be described mathematically as follows.

(a) Realize the classical algebra of observables $C^\infty(\Gamma)$ as multiplication operators $f^{\rm cl}$ on $\mathscr{H}_\Gamma$,

$$(f^{\rm cl}\Psi)(p,q) = f(p,q)\Psi(p,q), \tag{4.6}$$

for every $f\in C^\infty(\Gamma)$ and all $\Psi\in\mathscr{H}_\Gamma$.

(b) Identify, in $\mathscr{H}_\Gamma$, a reproducing kernel Hilbert subspace $\mathscr{H}_K = \mathbb{P}_K\mathscr{H}_\Gamma$ (denoted $\mathscr{H}_\xi$ in Sec. II), with kernel K and projector $\mathbb{P}_K$, and having the further property that $K(\zeta,\zeta')$ is separately continuous in ζ and $\zeta'\in T^*M$.

(c) Quantize, using the linear map

$$\pi_K^*: f\in C^\infty(\Gamma)\mapsto \hat f_K \equiv \mathbb{P}_K f^{\rm cl}\mathbb{P}_K \in \mathscr{L}(\mathscr{H}_K). \tag{4.7}$$

Associated to $\mathscr{H}_K$, there is a POV measure a_K, defined on the Borel sets Δ of Γ, and it can be shown[2] that

$$\hat f_K = \int_\Gamma f(\zeta)\,da_K(\zeta). \tag{4.8}$$

From the general theory of reproducing kernels,[9–12] it follows that the set of vectors

$$\mathscr{S}_K \equiv \{\xi_K^\zeta\in\mathscr{H}_K\,|\,\zeta\in\Gamma,\ \xi_K^\zeta(\zeta')\equiv K(\zeta',\zeta)\} \tag{4.9}$$

[cf. (2.53)] is overcomplete in $\mathscr{H}_K$, and, in fact,

$$a_K(\Delta) = \int_\Delta d\nu(\zeta)\,|\xi_K^\zeta\rangle\langle\xi_K^\zeta|. \tag{4.10}$$

We call the vectors $\xi_K^\zeta\in\mathscr{S}_K$ generalized coherent states, and indeed, as noted earlier, in the case at hand, vectors of the type (2.38), which enter into the definition (2.43) of K_ξ, are coherent states in the sense of Klauder[9] and Perelomov.[10] Since the resolution generator ξ in (2.38) determines all the coherent states $\xi^{p,q}$, we denote the kernel by K_ξ in this case, and see now clearly the crucial role played by the coherent states in the quantization procedure. Indeed a choice of K amounts to a choice of ξ, and hence to a determination of the family of coherent states $\mathscr{S}_K$, and vice versa.

We further want to draw attention to the following additional features of the quantization procedure outlined in (a)–(c) above. First [cf. (2.31)],

$$[\pi_K^*(p), \pi_K^*(q)] = -i\hbar\pi_K^*(1),$$
$$\text{with } \pi_K^*(1) = I_K, \tag{4.11}$$

where I_K is the identity operator on $\mathscr{H}_K$; in the above expression, the domain of π_K^* has been straightforwardly extended beyond $C^\infty(\Gamma)$ using (4.8). Second, going back to the classical operators $P^{\rm cl}$ and $Q^{\rm cl}$ [in (2.25)] and to the prequantized operators P_Γ and Q_Γ [in (2.4)], when the reproducing kernel K, with generator ξ, is chosen (cf. the discussion at the end of Sec. II) so that $\langle P\rangle_\xi = 0 = \langle Q\rangle_\xi$, we have

$$\mathbb{P}_K P^{\rm cl}\mathbb{P}_K = \mathbb{P}_K P_\Gamma \mathbb{P}_K, \quad \mathbb{P}_K Q^{\rm cl}\mathbb{P}_K = \mathbb{P}_K Q_\Gamma \mathbb{P}_K. \tag{4.12}$$

This demonstrates explicitly the sense in which a choice of K implies a choice of polarization.

(v) A last comment on the physical implication of the choice of the generator ξ of K, and hence of the polarization, ought to be made here. For an arbitrary classical observable f, its quantized form $\mathbb{P}_\xi f^{\rm cl}\mathbb{P}_\xi$ involves a specific ordering of the operators P_ξ and Q_ξ. To see this more clearly (cf. Ref. 2 for a detailed discussion, for the realization of specific orderings by "closed form" integral transforms, and for references to the extensive literature on the ordering problem), let f be a finite-degree polynomial

$$f(p,q) = \sum_{m,n} c_{mn} p^m q^n, \tag{4.13}$$

where c_{mn} are constants. Then

$$\mathbb{P}_\xi f^{\rm cl}\mathbb{P}_\xi = \sum_{m,n} c_{m,n}\widehat{G}_\xi^{m,n}(P_\xi, Q_\xi), \tag{4.14}$$

where [cf. (4.8)]

$$\widehat{G}_\xi^{m,n}(P_\xi, Q_\xi) = \int_\Gamma p^m q^n \, da_\xi(p,q), \tag{4.15}$$

is a polynomial of highest degree m in P_ξ and n in Q_ξ. The order in which the operators P_ξ and Q_ξ appear in $\widehat{G}_\xi^{m,n}(P_\xi, Q_\xi)$ depends on the POV measure a_ξ, and hence on the resolution generator ξ. *For instance*, if ξ is taken to be the ground state wave function of the harmonic oscillator, then the quantized operators (4.14) are antinormally ordered, i.e., each monomial of the form

$$(z^*)^m z^n \quad \text{with} \quad z = (2\hbar)^{-1/2}(q + ip) \tag{4.16}$$

is mapped to the operator monomial

$$(\hat{a}^*)^m \hat{a}^n \quad \text{with} \quad \hat{a} = (2\hbar)^{-1/2}(Q_\xi + iP_\xi). \tag{4.17}$$

As a further observation, we point out that a specific ordering is a reflection of the measuring apparatus used in a joint determination of position and momentum of the physical system. If the wave function of the state of the system is $\Psi_\xi \in \mathscr{H}_\xi$, then[2]

$$\int_{\mathbb{R}} dp |\Psi_\xi(p,q)|^2 = \int_{\mathbb{R}} dq' \, \chi_q(q')|\widetilde{\Psi}(q')|^2 \tag{4.18}$$

and

$$\int_{\mathbb{R}} dq |\Psi_\xi(p,q)|^2 = \int_{\mathbb{R}} dp' \, \widehat{\chi}_p(p')|\widehat{\widetilde{\Psi}}(p')|^2, \tag{4.19}$$

where $\widetilde{\Psi} = V_\xi^{-1}\Psi_\xi \in \mathscr{H}_s$, $\widehat{\widetilde{\Psi}}$ is its Fourier transform, and

$$\chi_q(q') = \chi_0(q'-q) = \hbar^{-1}|\widetilde{\xi}(q'-q)|^2, \tag{4.20}$$

$$\widehat{\chi}_p(p') = \widehat{\chi}_0(p'-p) = \hbar^{-1}|\widehat{\widetilde{\xi}}(p'-p)|^2, \tag{4.21}$$

$\widehat{\widetilde{\xi}}$ being the Fourier transform of $\widetilde{\xi} \in \mathscr{H}_s$. These equations show that if we interpret $|\Psi_\xi(p,q)|^2$ as the joint probability density for locating the system at the point $(p,q)\in\Gamma$, our apparatus allows us to measure it only up to an imprecision given by the probability distribution $\widehat{\chi}_p$ in momentum, and the probability distribution χ_q in position. Furthermore, consistent with the uncertainty principle, the product of the standard deviations satisfies

$$\sigma(\widehat{\chi}_p)\cdot\sigma(\chi_q) \geqslant \hbar/2. \tag{4.22}$$

In the equivalent quantization procedure, as it now emerges from Ref. 1, the above discussion translates into the statement that the choice of a maximal Abelian, atomic von Neumann subalgebra $\mathscr{A}$ of $\mathscr{W}_\Gamma$ corresponds to the choice of a class of orderings, one for each ξ_n appearing in (3.13). Choosing a specific ξ_n (see Proposition 3 in Ref. 1) to get one irreducible subrepresentation $\mathscr{W}_n$, from the decomposition of $\mathscr{W}_\Gamma$ associated to $\mathscr{A}$, amounts then to selecting one specific ordering, and thus one specific classical measuring apparatus. Note that if the system has been prepared in a faithful, normal, nondegenerate state φ, then the choice of the maximal Abelian, atomic von Neumann subalgebra $\mathscr{A}$ of $\mathscr{W}$ is uniquely determined by the requirement that $\varphi_{vN}^{\mathscr{A}} = \varphi$, where $\varphi_{vN}^{\mathscr{A}}$ is the state obtained from φ by the finest partitioning compatible with $\mathscr{A}$, defined in accordance with Ref. 15 by

$$\varphi_{vN}^{\mathscr{A}} = \sum_{n\in\mathbb{Z}^+} \lambda_n \varphi_n, \tag{4.23}$$

with

$$\lambda_n \equiv \langle\varphi; F_n\rangle, \quad \varphi_n\colon \ W\in\mathscr{W} \mapsto \frac{\langle\varphi; F_n W F_n\rangle}{\langle\varphi; F_n\rangle} \in \mathbb{C}, \tag{4.24}$$

and $\{F_n \mid n\in\mathbb{Z}^+\}$ is the partition of the unit $I\in\mathscr{W}$ into minimal projectors (i.e., atoms) of $\mathscr{A}$. Indeed, the requirement $\varphi_{vN}^{\mathscr{A}} = \varphi$ is equivalent[16] to having $\mathscr{A} = \mathscr{M}_\varphi$, the centralizer of $\mathscr{W}$ with respect to φ defined in (3.7). As opposed to a von Neumann reduction $\varphi \to \varphi_{vN}^{\mathscr{A}}$, a measurement is the subsequent filtering out of all but one φ_n, or equivalently one F_n, and thus here one ξ_n. Hence, the polarization is indeed determined by the physical situation at hand.

ACKNOWLEDGMENTS

One of us (STA) was supported in part by the Natural Science and Engineering Research Council of Canada. The other (GGE) was supported in part by the Akademie der Wissenschaften zu Göttingen.

[1] G. G. Emch, Int. J. Theor. Phys. 20, 891 (1981).
[2] S. T. Ali, Riv. Nuovo Cimento 8 (11), 1 (1985).
[3] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, Reading, MA, 1978), 2nd ed.
[4] J. Sniatycki, *Geometric Quantization and Quantum Mechanics* (Springer, Berlin, 1979).
[5] D. J. Simms and N. M. J. Woodhouse, *Lectures on Geometric Quantization* (Springer, Berlin, 1976).
[6] M. Takesaki, *Tomita's Theory of Modular Hilbert Algebras and its Applications* (Springer, New York, 1970).
[7] S. T. Ali and E. Prugovečki, Acta Appl. Math. 6, 19 (1986).
[8] B. O. Koopman, Proc. Natl. Acad. Sci. USA 17, 315 (1931).

[9]J. R. Klauder, J. Math. Phys. **4**, 1055 (1963).
[10]A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[11]N. Aronszajn, Trans. Am. Math. Soc. **68**, 337 (1950).
[12]U. Cattaneo, J. Math. Phys. **23**, 659 (1982).
[13]R. V. Kadison and J. R. Penrose, *Fundamentals of the Theory of Operator Algebras* (Academic, New York, 1983).
[14]G. G. Emch, J. Math. Phys. **23**, 1785 (1982).
[15]J. von Neumann, *Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).
[16]G. G. Emch, Commun. Math. Phys. **49**, 191 (1976).

Coulomb Green's functions, in an n-dimensional Euclidean space

L. Chetouani and T. F. Hammann
Faculté des Sciences et Techniques, Université de Haute Alsace, 4, rue des Frères Lumière, 68093 Mulhouse Cédex, France

(Received 25 March 1986; accepted for publication 6 August 1986)

The H-atom Green's function is calculated in an n-dimensional Euclidean space, following the Feynman Lagrangian formulation. The use of generalized polar coordinates allows the expansion of the propagator into partial propagators, and the separation of the angular and radial variables. The angular part is shown to be a generalized Legendre polynomial while the radial part may be transformed in that of the harmonic oscillator. The H-atom spectrum is given by the poles of the Green's function.

I. INTRODUCTION

The Feynman functional integrals are used to obtain the Green's function of an H-atom in an Euclidean space having dimension $n = p + 2$ ($n \geqslant 2$). Within an n-dimensional polar coordinate system, we first split the propagator into partial propagators, in order to separate the radial part from the angular one. The angular part merely is a generalization of the Legendre polynomials, and by means of a variable change, the radial part can be reduced into that of the harmonic oscillator in a centrifugal potential. The H-atom spectrum is obtained from the poles of the Green's function. The case $n = 1$ is deduced from the general case. The present paper gives, to the best of our knowledge, the first path integral derivation of the n-dimensional Coulomb Green's function.

II. PARTIAL PROPAGATOR DECOMPOSITION OF THE n-DIMENSIONAL PROPAGATOR OF A CENTRAL POTENTIAL $V(r) = -\alpha/r$

Consider a particle, having the mass m, and moving in an n-dimensional Euclidean space, in an n-dimensional central potential $V(r) = -\alpha/r$, where r stands for the n-dimensional radius. Let $n = p + 2 \geqslant 2$. The Cartesian coordinates $(x^{(1)}, x^{(2)}, \ldots, x^{(p+2)})$, are related to the polar coordinates $(r, \theta, \phi^{(1)}, \ldots, \phi^{(p)})$, by means of the following transformation:

$$
\begin{aligned}
x^{(1)} &= r \cos \theta, \\
x^{(2)} &= r \sin \theta \cos \phi^{(1)}, \\
x^{(3)} &= r \sin \theta \sin \phi^{(1)} \cos \phi^{(2)}, \\
&\vdots \\
x^{(p+1)} &= r \sin \theta \sin \phi^{(1)} \sin \phi^{(2)} \cdots \sin \phi^{(p-1)} \cos \phi^{(p)}, \\
x^{(p+2)} &= r \sin \theta \sin \phi^{(1)} \sin \phi^{(2)} \cdots \sin \phi^{(p-1)} \sin \phi^{(p)},
\end{aligned} \tag{1}
$$

with

$$0 < r < \infty, \quad 0 \leqslant \theta \leqslant \pi,$$

$$0 \leqslant \phi^{(i)} \leqslant \pi, \quad i = 1,2,\ldots,p-1, \quad 0 \leqslant \phi^{(p)} \leqslant 2\pi,$$

and

$$r^2 = \Sigma_{i=1}^{p+2} x^{(i)2}.$$

It is readily shown that the Jacobian of this transformation is

$$
\begin{aligned}
\mathrm{J} &= \frac{\partial(x^{(1)}, x^{(2)}, \ldots, x^{(p+2)})}{\partial(r, \theta, \phi^{(1)}, \ldots, \phi^{(p)})} \\
&= r^{p+1} \sin^p \theta \sin^{p-1} \phi^{(1)} \sin^{p-2} \phi^{(2)} \cdots \sin \phi^{(p-1)}.
\end{aligned}
$$

In Feynman's Lagrangian formulation, the kernel is given, in Cartesian coordinates, by the following path integral[1]:

$$
\begin{aligned}
K(\mathbf{r}_f, \mathbf{r}_i; T) &= \int \mathscr{D}x^{(1)}(t) \mathscr{D}x^{(2)}(t) \cdots \mathscr{D}x^{(p+2)}(t) \exp\left\{\frac{i}{\hbar} \int_0^T \left(\frac{m}{2} \dot{\mathbf{r}}^2 - V(r)\right) dt\right\} \\
&= \lim_{N \to \infty} \left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N(p+2/2)} \int \prod_{j=1}^{N-1} dx_j^{(1)} dx_j^{(2)} \cdots dx_j^{(p+2)} \exp\left\{\frac{i}{\hbar} \sum_{j=1}^{N} \frac{m}{2\epsilon} [(x_j^{(1)} - x_{j-1}^{(1)})^2 \right. \\
&\quad \left. + (x_j^{(2)} - x_{j-1}^{(2)})^2 + \cdots + (x_j^{(p+2)} - x_{j-1}^{(p+2)})^2] - \epsilon V(r_j)\right\}.
\end{aligned}
$$

In polar coordinates, this propagator is written as follows:

$$
\begin{aligned}
K(\mathbf{r}_f, \mathbf{r}_i; T) = \lim_{N \to \infty} &\left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N(p+2)/2} \int \prod_{j=1}^{N-1} (dr_j\, r_j^{p+1})(d\theta_j \sin^p \theta_j)(d\phi_j^{(1)} \sin^{p-1} \phi_j^{(1)}) \\
&\times (d\phi_j^{(2)} \sin^{p-2} \phi_j^{(2)}) \cdots (d\phi_j^{(p-1)} \sin^{(p-1)} \phi_j)(d\phi_j^{(p)}) \\
&\times \exp\left\{\frac{i}{\hbar} \sum_{j=1}^{N} \left[\frac{m}{2\epsilon}(r_j^2 + r_{j-1}^2 - 2 r_j r_{j-1} \cos \Theta_{j,j-1})\right] - \epsilon V(r_j)\right\},
\end{aligned} \tag{2}
$$

 0022-2488/86/122944-05$02.50

where

$$\mathbf{r}_f=\mathbf{r}(t_f)=(r_f,\theta_f,\phi_f^{(1)},...,\phi_f^{(p)}),\quad \mathbf{r}_i=\mathbf{r}(t_i)=(r_i,\theta_i,\phi_i^{(1)},...,\phi_i^{(p)}),$$
$$y_j=y(t_j),\quad y=(r,\theta,\phi^{(1)},...,\phi^{(p)}),\quad T=N\epsilon=N(t_j-t_{j-1})=t_f-t_i,$$

and

$$\begin{aligned}
\cos\Theta_{j,j-1}&=\frac{\mathbf{r}_j\cdot\mathbf{r}_{j-1}}{r_j\cdot r_{j-1}}=\cos\theta_j\cos\theta_{j-1}+\sin\theta_j\sin\theta_{j-1}\cos\gamma_{j,j-1}^{(1)},\\
\cos\gamma_{j,j-1}^{(1)}&=\cos\phi_j^{(1)}\cos\phi_{j-1}^{(1)}+\sin\phi_j^{(1)}\sin\phi_{j-1}^{(1)}\cos\gamma_{j,j-1}^{(2)},\\
&\vdots\\
\cos\gamma_{j,j-1}^{(r)}&=\cos\phi_j^{(r)}\cos\phi_{j-1}^{(r)}+\sin\phi_j^{(r)}\sin\phi_{j-1}^{(r)}\cos\gamma_{j,j-1}^{(r+1)},\\
&\vdots\qquad\qquad (1\leqslant r\leqslant p-2),\\
\cos\gamma_{j,j-1}^{(p-1)}&=\cos\phi_j^{(p-1)}\cos\phi_{j-1}^{(p-1)}+\sin\phi_j^{(p-1)}\sin\phi_j^{(p-1)}\cos(\phi_j^{(p)}-\phi_{j-1}^{(p)}).
\end{aligned}\tag{3}$$

However, this expression (2) is not suitable for integrating, because of the occurrence of the mixed term $-(i/\hbar)(m/\epsilon)r_jr_{j-1}\cos\Theta_{j,j-1}$ in the action. However, by means of the formula[2,3]

$$\exp(u\cos\Theta)=(2/u)^\nu\Gamma(\nu)\sum_{l=0}^{\infty}(l+\nu)I_{l+\nu}(u)C_l^\nu(\cos\Theta),\quad \nu\neq 0,-1,-2,...,$$

where the $I_{l+\nu}(u)$ are the modified Bessel functions[4] and the $C_l^\nu(\cos\Theta)$ are the Gegenbauer polynomials[5] that generalize the Legendre polynomials, the expression (2) is split into a radial part and an angular part:

$$\begin{aligned}
K(\mathbf{r}_f,\mathbf{r}_i;T)&=\lim_{N\to\infty}\left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N(p+2)/2}\int\prod_{j=1}^{N-1}(dr_j\,r_j^{p+1})(d\theta_j\sin^p\theta_j)\cdot(d\phi_j^{(1)}\sin^{(p-1)}\phi_j^{(1)})\cdots(d\phi^{(p-1)}\sin\phi_j^{(p-1)})(d\phi_j^{(p)})\\
&\quad\times\exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left[\frac{m}{2\epsilon}(r_j^2+r_{j-1}^2)-\epsilon V(r_j)\right]\right\}\prod_{j=1}^{N}\left[\left(\frac{2i\hbar\epsilon}{mr_jr_{j-1}}\right)^\nu\Gamma(\nu)\right.\\
&\quad\left.\times\sum_{l_j=0}^{\infty}(l_j+\nu)\mathrm{I}_{l_j+\nu}\left(\frac{mr_jr_{j-1}}{i\hbar\epsilon}\right)C_{l_j}^\nu(\cos\Theta_{j,j-1})\right]\\
&=\lim_{N\to\infty}\left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N(p+2)/2}\sum_{l_1,l_2,...,l_N=0}^{\infty}\int\prod_{j=1}^{N-1}(dr_jr_j^{p+1})\\
&\quad\times\exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left[\frac{m}{2\epsilon}(r_j^2+r_{j-1}^2)-\epsilon V(r_j)\right]\right\}\prod_{j=1}^{N}I_{l_j+\nu}\left(\frac{mr_jr_{j-1}}{i\hbar\epsilon}\right)\left(\frac{2i\hbar\epsilon}{mr_jr_{j-1}}\right)^\nu\\
&\quad\times\int\prod_{j=1}^{N-1}(d\theta_j\sin^p\theta_j)(d\phi_j^{(1)}\sin^{p-1}\phi_j^{(1)})\cdots(d\phi_j^{(p-1)}\sin\phi_j^{(p-1)})d\phi_j^{(p)}\prod_{j=1}^{N}\Gamma(\nu)(l_j+\nu)C_{l_j}^\nu(\cos\Theta_{j,j-1}).
\end{aligned}\tag{4}$$

Let

$$\overline{C}_n^\nu(\cos\Theta_{j,j-1})=\Gamma(\nu)(n+\nu)C_n^\nu(\cos\Theta_{j,j-1}),\ \widetilde{C}_n^\nu(\cos\alpha)=\left(\frac{n!(n+\nu)2^{2\nu-1}}{\pi\Gamma(2\nu+n)}\right)^{1/2}\Gamma(\nu)C_n^\nu(\cos\alpha).$$

In order to integrate over the angular part, we give to ν the fixed value $p/2$, and we separate all the $\theta_j,\phi_j^{(1)},...,\phi_j^{(p-2)},\phi_j^{(p-1)},\phi_j^{(p)}$ integrals, by applying the addition theorem[6] to the relation (3):

$$\begin{aligned}
&\overline{C}_n^{p/2}(\cos\Theta_{j,j-1})\\
&=2\pi\cdot\pi^{p/2}\sum_{k_1=0}^{n}\sum_{k_2=0}^{k_1}\cdots\sum_{k_{p-1}=0}^{k_{p-2}}\sum_{m=-k_{p-1}}^{k_{p-1}}(\sin\theta_j\sin\theta_{j-1})^{k_1}(\sin\phi_j^{(1)}\sin\phi_{j-1}^{(1)})^{k_2}\cdots(\sin\phi_j^{(p-2)}\sin\phi_{j-1}^{(p-2)})^{k_{p-1}}\\
&\quad\times\widetilde{C}_{n-k_1}^{p/2+k_1}(\cos\theta_j)\widetilde{C}_{n-k_1}^{p/2+k_1}(\cos\theta_{j-1})\widetilde{C}_{k_1-k_2}^{(p-1)/2+k_2}(\cos\phi_j^{(1)})\\
&\quad\times\widetilde{C}_{k_1-k_2}^{(p-1)/2+k_2}(\cos\phi_{j-1}^{(1)})\cdots\widetilde{C}_{k_{p-2}-k_{p-1}}^{1+k_{p-1}}(\cos\phi_j^{(p-2)})\,\widetilde{C}_{k_{p-2}-k_{p-1}}^{1+k_{p-1}}(\cos\phi_{j-1}^{(p-2)})\\
&\quad\times Y_{k_{p-1}}^{m*}(\phi_j^{(p-1)},\phi_j^{(p)})Y_{k_{p-1}}^{m}(\phi_{j-1}^{(p-1)},\phi_{j-1}^{(p)}).
\end{aligned}$$

The orthogonality of the $\widetilde{C}_m^\nu$ functions, and of the spherical harmonics $Y_l^m(\Omega)$,

$$\int_0^\pi d\alpha\sin^{2\nu}\alpha\widetilde{C}_n^\nu(\cos\alpha)\widetilde{C}_m^\nu(\cos\alpha)=\delta_{n,m},\quad\int Y_l^{m*}(\Omega)Y_{l'}^{m'}(\Omega)d\Omega=\delta_{ll'}\delta_{mm'},$$

allow us to calculate the angular part of the propagator (4):

$$\int \prod_{j=1}^{N-1} (d\theta_j \sin^p \theta_j)(d\phi_j^{(1)} \sin^{p-1} \phi_j^{(1)}) \cdots d\Omega_j \prod_{j=1}^{N} \overline{C}_{l_j}^{p/2}(\cos \Theta_{j,j-1}) = (2\pi\pi^{p/2})^{N-1} \overline{C}_l^{p/2}(\cos \Theta_{0,N}) \delta_{l,l_1} \prod_{j=2}^{N} \delta_{l_j,l_{j-1}},$$

with $\Theta_{0,N} = (\mathbf{r}_0,\mathbf{r}_N) = (\mathbf{r}_i,\mathbf{r}_f)$. The propagator (4) is then expanded into partial kernels:

$$K(\mathbf{r}_f,\mathbf{r}_i;T) = \sum_{l=0}^{\infty} K_l(r_f,r_i;T) \frac{(2l+n-2)}{4(\pi)^{n/2}} \Gamma\left(\frac{n}{2}-1\right) C_l^{n/2-1}(\cos \Theta_{i,f}), \tag{5}$$

where

$$K_l(r_f,r_i;T) = \lim_{N\to\infty} \left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N(p+2)/2} (2^{p+1}\pi)^N \int \prod_{j=1}^{N-1} dr_j\, r_j^{p+1}$$
$$\times \left[\prod_{j=1}^{N} \left(\frac{i\hbar\epsilon\pi}{2mr_jr_{j-1}}\right)^{1/2} \mathrm{I}_{l+p/2}\left(\frac{mr_jr_{j-1}}{i\hbar\epsilon}\right)\right] \exp\left[\frac{i}{\hbar}\sum_{j=1}^{N}\left(\frac{m}{2\epsilon}(r_j^2+r_{j-1}^2) - \epsilon V(r_j)\right)\right]. \tag{6}$$

The radial part in Eq. (6) can be simplified further, by taking into account the asymptotic behavior[7] of the modified Bessel functions:

$$I_\nu\left(\frac{u}{\epsilon}\right) \underset{\epsilon\to 0}{\to} \left(\frac{\epsilon}{2\pi u}\right)^{\frac{1}{2}} \exp\left\{\frac{u}{\epsilon} - \frac{1}{2}\frac{\epsilon}{u}\left(\nu^2 - \frac{1}{4}\right)\right\}.$$

This radial part becomes

$$K_l(r_f,r_i;T)$$
$$= \left(\frac{1}{r_fr_i}\right)^{(p+1)/2} \lim_{N\to\infty} \left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N/2} \int \prod_{j=1}^{N-1} dr_j \exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left[\frac{m}{2\epsilon}(r_j-r_{j-1})^2 - \frac{\hbar^2\epsilon}{2mr_jr_{j-1}}\left(\left(l+\frac{p}{2}\right)^2 - \frac{1}{4}\right) - \epsilon V(r_j)\right]\right\}$$
$$= \left(\frac{1}{r_fr_i}\right)^{(n-1)/2} \int \mathscr{D}r(t) \exp\left\{\frac{i}{\hbar}\int_0^T \left\{m\frac{\dot{r}^2}{2} - \frac{\hbar^2}{2mr^2}\left(\left(l+\frac{n}{2}-1\right)^2 - \frac{1}{4}\right) + \frac{\alpha}{r}\right\}dt\right\}. \tag{7}$$

III. GREEN'S FUNCTION FOR THE COULOMB POTENTIAL

In order to calculate the radial part (7), we make use of the transformation $(r,t) \to (u,s)$ defined as follows:

$$r = u^2, \quad \frac{dt}{ds} = 4u^2(s). \tag{8}$$

Then, by discretizing,

$$r_j = u_j^2 \quad r_{j-1} = u_{j-1}^2, \quad \epsilon = t_j - t_{j-1} = 4(s_j - s_{j-1})u_ju_{j-1} = 4\sigma_ju_ju_{j-1}.$$

The measure is then

$$\left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N/2} \prod_{j=1}^{N} dr_j = \left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N/2} \sum_{j=1}^{N-1} (2u_j\,du_j) = \left(\frac{m}{2i\pi\hbar\epsilon}\right)^{N/2} \frac{1}{(4u_Nu_0)^{1/2}} \prod_{j=1}^{N} (4u_ju_{j-1})^{1/2} \prod_{j=1}^{N-1} du_j$$
$$= \frac{1}{(4u_Nu_0)^{1/2}} \prod_{j=1}^{N} \left(\frac{m}{2i\pi\hbar\sigma_j}\right)^{1/2} \prod_{j=1}^{N-1} du_j. \tag{9a}$$

In the same way, the action S, when expanded around the mean position of the time interval $[j,j-1]$, is equal to

$$S(j,j-1) = \frac{m}{2\epsilon}(r_j-r_{j-1})^2 - \epsilon\left[\frac{\hbar^2}{2mr_jr_{j-1}}\left(\left(l+\frac{n}{2}-1\right)^2 - \frac{1}{4}\right) + V(\tilde{r}_j)\right]$$
$$\simeq \frac{m}{2\sigma_j}\frac{\tilde{u}_j^2}{u_ju_{j-1}}\Delta u_j^2 - \sigma_j\left[\frac{\hbar^2}{2m\tilde{u}_j^2}((2l+n-2)^2-1) - 4\alpha\right]$$
$$\simeq \frac{m}{2\sigma_j}\Delta u_j^2 - \sigma_j\left[\frac{\hbar^2}{2m\tilde{u}_j^2}((2l+n-2)^2-1) - 4\alpha\right] + \frac{m}{8\sigma_j}\frac{\Delta u_j^4}{\tilde{u}_j^2}, \tag{9b}$$

where

$$\tilde{x}_j = (x_j + x_{j-1})/2, \quad \Delta x = x_j - x_{j-1}.$$

We introduce the energy E by means of the Green's function (Fourier transformed of the propagator):

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = \int_0^\infty dT \exp\left(\frac{iET}{\hbar}\right) K(\mathbf{r}_f,\mathbf{r}_i;T) = \sum_{l=0}^{\infty} G_l(r_f,r_i;E) \frac{\Gamma(n/2-1)(2l+n-2)}{4\pi^{n/2}} C_l^{n/2-1}(\cos \Theta_{i,f}), \tag{10}$$

where the radial part G_l is obtained by using the identity

$$4u_i u_f \int_0^\infty ds'\, \delta\Big(\mathrm{T} - 4\int_0^{s'} ds\, u^2(\mathrm{s})\Big) = 1\,,$$

and the Mc Laughlin–Schulman procedure,[8] which allows us to replace the term $(m/8\sigma_j)\Delta u_j^4/\tilde{u}_j^2$ in Eq. (9b) by a pure quantum correction $(-3\hbar^2\sigma_j/8m)$:

$$G_l(r_f,r_i;E) = G_l(u_f^2,u_i^2;E) = \frac{2}{(u_f u_i)^{n-3/2}}\int_0^\infty ds' e^{i4\alpha s'/\hbar} P_E(u_f,u_i;s')ds'\,, \tag{11}$$

with

$$P_E(u_f,u_i;s') = \int \mathscr{D}u(s)\exp\left\{\frac{i}{\hbar}\int_0^{s'} ds\left[\frac{m\dot{u}^2}{2} - \frac{\hbar^2}{2mu^2}\Big((2l+n-2)^2 - \frac{1}{4}\Big) + 4Eu^2\right]\right\}. \tag{12}$$

If we change $2l+n-2$ into $\lambda' + \nu'/2 - 1$, then $P_E(u_f,u_i;s')$ represents, except for the missing factor $(1/u_f u_i)^{(\nu'-1)/2}$, the radial kernel[9] of the harmonic oscillator, having the angular momentum λ', in a ν'-dimensional space. After integrating $P_E(u_f,u_i;s')$,[10] Eq. (11) becomes

$$G_l(u_f^2,u_i^2;E) = \frac{2}{(u_f u_i)^{n-2}}\frac{m\omega}{i\hbar}\int_0^\infty ds' \frac{e^{i4\alpha s'/\hbar}}{\sin(\omega s')} I_{2l+n-2}\left(\frac{m\omega u_f u_i}{i\hbar\sin(\omega s')}\right)\exp\left\{\frac{im\omega}{2\hbar}(u_f^2+u_i^2)\cot(\omega s')\right\}, \tag{13}$$

where $\frac{1}{2}m\omega^2 = -4E$.

Let $\omega = 2i\hbar k/m$, $p' = -i\alpha m/\hbar^2 k$. Then, with the help of the formula[11]

$$\int_0^\infty dq\, \frac{e^{-2p'q}}{\sinh q} e^{-(1/2)(x+y)\coth q} I_{2\gamma}\left(\frac{(xy)^{1/2}}{\sinh q}\right) = \frac{\Gamma(p'+\gamma+\frac{1}{2})}{(xy)^{1/2}\Gamma(2\gamma+1)} M_{-p',\gamma}(x) W_{-p',\gamma}(y)\,,$$

where $M_{-p',\gamma}(x)$ and $W_{-p',\gamma}(y)$ are Whittaker functions[12], the expression (13) can also be integrated, and the n-dimensional H-atom Green's function is written as follows:

$$\begin{aligned} G(\mathbf{r}_f,\mathbf{r}_i;E) = {} & \frac{m}{\hbar k}(r_i r_f)^{(1-n)/2}\sum_{l=0}^\infty \frac{\Gamma(l+n/2-\frac{1}{2}+p')}{\Gamma(2l+n-1)} M_{-p',l+n/2-1}(-2ikr_i) \\ & \times W_{-p',l+n/2-1}(-2ikr_f)\,\frac{(2l+n-2)\Gamma(n/2-1)}{4\pi^{n/2}} C_l^{(n/2)-1}(\cos\Theta_{i,f}) \quad (r_f>r_i)\,. \end{aligned} \tag{14a}$$

The poles of the Euler function $\Gamma(l+n/2-\frac{1}{2}+p')$, $l+n/2-\frac{1}{2}+p' = -n_1$, with $n_1 = 0,1,2,..,\infty$, determine the n-dimensional H-atom spectrum:

$$E_{n',l}^{(n)} = -\alpha^2 m/2\hbar^2(n'+n/2-\tfrac{3}{2})^2\,,$$

where

$$n' = n_1 + l + 1 = l+1, l+2,\ldots,\infty\,. \tag{14b}$$

The expressions (14a) and (14b) generalize the results obtained in Ref. 13. It should be noted that Green's function (10)–(13) can be put into compact form.

First case $(n=2)$: With the help of the formulas[14]

$$\lim_{\lambda\to 0}\Gamma(\lambda)C_n^\lambda(\cos\phi) = (2/n)\cos n\phi, \quad C_0^0(\cos\phi) = 1\,,$$

$$\begin{aligned} \cos(z\cos\phi) &= 2^\nu\Gamma(\nu)\sum_{m=0}^\infty(-)^m(\nu+2m)\frac{J_{\nu+2m}(z)}{z^\nu}C_{2m}^\nu(\cos\phi) \\ &= J_0(z) + 2\sum_{m=1}^\infty(-)^m J_{2m}(z)\cos(2m\phi) \quad (\nu=0)\,, \end{aligned}$$

Eq. (14a) leads to

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = \frac{m}{\hbar k}(r_i r_f)^{-1/2}\sum_{m=-\infty}^{+\infty}\frac{\Gamma(|m|+p'+\frac{1}{2})}{\Gamma(2|m|+1)} M_{-p',|m|}(-2ikr_i) W_{-p',|m|}(-2ikr_f)\frac{e^{im(\theta_f-\theta_i)}}{2\pi}\,,$$

$\theta_f - \theta_i = (\mathbf{r}_f,\mathbf{r}_i)$ and $r_f > r_i$.

The Green's function can then be written as follows:

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = \frac{-im}{\pi\hbar}\int_0^\infty dq\,\frac{e^{-2p'q+ik(r_f+r_i)\coth q}}{\sinh q}\cos\left[\frac{2k(r_i r_f)^{1/2}}{\sinh q}\cos\left(\frac{\Theta i,f}{2}\right)\right]. \tag{15}$$

Second case $(n \geqslant 3)$: The use of the formulas[15]

$$\left(\frac{1}{2}kz\right)^{\mu-\nu} J_\nu(kz) = k^\mu \sum_{l=0}^{\infty} \frac{\Gamma(\mu+l)}{l!\,\Gamma(\nu+1)}\, {}_2F_1(\mu+l,-l;\nu+1;k^2)(\mu+2l)J_{\mu+2l}(z)\,,$$

$$C_n^\lambda(\cos\Theta) = \frac{\Gamma(2\lambda+n)}{\Gamma(n+1)\Gamma(2\lambda)}\, {}_2F_1\left(2\lambda+n,-n;\lambda+\frac{1}{2};\sin^2\left(\frac{\Theta}{2}\right)\right)$$

$$= (-)^n \frac{\Gamma(2\lambda+n)}{\Gamma(n+1)\Gamma(2\lambda)}\, {}_2F_1\left(2\lambda+n,-n;\lambda+\frac{1}{2};\cos^2\left(\frac{\Theta}{2}\right)\right)$$

allows the Green's function to be put, in this case, into the following form:

$$G(\mathbf{r}_f,\mathbf{r}_i;E) = (-i)^{n-1}\frac{m}{\hbar}(r_f r_i)^{(1/4)(3-n)} k^{(n-1)/2}\left[\cos\left(\frac{\Theta_{i,f}}{2}\right)\right]^{(3-n)/2} 2^{2-n}$$

$$\times \pi^{(1-n)/2}\int_0^\infty dq\, \frac{\exp\{-2p'q + ik(r_f+r_i)\coth q\}}{(\sinh q)^{(n+1)/2}} J_{(n-3/2)}\left(\frac{2k(r_f r_i)^{1/2}}{\sinh q}\cos\left(\frac{\Theta_{i,f}}{2}\right)\right). \qquad (16)$$

Since $J_{-1/2}(z) = [2/\pi z]^{1/2}\cos z$, it can be seen that the expression (15) is included in (16).

The n-dimensional H-atom Green's function (16) ($n \geqslant 2$) is a generalization of the cases where $n = 2$ and $n = 3$, corresponding to the Levi-Civita and Kustaanheimo–Stiefel transformations.[16,17] In the case $n = 1$, the Green's function may be obtained from Eq. (16) ($\Theta_{i,f} \to 2\pi$)

$$G(x_f,x_i;E) = -\frac{2m}{\hbar}(x_f x_i)^{1/2}\int_0^\infty \frac{dq}{\sinh q}\exp\{-2p'q + ik(x_f+x_i)\coth q\} J_{-1}\left(\frac{-2k(x_f x_i)^{1/2}}{\sinh q}\right).$$

Since

$$J_{-1}(-z) = -i\mathrm{I}_1(-iz), \quad z = \frac{2k(x_f x_i)^{1/2}}{\sinh q},$$

then

$$G(x_f,x_i;E) = (m/\hbar k)\Gamma(p'+1)M_{-p',1/2}(-2ikx_i)W_{-p',1/2}(-2ikx_f)\,,$$

with $x_f > x_i$. This result was already obtained in Ref. 18, by standard wave mechanics. This one-dimensional Coulomb problem may have degenerate levels.[19] Furthermore, it is easy to show, by using the identity[20]

$$\left(\frac{d}{z\,dz}\right)^m\left(\frac{I_\nu(z)}{z^\nu}\right) = \frac{I_{\nu+m}(z)}{z^{\nu+m}},$$

that the Green's function (16) of the H-atom in an n-dimensional Euclidean space, is connected with the $(n-2)$-dimensional Green's function, by

$$G_n(x,y) = -\frac{1}{2\pi y}\frac{\partial}{\partial y}G_{n-2}(x,y)\,,$$

where $x = r_f + r_i$ and $y = |\mathbf{r}_f - \mathbf{r}_i|$. This relation was already given by Hostler.[9]

IV. CONCLUSION

The n-dimensional H-atom Green's function has been obtained as an expansion in partial Green's function (14a) and in a compact form (16), in the functional integral formalism. To our knowledge this is the first functional derivation of this n-dimensional Coulomb Green's function. The known results of the hydrogenic literature are recovered, in the case $n = 1$ (Schrödinger formulation[18]), in the cases $n = 2,3$ (Schrödinger and path integral[13,16] formalisms, or phase space formalism[17]), and in the general n-dimensional case (Schrödinger formulation[9]).

[1] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965), p. 33, Eq. (2-22).

[2] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1965), p. 980, Eq. (8-534).

[3] H. Bateman, *Higher Transcendental Functions,* (McGraw-Hill, New York, 1953), Vol. 2, p. 98, Eq. (1).

[4] E. Whittaker and G. Watson, *A Course of Modern Analysis* (Cambridge U.P., Cambridge, 1952), p. 355.

[5] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1966), p. 77; Ref. 2, pp. 1029–31.

[6] Reference 3, p. 178, Eq. (34).

[7] Reference 2, p. 961, Eq. (8-41).

[8] D. McLaughlin and L. S. Schulman, J. Math. Phys. **12**, 2520 (1971).

[9] S. M. Blinder, J. Math. Phys. **25**, 905 (1984); L. C. Hostler, J. Math. Phys. **11**, 2966 (1970); D. Bergmann and Y. Frishman, J. Math. Phys. **6**, 1855, (1985); A. Giovannini and T. Tonietti, Nuovo Cimento A **54**, 1 (1968).

[10] D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969); W. Langguth and A. Inomata, J. Math. Phys. **20**, 499 (1979).

[11] Reference 2, p. 729, Eq. (6-699-4).

[12] Reference 4, p. 337.

[13] N. K. Pak and I. Sokmen, Phys. Rev. **30**, 1629 (1984); F. Steiner, Phys. Lett. A **106**, 356, 363 (1984).

[14] Reference 2, p. 1030, Eq. (8-934-4); Ref. 3, p. 102, Eq. (44).

[15] Reference 5, p. 140, Eq. (3); Ref. 2, p. 1030, Eq. (8-932-1); Ref. 3, p. 176, Eq. (20).

[16] I. H. Duru and H. Kleinert, Phys. Lett. B **84**, 185 (1979); Fortschr. Phys. **30**, 401 (1982); A. Inomata, Phys. Lett. A **87**, 387 (1982); R. Ho and A. Inomata, Phys. Rev. Lett. **48**, 231 (1982).

[17] J. M. Gracia-Bondia, Phys. Rev. A **30**, 691 (1984).

[18] J. Meixner, Math. Z. **36**, 677 (1933).

[19] R. Loudon, Am. J. Phys. **27**, 649 (1959).

[20] Reference 5, p. 79, Eq. (6).

Time-dependent invariant associated to nonlinear Schrödinger–Langevin equations

Antônio B. Nassar
Department of Physics, University of California, Los Angeles, California 90024

(Received 17 March 1986; accepted for publication 20 August 1986)

Via the quantum-hydrodynamical method, a time-dependent invariant associated to the quantum dissipative time-dependent harmonic oscillator (TDHO), described by two classes of nonlinear Schrödinger–Langevin equations with the following frictional nonlinear terms is constructed: (i) $W_1 = -i\hbar\nu(\ln\psi - \langle \ln\psi\rangle)$, which is the Schuch–Chung–Hartman frictional nonlinear term, and (ii) $W_2 = \nu\{[x - \langle x\rangle][c\hat{p} + (1-c)\langle\hat{p}\rangle] - \frac{1}{2}i\hbar c\}$, which includes the Süssmann ($c=1$), the Hasse ($c=\frac{1}{2}$), and the Albrecht–Kostin ($c=0$) frictional nonlinear operators. The associated invariant found is *exact* for the Schuch–Chung–Hartman and Hasse models, and only *approximate* for the Süssman and Albrecht–Kostin models.

I. INTRODUCTION

Among the remarkable successes of the theory of time-dependent invariants, as a manifold problem-solving tool in several areas of physics, we point out its prime relevance in connection with quantum physics.[1–12] For instance, the invariant has been used as an artifact to construct an exact solution to the associated time-dependent Schrödinger equation for certain types of potentials.

Methods that have been used for deriving invariants for time-dependent Lagrangian/Hamiltonian systems include Noether's theorem, the Lie theory of extended groups, Ermakov's method, the theory of canonical transformations, and the direct method.[13,14] In addition, the present author has recently introduced a new method, by extracting an invariant associated with the quantum time-dependent harmonic oscillator, via the hydrodynamical formulation of quantum mechanics: the quantum-hydrodynamical method.[11] This method provides a physically intuitive and mathematically transparent connection between the quantum problem and its associated classical counterpart.

We have exploited further the quantum-hydrodynamical method in the search of invariants associated to the quantum dissipative time-dependent harmonic oscillator described by two different models[15]: (1) an explicitly time-dependent linear Schrödinger–Langevin equation (the Caldirola–Kanai model), and (2) a logarithmic nonlinear Schrödinger–Langevin equation (the Kostin model). For the former model, we obtained an exact associated time-dependent invariant. For the latter model, we showed instead that *neither* an exact nor an approximate time-dependent invariant can be constructed.[15]

At this point one can ask a most pertinent question, tacitly suggested at the conclusion of our previous work.[15] *Does any other nonlinear Schrödinger–Langevin equation admit an associated time-dependent invariant*? In this paper we answer this question *affirmatively*.

Via the quantum-hydrodynamical method, we show that it makes it possible to find an exact or approximate configurational invariant associated to the quantum dissipative time-dependent harmonic oscillator (TDHO) described by certain classes of nonlinear Schrödinger–Langevin equations (NLSLE's), *which do not admit usual quantization prescription*. These equations will be presented and studied later (see Secs. II and III). In doing so, we demonstrate the essential feature of our method[15]: that is, one can deal with quantum dissipative time-dependent systems, *irrespective of* whether or not these systems can be defined by *an underlying Lagrangian/Hamiltonian* formalism. Apparently, this is an ingredient which makes our method appealing *vis à vis* the other aforementioned methods, since the problem under consideration is outside their realm of operationality.

In passing, we point out an interesting protocol developed by Remaud and Hernandez[16] in the construction of an exact invariant for the dissipative classical time-dependent harmonic oscillator. There, by generalizing a prescription originally given by Symon[17] and without any particular Hamiltonian formalism, they proceed further on quantizing their classical invariant aiming a close connection with a quantum description of the associated dissipative process. As these same authors point out, however, the quantization of this invariant is *not* obvious as one is forced to take into account sizable fluctuations. On the other hand, our work takes the reverse direction. We start from different quantum formulations of the dissipative time-dependent harmonic oscillator problem and then interconnect them with their classical counterpart description through two fluid-dynamical equations: *a continuity equation and an Euler-type equation.*[15] As we will see below, the latter equation is the main bridge for the transition from the quantum to its classical counterpart problem: it generates the proper pair of equations (the Ermakov system) which lead to the companion invariant of motion.

In Sec. II, we study the quantum dissipative time-dependent harmonic oscillator described by a nonlinear Schrödinger–Langevin equation with frictional logarithmic nonlinear operator $W_1 = -i\hbar\nu(\ln\psi - \langle\ln\psi\rangle)$, that is, the Schuch–Chung–Hartmann model.[18] For this model we derive an *exact* associated time-dependent invariant. In Sec. III, by proceeding along the lines of Sec. II, we study another

 0022-2488/86/122949-04$02.50

class of a nonlinear Schrödinger–Langevin equation with frictional nonlinear operator

$$W_2 = \nu\{[x - \langle x\rangle][c\hat{p} + (1-c)\langle \hat{p}\rangle] - \tfrac{1}{2}i\hbar c\},$$

which includes the Süssmann ($c=1$), Hasse ($c=\frac{1}{2}$), and Albrecht–Kostin ($c=0$) models.[19–22] We show that the *same* invariant, derived in Sec. II, is also *exact* for the Hasse model, while being only *approximate* for the Süssmann and Albrecht–Kostin models. In the last section, we study the quantum-hydrodynamical energy dissipation theorem, for each of the models under consideration, and comment about the possible implications of the results found.

II. NONLINEAR SCHRÖDINGER–LANGEVIN EQUATION WITH FRICTIONAL NONLINEAR TERM $W_1 = -\hbar\nu(\ln\psi - \langle\ln\psi\rangle)$: AN EXACT ASSOCIATED INVARIANT

We begin with the quantum formulation of the dissipative time-dependent harmonic oscillator described by the Schuch–Chung–Hartmann model,[18] namely,

$$i\hbar\frac{\partial\psi}{\partial t}(x,t) = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}(x,t) + \left(\frac{\hbar\nu}{i}(\ln\psi(x,t) - \langle\ln\psi(x,t)\rangle) + \frac{1}{2}m\omega^2(t)x^2\right)\psi(x,t), \tag{2.1}$$

where $\psi(x,t),\nu$, and $w(t)$ are the wave function, constant friction coefficient, and time-dependent harmonic oscillator frequency, respectively. The reader is referred to the works of Schuch, Chung, and Hartmann[18] for a detailed discussion of the compelling physical reasons that motivated the study of this interesting equation.

To obtain the quantum-hydrodynamical description of Eq. (2.1), we write the wave function in the form[15]

$$\psi(x,t) = \phi(x,t)\exp[iS(x,t)]. \tag{2.2}$$

After substitution of Eq. (2.2) into Eq. (2.1), we obtain from its real and imaginary parts

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \nu v + w^2(t)x = -\frac{1}{m}\frac{\partial V_{qu}}{\partial x}, \tag{2.3}$$

and

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) = -\nu\rho(\ln\rho - \langle\ln\rho\rangle), \tag{2.4}$$

where $\rho\equiv\phi^2$, $v\equiv(\hbar/m)(\partial S/\partial x)$, and $V_{qu}\equiv-(\hbar^2/2m)\rho^{-1/2}(\partial^2\rho^{1/2}/\partial x^2)$ are the quantum fluid density, the quantum fluid velocity, and the quantum potential, respectively. Equation (2.3) is an Euler-type equation describing trajectories of a fluid particle, with momentum $p=mv$, in a quantum dissipative medium, whereas Eq. (2.4) will be shown below to take the form of a Fokker–Planck-type equation. The above set of Eqs. (2.3) and (2.4) constitutes our fundamental stepping stone in the search for the invariant of motion associated to the nonlinear Schrödinger–Langevin Eq. (2.1).

Likewise in an earlier work,[15] we assume that the expectation value of the quantum force vanishes for all times, i.e., $\langle F_{qu}\rangle \equiv -\langle\partial V_{qu}/\partial x\rangle = 0$, suggesting that (an ansatz) $F_{qu} \propto (x - \langle x\rangle)$. We denote $\langle A\rangle \equiv \int\rho A\,dz$ as the expectation (average) value of A taken over an ensemble of identical fluid-particles. In doing so, we may split (2.3) into

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} + \nu v + w^2(t)x = k(t)\{x - q(t)\}, \tag{2.5}$$

and

$$\frac{\partial}{\partial x}\left[\frac{\hbar^2}{2m^2}\rho^{-1/2}\frac{\partial^2\rho^{1/2}}{\partial x^2}\right] = k(t)\{x - q(t)\}, \tag{2.6}$$

where $q(t)$ is the expectation value of x [$\langle x\rangle = q(t)$], which will be determined in concomitance with $k(t)$.

We prepare the fluid-particle initially in a Gaussian wave packet centered at $x=0$, $\rho_0(x)\equiv\rho(x,0) = [\pi\sigma(0)]^{-1/2}\exp[-x^2/\sigma(0)]$, with an initial velocity $v_0(x) = v(x,0)$, and integrate (2.6) (assuming that ρ vanishes for $|x|\to\infty$ at any time) to obtain

$$\rho(x,t) = [\pi\sigma(t)]^{-1/2}\exp\{-[x - q(t)]^2/\sigma(t)\}, \tag{2.7a}$$

where

$$\sigma^2(t) \equiv \hbar^2/m^2k(t). \tag{2.7b}$$

Next, with the help of (2.7a), Eq. (2.4) can be rewritten as

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(\rho v) - D\frac{\partial^2\rho}{\partial x^2} = 0, \tag{2.8}$$

with $D\equiv\nu\sigma/4$. Equation (2.8) has the form of a Fokker–Plank-type equation.

After integrating (2.8), we find

$$\upsilon(x,t) = \dot\sigma/2\sigma[x - q(t)] + \dot q(t), \tag{2.9a}$$

with

$$\upsilon \equiv v - D\frac{1}{\rho}\frac{\partial\rho}{\partial x} = v + \frac{\nu}{2}[x - q(t)]. \tag{2.9b}$$

The constant of integration in Eq. (2.9a) must be zero since ρ vanishes for $|x|\to\infty$.

By inserting Eq. (2.9) into (2.5), one easily obtains

$$\left(\frac{\ddot\sigma}{2\sigma} - \frac{\dot\sigma^2}{4\sigma^2} + \left(\omega^2(t) - \frac{\nu^2}{4}\right) - \frac{\hbar^2}{m^2\sigma^2}\right)(x - q) + (\ddot q + \nu\dot q + \omega^2(t)q) = 0. \tag{2.10}$$

This equation is actually the final form of the Euler-type equation (2.3), and, as we show right below, represents the main bridge from the quantum to the classical counterpart problem. It is identically satisfied if

$$\ddot\alpha + [w^2(t) - \nu^2/4]\alpha = 1/\alpha^3, \tag{2.11}$$

and

$$\ddot q + \nu\dot q + w^2(t)q = 0, \tag{2.12a}$$

where we have made $\sigma = (\hbar/m)\alpha^2$. By the change of variable $q = ue^{-\nu t/2}$, we recast Eq. (2.12a) as

$$\ddot u + (w^2(t) - \nu^2/4)u = 0. \tag{2.12b}$$

By eliminating $w^2(t) - \nu^2/4$ between Eqs. (2.11) and (2.12b), and, after some simple manipulations, we end up with

$$\dot I = 0, \tag{2.13a}$$

where

$$I = \tfrac{1}{2}[(\dot u\alpha - \dot\alpha u)^2 + (u/\alpha)^2] \qquad (2.13b)$$

or

$$I = \tfrac{1}{2}e^{\nu t}[(\dot q\alpha - \dot\alpha q + (\nu/2)q\alpha)^2 + (q/\alpha)^2], \qquad (2.13c)$$

which is the *exact* time-dependent invariant associated to the Schuch–Chung–Hartmann model.

Another interesting way of arriving at (2.13) is to write

$$q(t) = e^{-\nu t/2}\alpha(t)r(\tau), \qquad (2.14)$$

where τ is related to t through $[d\tau = \mu(t)dt]$:

$$\tau(t) = \int^t \mu(\lambda)d\lambda. \qquad (2.15)$$

Substitution of (2.14) into (2.12a) yields

$$(\alpha\mu^2)r'' + (2\mu\dot\alpha + \dot\mu\alpha)r' + (\ddot\alpha + (w^2(t) - \nu^2/4)\alpha)r = 0. \qquad (2.16)$$

This equation can be reduced to

$$r'' + r = 0, \qquad (2.17)$$

by making

$$2\dot\alpha/\alpha + \dot\mu/\mu = 0 \Rightarrow \mu = 1/\alpha^2 \qquad (2.18a,b)$$

and

$$\ddot\alpha + [w^2(t) - \nu^2/4]\alpha = 1/\alpha^3. \qquad (2.11)$$

By multiplying Eq. (2.17) by r' and integrating we obtain

$$I = \tfrac{1}{2}[(r')^2 + r^2]. \qquad (2.19)$$

Equation (2.13c) can be recovered from the above invariant by replacing back $r = e^{\nu t/2}(q/\alpha)$ and $d\tau = dt/\alpha^2$. Thus, we conclude that Eqs. (2.11) and (2.12a,b) constitute an *Ermakov system*, since they generate an Ermakov–Lewis-type of invariant.[15]

III. FRICTION OPERATOR $W_2 = \nu\{[x - \langle x\rangle][c\hat p + (1-c)\langle\hat p\rangle] - \frac{1}{2}i\hbar c\}$: AN EXACT AND APPROXIMATE ASSOCIATED INVARIANT

Next, we conside the quantum dissipative time-dependent harmonic oscillator described by the following class of nonlinear Schrödinger–Langevin equation:

$$\begin{aligned} i\hbar\frac{\partial\psi}{\partial t}(x,t) &= -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2}(x,t) \\ &\quad + (\nu\{[x - \langle x\rangle][c\hat p + (1-c)\langle\hat p\rangle] - \tfrac{1}{2}i\hbar c\} \\ &\quad + \tfrac{1}{2}mw^2(t)x^2)\psi(x,t), \end{aligned} \qquad (3.1)$$

where the frictional nonlinear operator term $\nu\{[x - \langle x\rangle][c\hat p + (1-c)\langle\hat p\rangle] - \frac{1}{2}i\hbar c\}$ includes those proposed by Süssmann ($c = 1$), Hasse ($c = \frac{1}{2}$), and Albrecht–Kostin ($c = 0$), as the special cases.[19–22]

A comprehensive analysis and critique of the compelling physical reasons related to the formal mathematical attractiveness, which motivate the study of this class of nonlinear Schrödinger–Langevin equations, can be found in the works of Stocker and Albrecht,[19] Albrecht,[20] Hasse,[21] and Kostin.[22]

To obtain the quantum-hydrodynamical description of Eq. (3.1), we proceed along the lines of Sec. II $[\psi = \phi \times\exp(iS)]$, resulting in

$$\begin{aligned} &\frac{\partial\nu}{\partial t} + \nu\frac{\partial\nu}{\partial x} + \nu\langle\nu\rangle + \omega^2(t)\langle x\rangle \\ &\quad = -\frac{1}{m}\frac{\partial}{\partial x}\left(V_{qu} + \frac{1}{2}m\Omega^2(t)(x - \langle x\rangle)^2\right) \end{aligned} \qquad (3.2)$$

and

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial x}(\rho\nu) = 0, \qquad (3.3)$$

where $\Omega^2(t) \equiv \omega^2(t) - \nu^2c^2$, $\rho \equiv \phi^2$, $\nu \equiv v + \nu c(x - \langle x\rangle)$ $[v \equiv (\hbar/m)(\partial S/\partial x)]$, and $V_{qu} \equiv -(\hbar^2/2m)\rho^{-1/2} \times(\partial^2\rho^{1/2}/\partial x^2)$ are the shifted frequency, the quantum fluid density, the generalized quantum fluid velocity (quantum fluid velocity), and the quantum potential, respectively.

By following closely the protocol developed in Sec. II, we convert the Euler-type equation (3.2) into the form

$$\begin{aligned} &\left(\frac{\ddot\sigma}{2\sigma} - \frac{\dot\sigma^2}{4\sigma^2} + (\omega^2(t) - c^2\nu^2) - \frac{\hbar^2}{m^2\sigma^2}\right)(x - q) \\ &\quad + (\ddot q + \nu\dot q + \omega^2(t)q) = 0. \end{aligned} \qquad (3.4)$$

This equation is identically satisfied if

$$\ddot\alpha + [\omega^2(t) - c^2\nu^2]\alpha = 1/\alpha^3, \qquad (3.5)$$

and

$$\ddot q + \nu\dot q + \omega^2(t)q = 0, \qquad (3.6)$$

with $\sigma = (\hbar/m)\alpha^2$.

Notice that by simply setting $c = \frac{1}{2}$ in Eq. (3.5), we recover Eq. (2.11). Thus, the results obtained in Sec. II hold also true for the Hasse model. In fact, *the* Schuch–Chung–Hartmann and Hasse models are alike under the protocol developed above, since the solution set to their corresponding quantum problem, i.e., $\rho(x,t)$ and $v(x,t)$ [or $\nu(x,t)$], are the same. To the best of our knowledge, this has not been realized in previous works.[18]

In general, for any values of c and ν in Eqs. (3.5) and (3.6)—including the Süssmann ($c = 1$) and Albrecht–Kostin ($c = 0$) models—no *exact* time-dependent invariant can be found. It turns out, however, that if ν^2 (and $c^2\nu^2$) is *very small* compared to, say, $\omega^2(t)$, then equations (3.5) and (3.6) can be approximated to

$$\ddot\alpha + \omega^2(t)\alpha = 1/\alpha^3, \qquad (3.7)$$

and

$$\ddot u + \omega^2(t)u = 0, \qquad (3.8)$$

with $u = e^{\nu t/2}q$.

Hence, the invariant constructed from Eqs. (3.7) and (3.8) is the *same* as that obtained in Sec. II [Eq. (2.13)]. It should be stressed,however, that this invariant is only *approximate* for the Süssmann ($c = 1$) and Albrecht–Kostin ($c = 0$) models—that is, if $\nu^2 \ll \omega^2(t)$.

IV. FINAL REMARKS AND CONCLUSION

It is worthwhile noticing that, if $\omega(t) = \omega_0$ (a constant), Eq. (2.13a) for the associated classical *invariant* reduces to the well-known equation for the classical energy disspation theorem

$$\frac{d}{dt}\left(\frac{1}{2}(\dot{q}^2+\omega_0^2 q^2)\right) = -\nu\dot{q}^2. \tag{4.1}$$

So, one further important, suggestive point remains to be investigated, in light of the second law of thermodynamics. We must verify whether the nonlinear Schrödinger–Langevin equations studied in Secs. II and III also fulfill the quantum-hydrodynamical-energy-dissipation theorem.[19]

(1) For the Schuch–Chung–Hartman model, we can construct the following relation:

$$\frac{\partial U}{\partial t}+\frac{\partial \Pi}{\partial x} = -\nu\rho v^2 - \frac{\nu}{2}\frac{U}{\rho}\frac{\partial}{\partial x}[\rho(x-q)], \tag{4.2a}$$

or

$$\frac{\partial U}{\partial t}+\frac{\partial}{\partial x}\left[\Pi+\frac{\nu}{2}U(x-q)\right]$$
$$= -\nu\rho v^2 - \frac{\nu}{2}\rho(x-q)\frac{\partial}{\partial x}\left(\frac{U}{\rho}\right), \tag{4.2b}$$

where

$$U=\tfrac{1}{2}\rho v^2+\rho(V_{qu}+V)/m, \tag{4.3}$$

is the quantum-fluid energy density, and

$$\Pi=\rho v\left(\frac{v^2}{2}+\frac{V_{qu}+V}{m}\right)+\frac{\hbar^2}{2m^2}\left[\left(\phi\frac{\partial^2\phi}{\partial x\,\partial t}\right)-\frac{\partial\phi}{\partial t}\frac{\partial\phi}{\partial x}\right] \tag{4.4}$$

is the quantum-fluid energy flow density.

By averaging Eq. (4.2b), one obtains

$$\dot{E} = -\nu\langle v^2\rangle - \frac{\nu}{2}\left[\left\langle x\frac{\partial}{\partial x}\left(\frac{U}{\rho}\right)\right\rangle - \langle x\rangle\left\langle\frac{\partial}{\partial x}\left(\frac{U}{\rho}\right)\right\rangle\right], \tag{4.5}$$

where $\dot{E}\equiv\langle\langle\partial U/\partial t\rangle\rangle$, $\langle\cdot\rangle\equiv\int\cdot\rho\,dx$, and $\langle\langle\cdot\rangle\rangle\equiv\int\cdot dx$. Whether this expression *always* fulfills the correctly expected inequality $\dot{E}\leqslant 0$ (for $\omega=\omega_0$) is *not obvious* whatsoever. We leave this point, however, for a detailed examination in a future work.

(2) For the other class of nonlinear Schrödinger–Langevin equations, we may also construct its corresponding relation for the quantum-mechanical energy dissipation theorum, namely

$$\frac{\partial\mathscr{V}}{\partial t}+\frac{\partial\P}{\partial x} = -\nu\rho\mathscr{v}[\langle v\rangle-\nu c^2(x-\langle x\rangle)], \tag{4.6}$$

where now

$$\mathscr{V}=\tfrac{1}{2}\rho\mathscr{v}^2+\rho(V_{qu}+V)/m, \tag{4.7}$$

is the quantum-fluid energy density, and

$$\P=\rho\mathscr{v}\left(\frac{\mathscr{v}^2}{2}+\frac{V_{qu}+V}{m}\right)+\frac{\hbar^2}{2m^2}\left[\left(\phi\frac{\partial^2\phi}{\partial x\,\partial t}\right)-\frac{\partial\phi}{\partial t}\frac{\partial\phi}{\partial x}\right] \tag{4.8}$$

is the quantum-fluid energy flow density.

By averaging Eq. (4.6) and using (2.7a), one gets

$$\dot{\mathscr{E}} = -\nu\langle v\rangle^2+\nu^2c^2\langle(x-\langle x\rangle v\rangle$$
$$+\nu^3c^3\langle(x-\langle x\rangle)^2\rangle, \tag{4.9}$$

where $\dot{\mathscr{E}}\equiv\langle\langle\partial\mathscr{V}/\partial t\rangle\rangle$. So, in order to guarantee that $\dot{\mathscr{E}}\leqslant 0$ (for $\omega=\omega_0$), we should have $c=0$. Thus, this seems to be the most plausible choice for the parameter c (see Ref. 20).

Another interesting remark worthy of notice. If we make the change of variable $\alpha=\zeta e^{\nu t/2}$ in Eq. (2.11), we get

$$\ddot{\zeta}+\nu\dot{\zeta}+\omega^2(t)\zeta=e^{-2\nu t}/\zeta^3, \tag{4.10}$$

which, together with Eq. (2.12a), forms the pair of equations derived in our earlier work for the quantized Caldirola–Kanai model (an explicitly time-dependent, *linear* Schrödinger–Langevin equation).

At last, a purely speculative point, if not a further question, for a future and more elaborated investigation: the existence of an invariant indicates, perhaps, that the nonlinear Schrödinger–Langevin equations studied above (through the quantum-hydrodynamical method) may allow some sort of linearization on the Schwartz space S and may be transformed, thereby, to a Hamiltonian formalism. In passing, Taflin[23] has recently studied the Burgers equation (a dissipative, nonlinear hydrodynamical equation which has no solitons). He has linearized it on the Schwartz space, defined it by a Hamiltonian formalism, and proved that it has an infinity of time-independent invariants in involution.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge stimulating correspondence with Professor P. G. L. Leach, Professor J. M. F. Bassalo, Dr. B. K. Cheng, and Dr. D. Schuch, and fruitful conversations and advice from Professor S. J. Putterman. The last paragraph in Sec. IV was suggested to the author by Professor M. Flato during one of the enjoyable discussions on the subject of this work.

Partial financial support by CNPq (Brazil) is also acknowledged.

[1] H. R. Lewis and W. B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969).
[2] P. Bertrand, N. V. Tuan, M. Feix, and J. Gutierrez, J. Plasma Phys. **23**, 401 (1980).
[3] R. S. Kaushal and H. J. Korsch, J. Math. Phys. **22**, 1904 (1981).
[4] H. J. Korsch and H. Laurent, J. Phys. B **14**, 4213 (1981).
[5] J. G. Hartley and J. R. Ray, Phys. Rev. A **24**, 2873 (1981); D **25**, 382 (1982).
[6] H. J. Korsch, H. Laurent, and R. Möhlenkamp, J. Phys. B **15**, 1 (1982).
[7] J. R. Ray, Phys. Rev. A **26**, 729 (1982); A **28**, 2603 (1983); D **25**, 3417 (1982).
[8] H. Kohl and R. M. Dreizler, Phys. Lett. A **98**, 95 (1983).
[9] D. Besnard, J. R. Burgan, A. Munier, M. R. Feix, and E. Fijalkow, J. Math. Phys. **24**, 1123 (1983).
[10] A. K. Dhara and S. V. Lawande, Phys. Rev. A **30**, 560 (1984); J. Phys. A **17**, 2423 (1984).
[11] A. B. Nassar, Phys. Rev. A **32**, 1862 (1985); A **33**, 4433 (1986).
[12] A. B. Nassar, J. M. F. Bassalo, and P. T. S. Alencar, Phys. Lett. A **113**, 365 (1986).
[13] J. R. Ray and J. L. Reid, Phys. Rev. A **26**, 1042 (1982), and references therein.
[14] H. R. Lewis and P. G. L. Leach, Ann. Phys. (NY) **164**, 47 (1985); J. Math. Phys. **23**, 165 (1983), and references therein.
[15] A. B. Nassar, J. Math. Phys. **27**, 755 (1986). See references 1–26 therein for an account of the methods for the investigation of the time-dependent invariants.
[16] B. Remaud and E. S. Hernandez, Physica A **103**, 35 (1980).
[17] K. R. Symon, J. Math, Phys. **11**, 1320 (1971).
[18] D. Schuch, K. M. Chung, and H. Hartmann, J. Math, Phys. **24**, 1652 (1983); **25**, 3086 (1984); Ber. Bunsenges. Phys. Chem. **89**, 589 (1985).
[19] W. Stocker and K. Albrecht, Ann. Phys. (NY) **117**, 436 (1979).
[20] K. Alkbrecht, Phys. Lett. B **56**, 127 (1975).
[21] R. W. Hasse, J. Math. Phys. **16**, 2005 (1975).
[22] M. D. Kostin, J. Stat. Phys. **12**, 146 (1975).
[23] E. Taflin, Phys. Rev. Lett. **47**, 1425 (1981).

Equivalence between the Lagrangian and Hamiltonian formalism for constrained systems

C. Batlle
Departament de Fisica Teorica, Universitat de Barcelona, Barcelona, Spain

J. Gomis
Department of Physics, Queens College of the City University of New York, Flushing, New York 11367-0904 and Department of Physics, City College of the City, University of New York, New York, New York 10031

J. M. Pons
Departament de Fisica Teorica, Universitat de Barcelona, Barcelona, Spain

N. Roman-Roy
Departament de Fisica, Escola Tècnica Superior d'Enginyers Industrials de Barcelona, Universitat Politecnica de Catalunya, Barcelona, Spain

(Received 25 November 1985; accepted for publication 20 August 1986)

The equivalence between the Lagrangian and Hamiltonian formalism is studied for constraint systems. A procedure to construct the Lagrangian constraints from the Hamiltonian constraints is given. Those Hamiltonian constraints that are first class with respect to the Hamiltonian constraints produce Lagrangian constraints that are FL-projectable.

I. INTRODUCTION

The current interest in constrained systems was spawned by Dirac[1] and Bergmann[2] in their study of the canonical formalism of gravitational fields. Since that time several people contributed to the building of a mechanics for such systems.[3] In particular the Lagrangian,[4] Hamiltonian,[4,5] Hamilton–Jacobi,[6] and geometrical formalisms[7] have been studied. For a time this field of research had little more than mathematical interest, but now with the increasing interest in gauge theories (any theory with gauge transformations is a theory of constrained systems), more people are beginning to use this formalism at the classical and quantum level.

On the other hand, constrained systems with a finite number of degrees of freedom have been used to construct an N-body relativistic mechanics of direct interactions[8] whose corresponding quantum mechanics,[9] which is multitemporal, is related to the Bethe–Salpeter equation.

Despite increasing interest, the mechanics of these systems is not as elaborate as the corresponding mechanics for unconstrained systems. For example, the equivalence between the Lagrangian and Hamiltonian formalism has not been definitely established.[10–12]

In this paper we give an explicit and complete proof of this equivalence. We construct an implicit inverse relation between velocities and momenta, i.e., the inverse Legendre transformation. Using that we deduce the Hamilton–Dirac equations from Euler–Lagrange equations. Neither is a set of normal differential equations, therefore the uniqueness and existence theorem cannot be applied. This means that, at most, we will only have solutions in a submanifold of the respective spaces and in general these solutions will not be unique.

A careful analysis shows that given a solution of the Euler–Lagrange equations we can construct a solution of the Hamilton–Dirac equations and vice versa. Next we look for the appropriate submanifold of the tangent bundle (TQ) and a submanifold of the cotangent bundle (T^*Q) where the solutions exist. These submanifolds are constructed through an iterative procedure. In a given local chart they are characterized by a set of functions that are called constraints.

The Hamiltonian formalism as developed in this paper differs from the usual development.[1,2] The first class primary constraints play a privileged role. Other constraints are either first or second class with respect to them. These constraints that are first class with respect to the primary first class constraints can be associated with Lagrangian constraints that are FL-projectable (or weakly FL-projectable). Those that are second class in the Hamiltonian formalism have associated non-FL-projectable Lagrangian constraints. It is also shown that all constraints other that the primary constraints have either a symmetric or antisymmetric Poisson bracket (PB) structure with the first class primary constraints.

The paper is organized as follows. In Sec. II we show that if we have a solution of the Euler–Lagrange equations we can construct from it a solution of the Hamilton–Dirac equations and vice versa. In Sec. III we develop an algorithm for the determination of the Hamiltonian constraints. In Sec. IV we develop an analogous algorithm for the Lagrangian constraints and we relate the Lagrangian and Hamiltonian constraints.

II. THE EQUIVALENCE THEOREMS

We consider an N-dimensional configuration space Q and a function L, the Lagrangian, defined in its tangent bundle TQ. The Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^j}-\frac{\partial L}{\partial q^j}=0;\quad j=1,\ldots,n, \tag{2.1}$$

can be written in the normal form of ordinary second-order differential equations (SODE) only when the Hessian ma-

 0022-2488/86/122953-10$02.50

trix $W_{ij} \equiv \partial^2 L / \partial \dot{q}^i \partial \dot{q}^j$ is regular.

If the Hessian matrix is singular, neither the existence nor uniqueness theorems for SODE holds. This means that the possible solutions of (2.2) lie in a submanifold of TQ and given a point of that submanifold we can have more than one solution passing through that point. We shall assume in the following that the rank of the Hessian matrix W is constant in all TQ and is $n - m_1$. If this is not the case, our considerations will only hold in an open region of TQ where this condition is satisfied.

A. The map FL

The fiber derivative of the Lagrangian is the application (FL) of the tangent bundle on the cotangent bundle T^*Q

$$\mathrm{FL}: TQ \to T^*Q,$$

given by $\mathrm{FL}(q,\dot{q}) = (q,p)$, where

$$p_i = \frac{\partial L}{\partial \dot{q}^i} \equiv \mathscr{P}_i(q,\dot{q}) \quad i = 1,\ldots,n. \tag{2.2}$$

We shall also assume that $\mathrm{FL}(TQ) = M_0 \subset T^*Q$ is a submanifold of T^*Q, locally defined by the constraints

$$\Phi_\mu^{(0)}(q,p) = 0, \quad \mu = 1,2,\ldots,m_1, \tag{2.3}$$

which are the primary constraints.

We also assume

$$\operatorname{rank}\left|\frac{\partial \Phi_\mu^{(0)}}{\partial p_j}\right| = m_1. \tag{2.4}$$

This condition excludes ineffective constraints at this level. In the following we will disregard Lagrangians that have ineffective constraints at any level.

The primary Hamiltonian constraints (2.3) are identified at the Lagrangian level, i.e.,

$$\Phi_\mu^{(0)}(q,\mathscr{P}(q,\dot{q})) \equiv 0, \tag{2.5a}$$

or equivalently

$$\mathrm{FL}^*\Phi_\mu^{(0)} \equiv 0, \tag{2.5b}$$

where FL* is the pullback application. From (2.5) we deduce

$$\frac{\partial \Phi_\mu^{(0)}}{\partial p_i}(q,\mathscr{P}(q,\dot{q}))\frac{\partial \mathscr{P}_i}{\partial \dot{q}_j} = 0, \tag{2.6}$$

and since $\partial \mathscr{P}_i / \partial \dot{q}^j$ is the Hessian matrix element W_{ij}, we have a basis for the null vectors of W:

$$\gamma_\mu^i = \mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial p^i} \quad \mu = 1,\ldots,m_1, \quad i = 1,\ldots,n. \tag{2.7}$$

A basis for the kernel of differential application FL* can be written in terms of (2.7):

$$\Gamma_\mu = \gamma_\mu^i(q,\dot{q})\frac{\partial}{\partial \dot{q}_i}. \tag{2.8}$$

A function $f \in \Lambda^0(\mathrm{TQ})$ is FL-projectable if there exists some function $g \in \Lambda^0(T^*Q)$ such that $f = \mathrm{FL}^*g$. The necessary and sufficient condition for $f \in \Lambda^0(\mathrm{TQ})$ to be FL-projectable is that[7,10]

$$\Gamma_\mu f = 0, \quad \mu = 1,\ldots,m_1. \tag{2.9}$$

The energy function $E_L(q\dot{q}) = \dot{q}^i \mathscr{P}_i(q\dot{q}) - L(q\dot{q})$ verifies condition (2.9). Therefore there exits a function $H_c \in \Lambda^0(T^*Q)$ such that

$$\mathrm{FL}^*H_c = E_L, \tag{2.10}$$

where H_c is only unambiguously defined in M_0.

Let us proceed to the inversion of the Legendre transformation FL. Given a point $(q_0,\dot{q}_0) \in \mathrm{TQ}$ and its image under $\mathrm{FL}(q_0,p_0)$, $p_0 = (q_0,\dot{q}_0)$, we have the identity

$$H_c(q_0,\mathscr{P}(q_0\dot{q}_0)) \equiv q_0^i \, \mathscr{P}_i(q_0,\dot{q}_0) - L(q_0,\dot{q}_0), \tag{2.11}$$

from which, taking the derivative with respect to $\dot{q}_j$, we obtain

$$\frac{\partial H_c}{\partial p_i}(q_0 p_0)\frac{\partial \mathscr{P}_i}{\partial \dot{q}_j}(q_0\dot{q}_0) = \frac{\partial \mathscr{P}_i}{\partial \dot{q}_j}(q_0\dot{q}_0)\dot{q}_0^i. \tag{2.12}$$

Therefore, $\dot{q}_0^i - (\partial H_c/\partial p)(q_0 p_0)$ is a null vector of W and can be written in terms of (2.7):

$$\dot{q}_0^i - \frac{\partial H_c}{\partial p_i}(q_0,p_0) = v_\mu^0 \, \gamma_\mu^i(q_0\dot{q}_0), \tag{2.13}$$

for some parameters v_μ^0. Note that $(q_0\dot{q}_0)$ is a particular point of anti-image $\mathrm{FL}^{-1}(q_0,p_0)$. The whole anti-image is a leaf of foliation, defined in TQ by the equivalence relation

$$x \sim x^1 \leftrightarrow \mathrm{FL}x = \mathrm{FL}x^1; \quad x,x^1 \in TQ. \tag{2.14}$$

Consequently, Ker FL_* at every point $x \in TQ$ is given by the elements of $T_x(TQ)$ tangent to the leaves of the foliation previously defined. In other words, these leaves are the integral surfaces of the vector fields belonging to Ker FL_*. This means that $\mathrm{FL}^{-1}(q_0 p_0)$ will be generated from the point $(q_0,\dot{q}_0)$ by the exponential map $e^{u_\mu \Gamma_\mu}$ with u_μ arbitrary parameters and Γ_μ the vector fields (2.8). Therefore,

$$\mathrm{FL}^{-1}(q_0 p_0) = \{q_0,\dot{q}(v)\}, \tag{2.15}$$

where

$$\dot{q}^i(v) = \frac{\partial H_c}{\partial p_i}(q_0,p_0) + v_\mu \frac{\partial \Phi_\mu^{(0)}}{\partial p_i}(q_0 p_0), \tag{2.16}$$

with the arbitrary parameters v_μ given by $v_\mu = v_\mu^0 + u_\mu$. Due to the condition (2.4) given a point of this leaf we can determine the parameters v^μ in terms of the coordinates of this point. This means that for a given point $(q,p) \in M_0$ and all its possible anti-images we have the relation

$$\dot{q}_i = \frac{\partial H_c}{\partial p_i}(q,p) + v_\mu(q\dot{q})\frac{\partial \Phi_\mu^{(0)}}{\partial p_i}(q,p). \tag{2.17}$$

If we now consider Eq. (2.17) as a system with (q,p) as data and $\dot{q}$ as unknowns, we show in Appendix A that there are no solutions if the data are out of M_0; whereas if the $(qp) \in M_0$, the solutions of (2.17) are obviously given by (2.16). Therefore we conclude that the relations (2.17) and (2.2) are equivalent. Therefore Eq. (2.17) is the inverse Legendre transformation; note that Eq. (2.17) is an implicit equation for $\dot{q}$.

Let us observe that the application of Γ to both sides of (2.17) gives

$$\Gamma_\mu v_\nu = \delta_{\mu\nu}. \tag{2.18}$$

This means that all the functions $v_\mu(q\dot{q})$ are not FL-projectable. However, we shall see in Sec. III that some of these functions admit a canonical form when restricted to a suitable submanifold of TQ.

Now if we take the derivative of (2.11) with respect to q_i, we have

$$-\frac{\partial L}{\partial q^i}(q\dot{q}) = \frac{\partial H_c}{\partial q^i}(qp) - \left(\dot{q}_j - \frac{\partial H_c}{\partial p_j}(qp)\right)\frac{\partial \mathscr{P}_j}{\partial q}(q,\dot{q}), \tag{2.19}$$

where $(qp) = \mathrm{FL}(q\dot{q})$. If we use Eq. (2.17) we have

$$-\frac{\partial L}{\partial q^i}(q\dot{q}) = \frac{\partial H_c}{\partial q^i}(qp) - v_\mu(q\dot{q})\frac{\partial \Phi_\mu^{(0)}}{\partial p_j}(qp)\frac{\partial \mathscr{P}_j}{\partial q^i}(q\dot{q}), \tag{2.20}$$

but since $\mathrm{FL}^*\Phi_\mu = 0$ we have

$$\mathrm{FL}^*\left(\frac{\partial \Phi_\mu}{\partial p_j}\right)\frac{\partial \mathscr{P}_j}{\partial q^i} + \mathrm{FL}^*\left(\frac{\partial \Phi_\mu}{\partial q^i}\right) = 0. \tag{2.21}$$

Therefore, Eq. (2.20) can be written as

$$-\frac{\partial L}{\partial q^i}(q\dot{q}) = \frac{\partial H_c}{\partial q^i}(q,p) + v_\mu(q\dot{q})\frac{\partial \Phi_\mu^{(0)}}{\partial q^i}(q,p). \tag{2.22}$$

Let us now consider the equations of motion.

B. Equations of motion

A curve q: $I\in R \to Q$ is a solution of the Euler–Lagrange equations (2.1) if the function $p(t)$ defined by

$$p(t) = \mathscr{P}\left(q(t), \frac{dq(t)}{dt}\right) \tag{2.23}$$

satisfies

$$\frac{dp}{dt} = \frac{\partial L}{\partial q}\left(q(t), \frac{dq}{dt}\right). \tag{2.24}$$

Due to the equivalence between Eqs. (2.2) and (2.17) we can write an expression equivalent to (2.23), i.e.,

$$\frac{dq}{dt} = \frac{\partial H_c}{\partial p}(q(t),p(t)) + v_\mu\left(q(t), \frac{dq(t)}{dt}\right)\frac{\partial \Phi_\mu^{(0)}}{\partial p}(q(t),p(t)), \tag{2.25}$$

also due to (2.22) we have

$$-\frac{\partial L}{\partial q}(q(t),\dot{q}(t)) = \frac{\partial H_c}{\partial q}(q(t),p(t)) + v_\mu\left(q(t), \frac{dq(t)}{dt}\right) \times \frac{\partial \Phi_\mu^{(0)}}{\partial q}(q(t),p(t)), \tag{2.26}$$

and using (2.24), Eq. (2.26) is written as

$$-\frac{dp}{dt} = \frac{\partial H_c}{\partial q}(q(t),p(t)) + v_\mu\left(q(t), \frac{dq(t)}{dt}\right)\frac{\partial \Phi_\mu^{(0)}}{\partial q}(q(t),p(t)). \tag{2.27}$$

Equations (2.25) and (2.27) are the Hamilton–Dirac equations for the singular Lagrangian $L(q\dot{q})$, so we can formulate the following theorem.

Theorem: If $q(t)$ is a solution of Euler–Lagrange equations (2.1) in configuration space, the lifting to T^*Q given by $(q(t),p(t))$ with $p(t)$ defined by (2.2), is a solution of the Hamilton–Dirac equations (2.25) and (2.27).

Furthermore, in the inverse sense, if $(q(t),p(t))$ verify Eq. (2.25) and Eq. (2.27), then (2.23) is satisfied because Eq. (2.23) is equivalent to (2.25). Furthermore, from (2.26) and (2.27) we obtain (2.24), therefore the following theorem holds.

Theorem 2: If $(q(t),p(t))$ is a solution of the Hamilton–Dirac equations (2.25) and (2.27), then $q(t)$ verifies the Euler–Lagrange equations (2.1).

If we consider the canonical symplectic structure of T^*Q we can write Eqs. (2.25) and (2.27) in terms of PB as

$$\frac{dq}{dt} = \{q,H_c\} + v_\mu\left(q, \frac{dq}{dt}\right)\{q, \Phi_\mu^{(0)}\}, \tag{2.28}$$

$$\frac{dp}{dt} = \{p,H_c\} + v_\mu\left(q, \frac{dq}{dt}\right)\{p, \Phi_\mu^{(0)}\}. \tag{2.29}$$

These equations are not written in the normal form, in the same sense as the Euler–Lagrange equations of motion, (2.1), therefore the possible solutions of those equations lie in a submanifold of T^*Q and the solution passing through a point of that submanifold is not necessarily unique.

Equations (2.28) and (2.39) can be written in a normal form if one introduces m_1 arbitrary functions of the evolution parameter $\lambda_\mu(t)$, and also imposes from the outset the primary constraints

$$\frac{dq}{dt} \underset{M_0}{=} \{q,H_c\} + \lambda_\mu(t)\{q, \Phi_\mu^{(0)}\}, \tag{2.30}$$

$$\frac{dp}{dt} \underset{M_0}{=} \{p,H_c\} + \lambda_\mu(t)\{p, \Phi_\mu^{(0)}\}, \tag{2.31}$$

where $\underset{M_0}{=}$ means weak equality on the surface M_0. Equations (2.31) are the standard Hamilton–Dirac equations.[1,2,4,5]

III. HAMILTONIAN FORMALISM

In the preceding section we assumed the existence of solutions of the equations of motion and we have shown the equivalence between the Lagrangian and Hamiltonian formalism. Now we study the submanifold where those solutions exist, we will use an iterative procedure. Let us begin with the Hamiltonian formalism, the Hamilton–Dirac equations of motion are Eqs. (2.28) and (2.29):

$$\begin{aligned} \frac{dq}{dt} &= \{q, H_c\} + v_\mu\left(q, \frac{dq}{dt}\right)\{q, \Phi_\mu^{(0)}\}, \\ \frac{dp}{dt} &= \{p, H_c\} + v_\mu\left(q, \frac{dq}{dt}\right)\{p, \Phi_\mu^{(0)}\}, \end{aligned} \tag{3.1}$$

where $\Phi_\mu^{(0)}$ are the primary Hamiltonian constraints and $v_\mu(q,\dot{q})$ are known function of q and $\dot{q}$. We know, from Appendix A, that (3.1) have only solutions if the initial conditions belong to the submanifold $M_0 \subset T^*Q$. In that case, a curve passing through a point of M_0 will be a solution of (3.1) if

$$\frac{d\,\Phi_\mu^{(0)}}{dt} \underset{M_0}{=} 0, \quad \mu = 1,\ldots,m_1, \tag{3.2a}$$

that is, the solution $(q(t),p(t))$ must belong entirely to M_0. In general, Eqs. (3.2) will be restrictions for the initial conditions. We write (3.2) as

$$0 \underset{M_0}{=} \{\Phi_\mu^{(0)},H_c\} + v_\nu(q,\dot{q})\{\Phi_\mu^{(0)}\ \Phi_\nu^{(0)}\}. \tag{3.2b}$$

To discuss the content of (3.2) it is necessary to know the rank of the PB matrix between the primary constraints, i.e.,

$$\operatorname{rank}|\{\Phi^{(0)}_\mu, \Phi^{(0)}_\nu\}| = m_1 - m_2, \quad \mu,\nu = 1,...,m_1, \tag{3.3}$$

which we assume to be constant on M_0. It is convenient to introduce an equivalent set of constraints

$$\Phi^{(0)}_{\mu_0}, \quad \mu_0 = 1,...,m_2, \tag{3.4}$$

$$\Phi^{(0)}_{\mu_0'}, \quad \mu_0' = 1,...,m_1 - m_2, \tag{3.5}$$

with the properties

$$0 \underset{M_0}{\neq} \det|\{\Phi^{(0)}_{\mu_0'}, \Phi^{(0)}_{\nu_0'}\}| \equiv \det C^{(1)}_{\mu_0'\nu_0'}, \tag{3.6}$$

$$\{\Phi^{(0)}_{\mu_0'}, \Phi^{(0)}_{\mu_0}\} \underset{M_0}{=} 0, \quad \{\Phi^{(0)}_{\mu_0}, \Phi^{(0)}_{\nu_0}\} \underset{M_0}{=} 0, \tag{3.7}$$

therefore $\Phi^{(0)}_{\mu_0}$, $\Phi^{(0)}_{\mu_0'}$ are first and second class, respectively, on M_0. Note that $\Phi^{(0)}_{\mu_0'}$ are $m_1 - m_2$ of the old primary constraints $\Phi^{(0)}_\mu$, instead $\Phi^{(0)}_{\mu_0}$ are linear combinations of them. If we consider Eq. (3.2) for the second-class constraints $\Phi^{(0)}_{\mu_0'}$, we obtain a canonical expression for the functions $v_{\nu_0'}(q,\dot q)$:

$$v_{\nu_0'}(q\dot q) \underset{M_0}{=} -(C^{(1)})^{-1}_{\nu_0'\mu_0'}\{\Phi_{\mu_0'}, H_c\}. \tag{3.8}$$

Therefore the evolution for a generic quantity $A(q,p)$ in M_0 is given by

$$\frac{dA}{dt} \underset{M_0}{=} \{A, H^{(1)}_c\} + v_{\nu_0}(q\dot q)\{A, \Phi^{(0)}_{\nu_0}\}, \quad \nu_0 = 1,...,m_2, \tag{3.9}$$

where

$$H^{(1)}_c \equiv H_c - \{H_c\, \Phi^{(0)}_{\mu_0'}\}(C^{(1)})^{-1}_{\mu_0'\nu_0'}\, \Phi^{(0)}_{\nu_0'},$$

$$\mu_0', \nu_0' = 1,...,m_1 - m_2, \tag{3.10}$$

with the properties[13]

$$\{\Phi^{(0)}_{\mu_0'}, H^{(1)}_c\} \underset{M_0}{=} 0. \tag{3.11}$$

The evolution of the first class constraints $\Phi^{(0)}_{\mu_0}$ is given by

$$\frac{d}{dt}\Phi^{(0)}_{\mu_0} \underset{M_0}{=} \{\Phi^{(0)}_{\mu_0}, H^{(1)}_c\} \equiv \Phi^{(1)}_{\mu_0}, \tag{3.12}$$

the stability conditions for $\Phi^{(1)}_{\mu_0}$ are

$$\Phi^{(1)}_{\mu_0} = 0. \tag{3.13}$$

If all these conditions are satisfied on M_0 the analysis is finished, if this is not the case Eqs. (3.13) are new restrictions on the initial conditions, which we call secondary constraints. Note that some of these constraints can be automatically satisfied on M_0, but in order to use a more compact notation we will continue to use the subscript μ_0 for all secondary constraints.

Let M_1, be the new submanifold defined by

$$\begin{aligned} &\Phi^{(0)}_{\mu_0'} = 0, \quad \mu_0' = 1,...,m_1 - m_2, \\ &\Phi^{(0)}_{\mu_0} = 0, \quad \Phi^{(1)}_{\mu_0} = 0, \quad \mu_0 = 1,...,m_2. \end{aligned} \tag{3.14}$$

A curve passing through a point of M_1 will be a solution of (3.1) if

$$\frac{d\,\Phi^{(1)}_{\mu_0}}{dt} \underset{M_1}{=} 0. \tag{3.15a}$$

These stability conditions can be written explicitly

$$0 \underset{M_1}{=} \{\Phi^{(1)}_{\mu_0}, H^{(1)}_c\} + v_{\nu_0(q,\dot q)}\{\Phi^{(1)}_{\mu_0}, \Phi^{(0)}_{\nu_0}\}. \tag{3.15b}$$

In Eq. (3.15) a PB matrix appears between the primary first-class constraints on M_0 and the secondary constraints. As is shown in Appendix B this matrix is symmetric:

$$\{\Phi^{(1)}_{\mu_0}, \Phi^{(0)}_{\nu_0}\} \underset{M_1}{=} \{\Phi^{(1)}_{\nu_0}, \Phi^{(0)}_{\mu_0}\}. \tag{3.16}$$

Let $m_2 - m_3$ be the rank of this matrix. Due to this symmetry property we can introduce a new set of constraints

$$\begin{aligned} &\Phi^{(0)}_{\mu_1} = 0, \quad \Phi^{(1)}_{\mu_1} = 0, \quad \mu_1 = 1,...,m_3, \\ &\Phi^{(0)}_{\mu_1'} = 0, \quad \Phi^{(1)}_{\mu_1'} = 0, \quad \mu_1' = 1,...,m_2 - m_3, \end{aligned} \tag{3.17}$$

with the following properties:

$$0 \underset{M_1}{\neq} \det|\{\Phi^{(1)}_{\mu_1'}, \Phi^{(0)}_{\nu_1'}\}| \equiv \det C^{(2)}_{\mu_1'\nu_1'}, \tag{3.18}$$

$$\{\Phi^{(1)}_{\mu_1}, \Phi^{(0)}_{\mu_0}\} \underset{M_1}{=} 0, \tag{3.19}$$

$$\{\Phi^{(1)}_{\mu_0}, \Phi^{(0)}_{\mu_1}\} \underset{M_1}{=} 0. \tag{3.20}$$

Note that $\Phi^{(0)}_{\mu_1'}$ are $m_2 - m_3$ of the old $\Phi^{(0)}_{\mu_0}$, due to the symmetry property, Eq. (3.16), and $\Phi^{(1)}_{\mu_1'}$ are also the same $m_2 - m_3$ of the $\Phi^{(1)}_{\mu_0}$. The $\Phi^{(0)}_{\mu_1}$ are a linear combination of the $\Phi^{(0)}_{\mu_0}$ and the $\Phi^{(1)}_{\mu_1}$ are the same linear combination of the $\Phi^{(1)}_{\mu_0}$. This means that we also have

$$\begin{aligned} &\Phi^{(1)}_{\mu_1} = \{\Phi^{(0)}_{\mu_1}, H^{(1)}_c\}, \\ &\Phi^{(1)}_{\mu_1'} = \{\Phi^{(0)}_{\mu_1'}, H^{(1)}_c\}, \end{aligned} \tag{3.21}$$

This means that the labeling of new constraints is compatible with their stability, therefore we have a sort of hereditary property. It should be noted that $\Phi^{(0)}_{\mu_1}$ are first-class constraints on M_1.

At this point, we consider the stability condition for the constraints $\Phi^{(1)}_{\nu_1'}$. From this we obtain a canonical expression for the functions $v_{\nu_1'}(q\dot q)$:

$$v_{\nu_1'}(q\dot q) \underset{M_1}{=} -(C^{(2)})^{-1}_{\nu_1'\mu_1'}\{\Phi^{(1)}_{\mu_1'}, H^{(1)}_c\}. \tag{3.22}$$

The evolution on M_1 is given by

$$\frac{dA}{dt} \underset{M_1}{=} \{A, H^{(2)}_c\} + v_{\nu_1}(q\dot q)\{A, \Phi^{(0)}_{\nu_1}\}, \tag{3.23}$$

where

$$H^{(2)}_c = H^{(1)}_c + \{H^{(1)}_c, \Phi^{(1)}_{\mu_1'}\}(C^{(2)})^{-1}_{\mu_1'\nu_1'}\, \Phi^{(0)}_{\nu_1'}, \tag{3.24}$$

with the properties

$$\{\Phi^{(1)}_{\mu_1'}, H^{(2)}_c\} \underset{M_1}{=} 0. \tag{3.25}$$

Now consider the stability of the remaining secondary constraints

$$\frac{d}{dt}\Phi^{(1)}_{\mu_1} \underset{M_1}{=} \{\Phi^{(1)}_{\mu_1}, H^{(2)}_c\} \equiv \Phi^{(2)}_{\mu_1}, \tag{3.26}$$

the relations $\Phi^{(2)}_{\mu_1} = 0$ can be satisfied on M_1 in which case the analysis is finished. Otherwise

$$\Phi^{(2)}_{\mu_1} = 0, \quad \mu_1 = 1,...,m_3, \tag{3.27}$$

are tertiary constraints. At this level the evolution is restricted to the submanifold M_2:

$$\Phi^{(0)}_{\mu'_0} = 0, \quad \Phi^{(0)}_{\mu'_1} = 0, \quad \Phi^{(0)}_{\mu_1} = 0,$$
$$\Phi^{(1)}_{\mu'_1} = 0, \quad \Phi^{(0)}_{\mu_1} = 0, \quad \Phi^{(2)}_{\mu_1} = 0, \tag{3.28}$$

with

$$\mu'_0 = 1,...,m_1 - m_2, \quad \mu'_1 = 1,...,m_2 - m_3, \quad \mu_1 = 1,...,m_3. \tag{3.29}$$

In order to study the stability of tertiary constraints $\Phi^{(2)}_{\mu_1}$, we need to consider the PB matrix of the primary first-class constraints on M_1, $\Phi^{(0)}_{\nu_1}$, with the tertiary constraints $\Phi^{(2)}_{\mu_1}$. As is shown in Appendix B, this matrix is antisymmetric in the submanifold M_2:

$$\{\Phi^{(2)}_{\mu_1}, \Phi^{(0)}_{\nu_1}\} \underset{M_2}{=} -\{\Phi^{(2)}_{\nu_1}, \Phi^{(0)}_{\mu_1}\}. \tag{3.30}$$

Let $m_3 - m_4$ be the rank of that matrix. Due to the antisymmetry property we can introduce a set of constraints

$$\Phi^{(2)}_{\mu_2}, \Phi^{(2)}_{\mu'_2}, \quad \mu_2 = 1,...,m_4,$$
$$\Phi^{(0)}_{\mu_2}, \Phi^{(0)}_{\mu'_2}, \quad \mu'_2 = 1,...,m_3 - m_4, \tag{3.31}$$

which define the same surface as the set $\Phi^{(0)}_{\mu_1}$, $\Phi^{(2)}_{\nu_1}$, with the properties

$$\det C^{(3)}_{\mu'_2\nu'_2} \equiv \det\{\Phi^{(2)}_{\mu'_2}, \Phi^{(0)}_{\nu'_2}\} \underset{M_2}{\neq} 0, \tag{3.32}$$

$$\{\Phi^{(2)}_{\mu_2} \Phi^{(0)}_{\mu_1}\} \underset{M_2}{=} 0, \quad \{\Phi^{(2)}_{\mu_1}, \Phi^{(0)}_{\mu_2}\} \underset{M_2}{=} 0, \tag{3.33}$$

$$\Phi^{(1)}_{\mu_2} = \{\Phi^{(0)}_{\mu_2}, H^{(1)}_c\}, \quad \Phi^{(1)}_{\mu'_2} = \{\Phi^{(0)}_{\mu'_2}, H^{(1)}_c\}, \tag{3.34}$$

$$\Phi^{(2)}_{\mu_2} = \{\Phi^{(1)}_{\mu_2} H^{(2)}_c\}, \quad \Phi^{(2)}_{\mu'_2} = \{\Phi^{(1)}_{\mu'_2}, H^{(2)}_c\}. \tag{3.35}$$

The stability conditions for the tertiary constraints $\Phi^{(2)}_{\mu'_2}$, as in the previous case, enables us to obtain a canonical expression for the functions $v_{\nu'_2}$. Using that expression the evolution on M_2 is given by

$$\frac{dA}{dt} \underset{M_2}{=} \{A, H^{(3)}_c\} + v_{\nu 2}\{A, \Phi^{(0)}_{\nu_2}\}, \tag{3.36}$$

where

$$H^{(3)}_c = H^{(2)}_c - \{H^{(2)}_c, \Phi^{(2)}_{\mu'_2}\} (C^{(3)})^{-1}_{\mu'_2\nu'_2} \Phi^{(0)}_{\nu'_2}, \tag{3.37}$$

with the property

$$\{\Phi^{(2)}_{\mu'_2}, H^{(3)}_c\} \underset{M_2}{=} 0. \tag{3.38}$$

With respect to the stability of $\Phi^{(2)}_{\mu_2}$ we have the relations

$$\Phi^{(3)}_{\mu_2} = 0, \tag{3.39}$$

where

$$\Phi^{(3)}_{\mu_2} \equiv \{\Phi^{(2)}_{\mu_2}, H^{(3)}_c\}. \tag{3.40}$$

If the relations (3.39) are verified on M_2, the analysis is finished. Otherwise we have more constraints and therefore we need further to require the stability of those constraints and the procedure continues as before. Let assume that our Lagrangian has a final submanifold M_f where we have solution of the equations of motion (3.1). We write the constraints defining M_f as

$$\begin{array}{lllllll}
\Phi^{(0)}_{\mu'_0} & \Phi^{(0)}_{\mu'_1} & \Phi^{(0)}_{\mu'_2} & \cdots & \Phi^{(0)}_{\mu'_f}\,\Phi^{(0)}_{\mu_f}, & \text{primary;} \\
 & \Phi^{(1)}_{\mu'_1} & \Phi^{(1)}_{\mu'_2} & \cdots & \Phi^{(1)}_{\mu'_f}\,\Phi^{(1)}_{\mu_f}, & \text{secondary;} \\
 & & \Phi^{(2)}_{\mu'_2} & \cdots & \Phi^{(2)}_{\mu'_f}\,\Phi^{(2)}_{\mu_f}, & \text{tertiary;} \\
 & & & & \Phi^{(f)}_{\mu'_f}\,\Phi^{(f)}_{\mu_f}, & f\text{-ary,}
\end{array} \tag{3.41}$$

where the hereditary property is manifest. The equations of motion on M_f are

$$\frac{dA}{dt} \underset{M_f}{=} \{A, H^{(f+1)}_c\} + v_{\nu_f}(q\dot{q})\{A\Phi^{(f)}_{\nu_f}\}, \tag{3.42}$$

where

$$H^{(f+1)} = H^{(f)}_c + (-)^{(f+1)}\{H^{(f)}_c, \Phi^{(f)}_{\mu'_f}\}(C^{(f)})^{-1}_{\mu'_f\nu'_f} \Phi^{(0)}_{\nu'_f}. \tag{3.43}$$

The matrix $C^{(f)}$ is symmetric or antisymmetric depending on whether f is odd or even.

We have

$$\{\Phi^{(f)}_{\mu'_f}, H^{(f+1)}_c\} \underset{M_f}{=} 0, \tag{3.44}$$

and since the analysis is finished we have

$$\{\Phi^{(f)}_{\mu_f}, H^{(f+1)}_c\} \underset{M_f}{=} 0. \tag{3.45}$$

Note that in the equations of motion there appear functions $v_{\mu_f}(q\dot{q})$ that are not determined canonically and are associated with the primary first class constraints on the submanifold M_f.

Let us now study the relation between the procedure of Dirac brackets (DB) for second class constraints,[1] and the procedure developed here. Let us begin with the case with no tertiary constraints. The DB with respect to second-class constraints $\Phi^{(0)}_{\mu'_0}$ $\Phi^{(0)}_{\mu'_1}$ $\Phi^{(1)}_{\mu'_1}$ can be constructed in two steps. First we construct the DB for the constraints $\Phi^{(0)}_{\mu'_0}$, i.e.,

$$\{A,B\}^{H_1} \equiv \{A,B\} - \{A\Phi^{(0)}_{\mu'_0}\} (D^{(1)})^{-1}_{\mu'_0\nu'_0} \{\Phi^{(0)}_{\nu'_0}, B\}, \tag{3.46}$$

where $(\Delta^{(1)})^{-1}$ is the inverse matrix of $D^{(1)}$ defined by

$$D^{(1)}_{\mu'_0\nu'_0} = \{\Phi^{(0)}_{\mu'_0}, \Phi^{(0)}_{\nu'_0}\}, \tag{3.47}$$

which coincides with $C^{(1)}_{\mu'_0\nu'_0}$ [Eq. (3.6)]. The final DB is written as

$$\{A,B\}^{H_2} \equiv \{A,B\}^{H_1} - \{A\chi_{\mu'_1}\}^{H_1} D^{(2)}_{\mu'_1\nu'_1} \{\chi_{\nu'_1}, B\}^{H_1}, \tag{3.48}$$

where $\chi_{\mu'_1}$ indicates any one of the constraints $\Phi^{(0)}_{\mu'_1}$, $\Phi^{(1)}_{\mu'_1}$, and $(D^{(2)})^{-1}$ is the inverse of the matrix $D^{(2)}$ defined by $D^{(2)}_{\mu'_1\nu'_1} = \{\chi_{\mu'_1}, \chi_{\nu'_1}\}^{H_1}$. Explicitly

$$D^{(2)} \underset{M_0}{=} \begin{pmatrix} 0 & -C^{(2)} \\ C^{(2)} & K \end{pmatrix}, \tag{3.49}$$

where $C^{(2)}$ is the matrix defined in (3.18) and K is a matrix constructed with the $\Phi^{(1)}_{\mu'_1}$ constraints.

Let us consider (3.47) in the case $B = H_c$. Using Eqs. (3.48), (3.12), and (3.24) we have

$$\{A,H_c\}^{H_2} \underset{M_0}{=} \{A,H_c\}^{H_1} - \{A\Phi^{(0)}_{\mu_1'}\}^{H_c}[(C^{(2)})K(C^{(2)})^{-1}]_{\mu_1'\nu_1'}\{\Phi^{(0)}_{\nu_1'}H_c\}^{H_1}$$

$$- \{A\Phi^{(0)}_{\mu_1'}\}^{H_1}(C^{(2)})^{-1}_{\mu_1'\nu_1'}\{\Phi^{(1)}_{\nu_1'},H_c\}^{H_1} + \{A\Phi^{(1)}_{\mu_1'}\}^{H_1}(C^{(2)})^{-1}_{\mu_1'\nu_1'}\{\Phi^{(0)}_{\mu_1'},H_c\}^{H_1} \underset{M_1}{=} \{A,H^{(2)}_c\}. \quad (3.50)$$

Therefore our procedure on the M_1 surface is equivalent to the Dirac procedure. In Appendix C we explicitly prove this result for the case of no quartiary constraints. In order to give a proof in the general case, we need to consider a more geometrical formulation that takes into account the new structures we have found. Work in that direction is in progress.

Summing up, Eqs. (3.42) are equivalent on M_f to the equations of motion generated by the total Dirac Hamiltonian:

$$H_T = H^{H(f+1)} + v_{\mu_f}\Phi^{(0)}_{\mu_f}, \quad (3.51)$$

where $H^{H(f+1)}$ is the starred Hamiltonian[13] with respect to all second-class constraints. Therefore the DB is not the minimal structure to obtain the Hamiltonian equations of motion.

IV. LAGRANGIAN FORMALISM: RELATION BETWEEN THE LAGRANGIAN AND HAMILTONIAN CONSTRAINTS

In the previous section we have built a new scheme for the construction and classification of the submanifold of the Hamiltonian constraints. Now, we shall use these results to do the same with the Lagrangian constraints.

Using the Hessian matrix W_{ij}, we can write the Lagrangian equations of motion (2.1) as

$$W_{ij}\ddot{q}^j = \alpha_i, \quad (4.1)$$

where

$$\alpha_i \equiv \frac{\partial L}{\partial q^i} - \dot{q}^j\frac{\partial^2 L}{\partial q^j\,\partial\dot{q}^i}. \quad (4.2)$$

If the rank of W is $n-m_1$, $m_1>0$, the Hessian will have m_1 null vectors $\gamma_\mu(q,\dot{q})$ such that

$$W_{ij}\gamma^j_\mu = 0. \quad (4.3)$$

The contraction of Eq. (4.1) with a null vector gives

$$\chi^{(1)}_\mu \equiv \alpha_i\gamma^i_\mu = 0. \quad (4.4)$$

This is the first generation of Lagrangian constraints:

$$\chi^{(1)}_\mu = 0, \quad \mu = 1,\dots,m_1. \quad (4.5)$$

The submanifold in TQ locally defined by the vanishing of the $\chi^{(1)}_\mu$ is denoted by S_1. These Lagrangian constraints can also be obtained with the help of the operator K:

$$K = \dot{q}^i\mathrm{FL}^*\frac{\partial}{\partial q^i} + \frac{\partial L}{\partial q^i}\mathrm{FL}^*\frac{\partial}{\partial p_i}, \quad (4.6)$$

which takes a function in $\Lambda^0(T^*Q)$, differentiates it with respect to time, and gives the result in $\Lambda^0(TQ)$. As we demonstrate in Appendix C, one has

$$K\Phi^{(0)}_\mu = \chi^{(1)}_\mu. \quad (4.7)$$

From this equation, we can see that every primary constraint produces a Lagrangian constraint of the first generation. Because of the linearity of the relations (4.4) and (2.7), the classification (3.5) enables us to make the splitting

$$\chi^{(1)}_{\mu_0}, \quad \mu_0 = 1,\dots,m_2, \quad (4.8a)$$

$$\chi^{(1)}_{\mu_0'}, \quad \mu_0' = 1,\dots,m_1-m_2. \quad (4.8b)$$

Now we can demonstrate the following relations:

$$\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0'},H_c\} + v_{\nu_0'}(q,\dot{q})\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0'},\Phi^{(0)}_{\nu_0'}\} = \Phi^{(0)}_{\mu_0'}, \quad (4.9a)$$

$$\mathrm{FL}^*\Phi^{(1)}_{\mu_0} = \chi^{(1)}_{\mu_0} \quad (\mathrm{mod}\ \chi^{(1)}_{\mu_0'}). \quad (4.9b)$$

First, we demonstrate (4.9a), using (2.17), (2.20), (D4), and (D5). We have

$$\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0'},H_c\} = \chi_{\mu_0'} - v_{\nu_0}(q,\dot{q})\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0'},\Phi^{(0)}_{\nu_0}\} - v_{\nu_0'}(q,\dot{q})\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0'},\Phi^{(0)}_{\nu_0'}\}, \quad (4.10)$$

but $\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0'},\Phi^{(0)}_{\mu_{\nu_0}}\} = 0$, due to the fact $\{\Phi^{(0)}_{\mu_0'},\Phi^{(0)}_{\nu_0}\} \underset{M_0}{=} 0$, so (4.9a) is demonstrated. Now we can demonstrate (4.9b). We have $\Phi^{(1)}_{\mu_0} = \{\Phi^{(0)}_{\mu_0},H^{(1)}_c\}$ with $H^{(1)}_c$ given by (3.12); using (2.17) and (2.20) we have

$$\mathrm{FL}^*\frac{\partial H^{(1)}_c}{\partial p_i} = \dot{q}^i - v_{\mu_0}(q\dot{q})\mathrm{FL}^*\frac{\partial\Phi\,\mu_0^{(0)}}{\partial p_i} \quad (\mathrm{mod}\ \Phi^{(1)}_{\mu_0'}), \quad (4.11a)$$

$$\mathrm{FL}^*\frac{\partial H^{(1)}_c}{\partial q^i} = -\frac{\partial L}{\partial q^i} - v_{\mu_0}(q,\dot{q})\mathrm{FL}^*\frac{\partial\Phi^{(0)}_{\mu_0}}{\partial q_i} \quad (\mathrm{mod}\ \chi^{(1)}_{\mu_0'}), \quad (4.11b)$$

and therefore

$$\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0},H^{(1)}_c\} = \chi^{(1)}_{\mu_0} - v_{\nu_0}(q,\dot{q})\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0},\Phi^{(0)}_{\nu_0}\} \quad (\mathrm{mod}\,\chi^{(1)}_{\mu_0'}) = \Phi^{(1)}_{\mu_0} \quad (\mathrm{mod}\ \chi^{(1)}_{\mu_0'})$$

as desired.

Furthermore from an analogous equation to (4.10) we have that

$$\mathrm{FL}^*\{\Phi^{(0)}_{\mu_0},H_c\} = \chi^{(1)}_{\mu_0}. \quad (4.12)$$

From Eq. (4.9a) we see that the constraints $\chi^{(1)}_{\mu_0'}$ are not FL-projectable, instead the constraints $\chi^{(1)}_{\mu_0}$ are FL-projectable [see Eq. (4.12)]. These results can also be seen using the relation

$$\Gamma_\mu\chi^{(1)}_\nu = \mathrm{FL}^*\{\Phi^{(0)}_\nu,\Phi^{(0)}_\mu\} \quad (4.13)$$

[see Appendix C and Eq. (2.9)]. Therefore the Lagrangian constraints $\chi^{(1)}_{\mu_0'}$ associated to the primary second-class constraints by means of the operator K are not FL-projectable. Instead those associated the primary first-class constraints are FL-projectable.

Now we want to investigate the stability of the constraints that locally define S_1. We have

$$\frac{d}{dt}\chi_\mu^{(1)} = \dot q^i \frac{\partial \chi_\mu^{(1)}}{\partial q^i} + \ddot q^i \frac{\partial \chi_\mu^{(1)}}{\partial \dot q^i}. \tag{4.14}$$

We can obtain $\ddot q$ from the equation of motion introducing the completeness relation (see A5)

$$\delta^i_j = M^{ik} W_{kj} + \tilde\gamma_{\mu j}\gamma^i_\mu, \tag{4.15}$$

we have

$$\ddot q^i \underset{S_1}{=} M^{ij}\alpha_j + \beta_\mu \gamma^i_\mu, \tag{4.16}$$

where

$$\beta_\mu = \tilde\gamma_{\mu k}\ddot q_k, \tag{4.17}$$

are the accelerations that are undetermined by the equations of motion. Substitution of (4.16) in (4.14) gives

$$\frac{d}{dt}\chi_\mu^{(1)} \underset{S_1}{=} D^{(0)}\chi_\mu^{(1)} + \beta_\nu \Gamma_\nu \chi_\mu^{(1)}, \tag{4.18}$$

where

$$D^{(0)} \equiv \alpha_j M^{ji}\frac{\partial}{\partial \dot q^i} + \dot q^i \frac{\partial}{\partial q^i}. \tag{4.19}$$

If we consider the stability, Eq. (4.14), of the non-FL-projectable constraints, $\chi^{(0)}_{\mu_0'}$, we have

$$0 \underset{S_1}{=} D^{(0)}\chi^{(1)}_{\mu_0'} + \beta_{\nu_0'}\Gamma_{\nu_0'}\chi^{(1)}_{\mu_0'}, \tag{4.20}$$

with

$$\Gamma_{\nu_0'}\chi^{(1)}_{\mu_0'} = \mathrm{FL}^*(C^{(1)})_{\mu_0'\nu_0'}, \tag{4.21}$$

due to (4.13) and (3.6). Since $C^{(1)}$ has an inverse one can determine the undetermined accelerations $\beta_{\nu_0'}$ as a function of q and $\dot q$:

$$\beta_{\nu_0'} \underset{S_1}{=} -\mathrm{FL}^*[(C^1)^{-1}_{\nu\mu_0'}]D^{(0)}\chi^{(1)}_{\mu_0'}. \tag{4.22}$$

Let us now consider the stability of the FL-projectable constraints. We have

$$\frac{d}{dt}\chi^{(1)}_{\mu_0} \underset{S_1}{=} D^{(0)}\chi^{(1)}_{\mu_0'} \equiv \chi^{(2)}_{\mu_0}. \tag{4.23}$$

If the relations $\chi^{(2)}_{\mu_0} = 0$ are automatically verified in S_1, the analysis is finished. Otherwise the $\chi^{(2)}_{\mu_0}$ are the second generation of the Lagrangian constraints, which together with $\chi^{(1)}_\mu$ define the surface S_2.

Now it is necessary to study the stability of $\chi^{(2)}_{\mu_0}$. It is possible to show that

$$K\Phi^{(1)}_{\mu_0} \underset{S_1}{=} \chi^{(2)}_{\mu_0}, \tag{4.24}$$

where K is the operator defined in (4.6). So the $\chi^{(2)}_{\mu_0}$ are associated with the Hamiltonian constraints $\Phi^{(1)}_{\mu_0}$. Remembering the splitting (3.17), it follows that we have the following splitting at Lagrangian level:

$$\begin{aligned} &\chi^{(1)}_{\mu_1}, \quad \chi^{(2)}_{\mu_1}, \quad \mu_1 = 1,\dots,m_3,\\ &\chi^{(1)}_{\mu_1'}, \quad \chi^{(2)}_{\mu_1'}, \quad \mu_1' = 1,\dots,m_2 - m_3. \end{aligned} \tag{4.25}$$

We see that labeling is compatible with the stability.

Now we can show

$$\mathrm{FL}^*\{\Phi^{(1)}_{\mu_1'}, H^{(1)}_c\} + v_{\nu_1'}(q,\dot q)\mathrm{FL}^*\{\Phi^{(1)}_{\mu_1'}, \Phi^{(0)}_{\nu_1'}\} \underset{S_1}{=} \chi^{(2)}_{\mu_0'}, \tag{4.26a}$$

$$\mathrm{FL}^*\Phi^{(2)}_{\mu_1} \underset{S_1}{=} \chi^{(2)}_{\mu_1} \quad (\mathrm{mod}\ \chi^{(2)}_{\mu_1'}). \tag{4.26b}$$

The proof of (4.26) is more intricate than the first level and is not given here. One can also show

$$\mathrm{FL}^*\{\Phi^{(1)}_{\mu_1}, H^{(1)}_c\} \underset{S_1}{=} \chi^{(2)}_{\mu_2}. \tag{4.27}$$

From Eq. (4.26a) we see that the constraints $\chi^{(2)}_{\mu_1'}$ are not FL-projectable, instead the constraints $\chi^{(2)}_{\mu_1}$ are FL-projectable on the surface S_1, [see Eq. (4.27)]. We call these objects weakly FL-projectable on S_1.

Furthermore, using the results of Appendix C we have

$$\Gamma_{\mu_0}\chi^{(2)}_{\mu_1'} = \mathrm{FL}^*\{\Phi^{(1)}_{\mu_1'}, \Phi^{(0)}_{\mu_0}\} \underset{S_1}{\neq} 0, \tag{4.28}$$

instead of

$$\Gamma_{\mu_0}\chi^{(2)}_{\mu_1} = \mathrm{FL}^*\{\Phi^{(1)}_{\mu_1}, \Phi^{(0)}_{\mu_0}\} \underset{S_1}{=} 0. \tag{4.29}$$

This analysis suggests that the necessary and sufficient condition for a function f to be weakly FL-projectable on S_1 is

$$\Gamma_{\mu_0} f \underset{S_1}{=} 0, \quad \mu_0 = 1,\dots,m_1, \tag{4.30}$$

where Γ_{μ_0} are the vector vectors fields of Ker FL tangent to S_1. This result is proved in a separate paper.

Now we need to require the stability of $\chi^{(2)}_{\mu_0}$

$$\begin{aligned} \frac{d}{dt}\chi^{(2)}_{\mu_0} &\underset{S_2}{=} (D^0 + \beta_{\nu_0'}\Gamma_{\nu_0'})\chi^{(2)}_{\mu_0} + \beta_{\nu_0}\Gamma_{\nu_0}\chi^{(2)}_{\mu_0}\\ &\equiv D^{(1)}\chi^{(2)}_{\mu_0} + \beta_{\nu_0}\Gamma_{\nu_0}\chi^{(2)}_{\mu_0}. \end{aligned} \tag{4.31}$$

If we consider the stability condition for $\chi^{(2)}_{\mu_1'}$, we can express the undetermined accelerations $\beta_{\nu_1'}$ in terms of the coordinates and velocities:

$$\beta_{\nu_1'} \underset{S_2}{=} \mathrm{FL}^*[(C^{(2)})^{-1}_{\nu_1'\mu_1'}]D^{(1)}\chi^{(2)}_{\mu_1'}, \tag{4.32}$$

due to the nonsingular character of the matrix $C^{(2)}$, Eq. (3.18). Let us now consider the constraints $\chi^{(2)}_{\mu_1}$, we have

$$\frac{d}{dt}\chi^{(2)}_{\mu_1} \underset{S_2}{=} D^{(1)}\chi^{(2)}_{\mu_1} \equiv \chi^{(3)}_{\mu_1}. \tag{4.33}$$

If the relations $\chi^{(3)}_{\mu_1}$, are automatically verified in S_2 the analysis is finished. Otherwise $\chi^{(3)}_{\mu_1}$ are the third generation of the Lagrangian constraints and the procedure continues. At Hamiltonian level we have assumed the existence of a submanifold M_f where we can have solutions. This implies that the relations

$$\{\Phi^{(f)}_{\mu_f}, H^{(f+1)}\} \underset{M_f}{=} 0 \tag{4.34}$$

are identities on M_f. At the Lagrangian level we have

$$K\Phi^{(f)}_{\mu_f'} \underset{S_f}{=} \chi^{(f+1)}_{\mu_f'}, \tag{4.35}$$

$$K\Phi^{(f)}_{\mu_f} \underset{S_f}{=} 0, \tag{4.36}$$

and

$$\mathrm{FL}^*\{\Phi^{(f)}_{\mu_f'}, H^{(f)}_c\} + v_{\nu_f'}(q,\dot q)\{\Phi^{(f)}_{\mu_f'}\ \Phi^{(f)}_{\mu_f'}\} \underset{S_f}{=} \chi^{(f+1)}_{\mu_f'},$$

$$\mathrm{FL}^*\{\Phi_{\mu_f}^{(f)} H_c^{(f+1)}\} \underset{S_f}{=} 0.$$

This means that we have no more weakly FL-projectable constraints. However, we have the non-FL-projectable constraints $\chi_{\mu',f}^{(f+1)}$. If we consider the stability of these constraints, we can obtain the undetermined $\beta_{\mu_f'}$ acceleration in terms of the coordinate and momenta. At this point the analysis is finished.

Summing up, at every level a (weakly) FL-projectable constraint on a certain submanifold comes from the stability of a Hamiltonian constraint of the preceding level, which is first class with respect to the primary Hamiltonian constraint, while a non-FL-projectable constraint comes from the stability of a Hamiltonian constraint which converts a primary constraint to the second class. Also, if a certain number of velocities are canonically determined at a given level, the same number of accelerations are determined at the next level.

V. CONCLUSIONS

The equivalence between the Lagrangian and Hamiltonian formalism for constrained systems has been proved, in the sense that given a solution $q(+)$ of Euler–Lagrange equations of motion, the functions $q(t)$ and $p(t) = \mathscr{P}(q(t)[dq(t)/dt])$ are solutions of the Hamiltonian–Dirac equations of motion and vice versa. Note that neither of these equations is in normal form. This means that we can only have solutions in a submanifold of the respective space. These submanifolds are constructed through an interactive procedure. At the Hamiltonian level, our procedure differs from the standard one. All constraints are classified according to whether or not they are first class with respect to the primary constraints. We have seen that PB matrix of the primary first-class constraints on M_0 and the secondary, tertiary, ... constraints are either symmetric or antisymmetric. This implies that our final Hamiltonian $H_c^{(f+1)}$ differs from the starred Hamiltonian of Komar and Bergman, but on the final submanifold M_f they both yield the same evolution.

At the Lagrangian level, we have seen that the Lagrangian constraints can be obtained from the stability of the Hamiltonian constraints using the K operator (4.6). Furthermore, the Lagrangian constraints that are FL-projectable or weakly FL-projectable are the Lagrangian counterparts of the Hamiltonian constraints, which are first class or second class with respect to the primary Hamiltonian constraints. In fact at every level, a (weakly) FL-projectable constraint on a certain submanifold comes from the stability of a Hamiltonian constraint of the preceding level, which is first class with respect to the primary Hamiltonian constraint, while a non-FL-projectable constraint comes from the stability of a Hamiltonian constraint that converts a primary constraint to the second class. Also, if a certain number of velocities are canonically determined at a given level, the same number of accelerations are determined at the next level.

ACKNOWLEDGMENTS

We are grateful to Professor Kenneth Rafanelli for a careful reading of the manuscript. One of us (J.G.) would like to acknowledge Professor B. Sakita for the hospitality at the High Energy Group of City College.

APPENDIX A: COMMENTS ON THE INVERSE LEGENDRE TRANSFORMATION

Here we demonstrate that Eqs. (2.17),

$$\dot{q} = \frac{\partial H_c}{\partial p_i}(q,p) + v_\mu(q\dot{q})\frac{\partial \Phi_\mu^{(0)}}{\partial p_i}(q,p), \tag{A1}$$

for $\dot{q}$ in terms of q and p, have solutions only if (q,p) belongs to the submanifold M_0 of the primary constraints. We know that when (q,p) belongs to M_0, the solutions are given by (2.16). Therefore, we have the identities

$$\dot{q}^i \equiv \mathrm{FL}^* \frac{\partial H_c}{\partial p_i} + v_\mu(q\dot{q})\mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial p_i}. \tag{A2}$$

Now at fixed q we consider an infinitesimal displacement $p + dp$ from a point $(q,p)\in M_0$. We want to know if there exists a solution of (A1), $\dot{q} + d\dot{q}$, derived continuously from the solution, $\dot{q}$, with data $(q,p)\in M_0$. If such a solution exists, the following identities must be verified:

$$d\dot{q}^i = \mathrm{FL}^* \frac{\partial^2 H_c}{\partial p_i\, \partial p_j} dp_j + \frac{\partial v_\mu(q,\dot{q})}{\partial \dot{q}^j} \mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial p_i} d\dot{q}^j + v_\mu(q,\dot{q})\mathrm{FL}^* \frac{\partial^2 \Phi_\mu^{(0)}}{\partial p_i\, \partial p_j} dp_j. \tag{A3}$$

In order to study when these identities are verified, let us consider the completeness relation

$$\delta^i_j = \tilde{\gamma}_{\mu j}\gamma^i_\mu + M^{ik}W_{kj}, \tag{A4}$$

where

$$\tilde{\gamma}_{\mu j} = \frac{\partial v_\mu(q,\dot{q})}{\partial \dot{q}^j}, \quad \gamma^i_\mu = \mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial p_i},$$
$$M^{ik} = \mathrm{FL}^* \frac{\partial^2 H_c}{\partial p_i\, \partial p_k} + v_\mu(q,\dot{q})\mathrm{FL}^* \frac{\partial^2 \Phi_\mu}{\partial p_i\, \partial p}, \tag{A5}$$
$$W_{kj} = \frac{\partial^2 L}{\partial \dot{q}_k\, \partial q_j}.$$

Equation (A4) can be obtained by taking the derivative of (A2) with respect to q_j. Note that M^{ik} is not unambiguously defined in $\Lambda^0(T^*Q)$. This is due to the ambiguity of the definition of H_c out of the surface M_0.

The change

$$H_c(q,p) \to H'_c(q,p) = H_c(q,p) + \lambda_\mu(q,p)\Phi_\mu^{(0)}(q,p)$$

with λ_μ arbitrary describes the arbitrariness of the Hamiltonian. This change produces a new definition of the functions v_μ of (2.17),

$$v'_\mu(q,\dot{q}) = v_\mu(q,\dot{q}) - \mathrm{FL}^*\lambda_\mu, \tag{A6}$$

and consequently a change in the matrices M^{ik},

$$M'^{ik} = M^{ik} + \mathrm{FL}^* \frac{\partial \lambda_\mu}{\partial p_i}\gamma^k_\mu + \mathrm{FL}^* \frac{\partial \lambda_\mu}{\partial p_k}\gamma^i_\mu. \tag{A7}$$

Equation (A3) can be written [using (A4) and (A5)] as

$$M^{ik}(W_{kj}d\dot{q}^j - d\,\mathrm{FL}^*p_k) = 0, \tag{A8}$$

where M^{ik} is any element of the family (A7). Since in that family there always exists a nonsingular matrix, Eq. (A8) is verified only if

$$W_{kj}d\dot{q}^j = d\,\mathrm{FL}^*p_k. \tag{A9}$$

The necessary and sufficient condition for (A9) to be fulfilled is that

$$\gamma^k_\mu d\,\mathrm{FL}^*p_k = 0, \tag{A10}$$

where γ^k_μ are the null vectors of W, Eq. (A5). Therefore, we have

$$\mathrm{FL}^*d\Phi_\mu = 0, \tag{A11}$$

and therefore we conclude that the only displacement $p \to p + dp$ that allow changes on the solutions $\dot{q} \to \dot{q} + d\dot{q}$ are those made on the surface M_0.

APPENDIX B: STRUCTURE OF THE CONSTRAINTS POISSON BRACKET

In Sec. III we stated that the matrix of PB between a subset of primary first-class constraints on M_0 and the secondary, tertiary, ... constraints was either symmetric or antisymmetric. Here we demonstrate this property explicitly. Let us begin with the matrix of PB between the primary constraints $\Phi^{(0)}_{\mu_0}$ and the secondary constraints $\Phi^{(1)}_{\mu_0}$. Taking into account the definition of $\Phi^{(1)}_{\mu_0}$, the Jacobi identity, and the first-class character of $\Phi^{(0)}_{\mu_0}$ on M_0 we have

$$\begin{aligned}\{\Phi^{(1)}_{\mu_0},\Phi^{(0)}_{\nu_0}\} &= \{\{\Phi^{(0)}_{\mu_0},H^{(1)}_c\},\Phi^{(0)}_{\nu_0}\} \underset{M_0}{=} \{\{\Phi^{(0)}_{\mu_0},H_c\},\Phi^{(0)}_{\nu_0}\}\\ &= -\{\{\Phi^{(0)}_{\nu_0},\Phi^{(0)}_{\mu_0}\},H_c\} - \{\{H_c,\Phi^{(0)}_{\nu_0}\},\Phi^{(0)}_{\mu_0}\}\\ &\underset{M_0}{=} \{\Phi^{(1)}_{\nu_0},\Phi^{(0)}_{\mu_0}\} - \{\{\Phi^{(0)}_{\nu_0},\Phi^{(0)}_{\mu_0}\},H_c\}.\end{aligned}$$

Consider the last term of Eq. (B1), since the $\Phi^{(0)}_{\mu_0}$'s are first class on M_0 then

$$\{\{\Phi^{(0)}_{\mu_0},\Phi^{(0)}_{\nu_0}\}\Phi^{(0)}_{\nu_0'}\} \underset{M_0}{=} 0,$$

this implies that

$$\{\Phi_{\mu_0},\Phi^{(0)}_{\nu_0}\} = O(\Phi^{(0)}_{\mu_0}) + O^2(\Phi^{(0)}_{\mu_0'}),$$

where $O(\Phi^{(0)}_{\mu_0})$ is a function that contains a term linear in $\Phi^{(0)}_{\mu_0}$, and $O^2(\Phi^{(0)}_{\mu_0'})$ is a function of $\Phi^{(0)}_{\mu_0}$ that contains a term quadratic in $\Phi^{(0)}_{\mu_0'}$ as the lowest-order term. This means that the last term of Eq. (B1) vanishes on the surface M_1, therefore (B1) becomes

$$\{\Phi^{(1)}_{\mu_0},\Phi^{(0)}_{\nu_0}\} \underset{M_1}{=} \{\Phi^{(1)}_{\nu_0},\Phi^{(0)}_{\mu_0}\}.$$

This means that the above matrix is symmetric with respect to the interchange of μ_0 and ν_0 on M_1.

Let us now consider the matrix of the PB between the primary first class constraints on M_2 and the tertiary constraints. We have

$$\begin{aligned}\{\Phi^{(2)}_{\mu_1},\Phi^{(0)}_{\nu_1}\} &= \{\{\Phi^{(1)}_{\mu_1},H^{(2)}_c\}\Phi^{(0)}_{\nu_1}\} \underset{M_1}{=} \{\{\Phi^{(1)}_{\mu_1},H^{(1)}_c\}\Phi^{(0)}_{\nu_1}\}\\ &= -\{\{\Phi^{(0)}_{\nu_1},\Phi^{(1)}_{\mu_1}\}H^{(1)}_c\} + \{\Phi^{(1)}_{\nu_1},\Phi^{(1)}_{\mu_1}\}\\ &= -\{\{\Phi^{(0)}_{\nu_1},\Phi^{(1)}_{\mu_1}\}H^{(1)}_c\}\\ &\quad - \{H^{(1)}_c\{\Phi^{(1)}_{\nu_1},\Phi^{(0)}_{\mu_1}\}\} - \{\Phi^{(0)}_{\mu_1}\{H^{(1)}_c,\Phi^{(1)}_{\nu_1}\}\}\\ &\underset{M_1}{=} -\{\{\Phi^{(0)}_{\nu_1},\Phi^{(1)}_{\mu_1}\}H^{(1)}_c\}\\ &\quad + \{\{\Phi^{(1)}_{\nu_1},\Phi^{(0)}_{\mu_1}\}H^{(1)}_c\} - \{\Phi^{(2)}_{\nu_1},\Phi^{(0)}_{\mu_1}\}.\end{aligned}$$

The PB $\{\Phi^{(0)}_{\mu_1},\Phi^{(1)}_{\nu_1}\}$ vanishes on M_1, furthermore we have

$$\begin{aligned}&\{\{\Phi^{(0)}_{\mu_1},\Phi^{(1)}_{\nu_1}\}\Phi^{(1)}_{\mu_1}\}\\ &= -\{\{\Phi^{(0)}_{\mu_1'},\Phi^{(0)}_{\mu_1}\}\Phi^{(1)}_{\nu_1}\} - \{\{\Phi^{(1)}_{\nu_1}\Phi^{(0)}_{\mu_1'}\}\Phi^{(0)}_{\mu_1}\} \underset{M_1}{=} 0,\end{aligned}$$

therefore

$$\{\Phi^{(0)}_{\mu_2},\Phi^{(1)}_{\nu_1}\} = O(\Phi^{(0)}_{\mu_0'} + O(\Phi^{(0)}_{\mu_1'}) + O(\Phi^{(0)}_{\mu_1}) + O^2(\Phi^{(0)}_{\mu_1'}),$$

which implies

$$\{\{\Phi^{(0)}_{\nu_1},\Phi^{(1)}_{\mu_1}\}H^{(1)}_c\} \underset{M_2}{=} 0.$$

Equation (B5) becomes

$$\{\Phi^{(2)}_{\mu_1},\Phi^{(0)}_{\nu_1}\} \underset{M_2}{=} -\{\Phi^{(2)}_{\nu_1},\Phi^{(0)}_{\mu_1}\},$$

which means that this matrix of PB's is antisymmetric. This antisymmetric property is due to the twofold application of the Jacobi identity. In the general case for the primary first class constraints on M_k and the K-ary constraints, we will have

$$\{\Phi^{(k)}_{\mu_k},\Phi^{(0)}_{\nu_k}\} \underset{M_k}{=} (-1)^{k+1}\{\Phi^{(k)}_{\nu_k},\Phi^{(0)}_{\mu_k}\}.$$

APPENDIX C: RELATION TO DIRAC BRACKET

We want to study the relation between the procedure developed in the text and the standard Dirac bracket formalism for the second class constraints when there are no quartiary constraints. In this case the second class constraints are $\Phi^{(0)}_{\mu_0'},\Phi^{(0)}_{\mu_1'},\Phi^{(0)}_{\mu_2'},\Phi^{(1)}_{\mu_1'},\Phi^{(1)}_{\mu_2'},\Phi^{(2)}_{\mu_2'}$, if we use the Dirac bracket with respect to $\Phi^{(0)}_{\mu_0'}\ \Phi^{(0)}_{\mu_1'}\ \Phi^{(1)}_{\mu_1'}$, Eq. (3.48), we only need to consider the constraints $\Phi^{(0)}_{\mu_2'},\Phi^{(1)}_{\mu_2'},\Phi^{(2)}_{\mu_2'}$. Let $\chi_{\mu_2'}$ denote anyone of those constraints and matrix $D^{(3)}_{\mu_2'\nu_2'} \equiv \{\chi_{\mu_2'}\chi_{\nu_2'}\}^{H_2}$. Using the relations

$$\{\Phi^{(0)}_{\mu_2'},\Phi^{(2)}_{\nu_2'}\}^{H_2} \underset{M_1}{=} \{\Phi^{(0)}_{\mu_2'},\Phi^{(2)}_{\nu_2}\}, \tag{C1}$$

we have

$$D^{(3)} \underset{M_1}{=} \begin{pmatrix} 0 & 0 & -C^{(3)} \\ 0 & -C^{(3)} & -B \\ C^{(3)} & B & Q \end{pmatrix} \tag{C2}$$

where $C^{(3)}$ is given by Eq. (3.32). The inverse matrix $D^{(3)-1}$ is given by

$$D^{(3)-1} = \begin{pmatrix} M & C^{(3)-1}BC^{(3)-1} & C^{(3)-1} \\ -C^{(3)-1}BC^{(3)-1} & -C^{(3)-1} & 0 \\ -C^{(3)-1} & 0 & 0 \end{pmatrix}, \tag{C3}$$

where

$$M = C^{(3)-1}QC^{(3)-1} + C^{(3)-1}BC^{(3)-1}BC^{(3)-1}. \tag{C4}$$

At this point we can write the final Dirac bracket

$$\{A,B\}^{H_3} = \{A,B\}^{H_2} - \{A,\chi_{\mu_2'}\}^{H_2} D^{(3)-1}_{\mu_2'\nu_2'} \{\chi_{\nu_2'},B\}^{H_2}.$$

If we consider (C5) for the case $B = H_c$, we obtain

$$\{A,H_c\}^{H_3} \underset{M_2}{=} \{A,H_c^{(3)}\}.$$

Therefore the evolution on M_2 with the Dirac bracket formalism coincides with our procedure.

APPENDIX D: PROPERTIES OF THE OPERATOR K

Here we want to show that the operator K,

$$K\colon\ \Lambda^0(T^*Q) \to \Lambda^0(TQ), \tag{D1}$$

given by

$$K = \dot q^i \mathrm{FL}^* \frac{\partial}{\partial q^i} + \frac{\partial L}{\partial q^i} \mathrm{FL}^* \frac{\partial}{\partial p_i}, \tag{D2}$$

applied to the primary Hamiltonian constraints $\Phi_\mu^{(0)}$ produces the first generation of Lagrangian constraints $\chi_\mu^{(0)}$. In fact

$$K\Phi_\mu^{(0)} = \dot q^i \mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial q^i} + \frac{\partial L}{\partial q^i} \mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial p_i}, \quad \mu = 1,\ldots,m_1, \tag{D3}$$

and from Eqs. (2.7) and (2.21) we have

$$\mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial p_i} = \gamma_\mu^i, \tag{D4}$$

$$\mathrm{FL}^* \frac{\partial \Phi_\mu^{(0)}}{\partial q^i} = -\gamma^j_\mu \frac{\partial^2 L}{\partial \dot q^j\, \partial q^i} = -\Gamma_\mu \frac{\partial L}{\partial q^i}, \tag{D5}$$

therefore

$$K\Phi_\mu^{(0)} = \gamma_\mu^i \left(\frac{\partial L}{\partial q^i} - \frac{\partial^2 L}{\partial \dot q^i\, \partial q^j} \dot q^j \right) = \chi_\mu^{(1)}. \tag{D6}$$

In general, the operator K applied to a function $g \in \Lambda^0(T^*Q)$ gives its temporal derivative expressed in $\Lambda^0(TQ)$, which we denote by $f_g(q,\dot q)$. Let us now study the projectability of f_g:

$$(\Gamma_\mu f_g) = \left(\mathrm{FL}^* \frac{\partial g}{\partial q^i}\right) \Gamma_\mu q^i + \left(\mathrm{FL}^* \frac{\partial g}{\partial q^i}\right) \Gamma_\mu \frac{\partial L}{\partial q^i}. \tag{D7}$$

Using (D4) and (D5) we obtain

$$\Gamma_\mu f_g = \mathrm{FL}^* \{g, \Phi_\mu^{(0)}\}. \tag{D8}$$

Therefore $f_g \leftarrow \Lambda^0(TQ)$ will be FL-projectable if $g \leftarrow \Lambda^0(TQ)$ is a first class function with respect to the primary first class constraints on M_0. In particular using (D8) we have

$$\Gamma_\mu \chi_\nu^{(1)} = \mathrm{FL}^* \{\Phi_\nu^{(0)}, \Phi_\mu^{(0)}\}, \tag{D9}$$

which states only that the Lagrangian constraints associated with the first class primary constraints are FL-projectable. Another consequence of (D8) is

$$\Gamma_{\mu_0} \chi_{\nu_0}^{(2)} = \mathrm{FL}^* \{\Phi_{\nu_0}^{(1)}, \Phi_{\mu_0}^{(0)}\},$$

which tells us that the only Lagrangian constraints associated with the secondary Hamiltonian constraints which are first class with respect to the primary first-class constraints on M_0 are weakly FL-projectable.

[1]P. A. M. Dirac, Can. J. Math. 2, 129 (1950); Proc. R. Soc. London Ser. A 246, 326 (1958).

[2]J. L. Anderson and P. G. Bergmann, Phys. Rev. 83, 1018 (1951); P. G. Bergmann and I. Goldberg, Phys. Rev. 98, 531 (1955).

[3]It is impossible to quote here all relevant contributions. Please see references to the original literature in the quoted review papers.

[4]E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A Modern Perspective* (Wiley, New York, 1974).

[5]K. Sundermeyer, "Constrained dynamics," in *Lecture Notes in Physics*, Vol. 169 (Springer, Berlin, 1982).

[6]See, for example, D. Dominici, G. Longhi, J. Gomis and J. M. Pons, J. Math. Phys. 25, 2439 (1984).

[7]G. Marmo, N. Mukunda, and J. Samuel, Riv. Nuovo Cimento 6, 4 (1983).

[8]J. Llosa (Editor), "Relativistic action at a distance: Classical and quantum aspects," *Lecture Notes in Physics*, Vol. 162, edited by J. Llosa (Springer, Berlin, 1982).

[9]H. Sazdjian, Phys. Lett. B 156, 381 (1985).

[10]M. Gotay and J. M. Nester, Ann. Inst. H. Poincarè 30, 129 (1979).

[11]K. Kamimura, Nuovo Cimento B 69, 33 (1982).

[12]P. G. Bergmann and A. Komar, Phys. Rev. Lett. 4, 432 (1960).

[13]R. Skinner and R. Rusk, J. Math. Phys. 24, 2589 (1983).

Chebyshev polynomials and quadratic path integrals

Patrick L. Nash
Division of Earth and Physical Sciences, The University of Texas at San Antonio, San Antonio, Texas 78285-0663

(Received 25 March 1986; accepted for publication 14 August 1986)

A simple method for the evaluation of path integrals associated with quadratic Lagrangians is discussed. This approach makes use of a relationship between the Van Vleck–Morette determinant and a limit that involves the Chebyshev polynomials of the second kind.

I. INTRODUCTION

In this brief paper we would like to point out a connection between the Van Vleck–Morette determinant of a path integral over a quadratic Lagrangian in one spatial dimension, and the Chebyshev polynomials of the second kind $U_N(x)$. Let the Lagrangian be given by

$$L = (m/2)\{\dot{x}^2 - \omega^2 x^2\}\,. \tag{1}$$

Following Schulman[1] [see Eqs. (6.16) and (6.17) on p. 34], we write the propagator as

$$\begin{aligned} G(x_b,t_b;x_a,t_a) &= \exp\{iS(x_b,t_b;x_a,t_a)\} \\ &\quad\times \lim_{N\to\infty}\left(\frac{m}{2\pi i\epsilon}\right)^{1/2}\int_{-\infty}^{\infty}\prod_{n=1}^{N}(\pi^{-1/2}\,dy_n) \\ &\quad\times \exp\left\{-\sum_{j=0}^{N}\left[(y_{j+1}-y_j)^2-\epsilon^2\omega^2 y_j^2\right]\right\}, \end{aligned} \tag{2}$$

where $\epsilon = (t_b - t_a)/(N+1)$ and $y_0 = 0 = y_{N+1}$. To evaluate the N-dimensional integral we first define a vector y in N dimensions

$$\tilde{y} = (y_1,\ldots,y_N) \tag{3}$$

(the tilde denotes transpose; our y is denoted as η by Schulman), and a real symmetric $N\times N$ matrix T_N such that

$$\tilde{y}T_N y = \sum_{j=0}^{N}(y_{j+1}-y_j)^2\,.$$

As is well known, and as Schulman explicitly shows, T_N has diagonal matrix elements equal to 2, matrix elements equal to -1 directly above and below the main diagonal, and matrix elements equal to zero elsewhere. We define a real symmetric $N\times N$ matrix $T_N(x)$ by replacing the diagonal matrix elements of T_N, namely 2, by $2x$. We have $T_N(1) = T_N$. We note that the argument of the exponential in the integrand of Eq. (2) can be written as

$$\begin{aligned} -\sum_{j=0}^{N}\left[(y_{j+1}-y_j)^2-\epsilon^2\omega^2 y_j^2\right] &= -\tilde{y}T_N y + \epsilon^2\omega^2\tilde{y}y = -\tilde{y}\{T_N(1)-\epsilon^2\omega^2 I\}y \\ &= -\tilde{y}T_N(1-\epsilon^2\omega^2/2)y. \end{aligned}$$

Hence the propagator of Eq. (2) is given by

$$\begin{aligned} G(x_b,t_b;x_a,t_a) &= \exp[iS(x_b,t_b;x_a,t_a)] \\ &\quad\times\lim_{N\to\infty}\{m/2\pi i\epsilon\det[T_N(1-\epsilon^2\omega^2/2)]\}^{1/2}\,. \end{aligned} \tag{4}$$

We shall prove that $\det T_N(x) = U_N(x)$, so that in this simple case the Van Vleck–Morette determinant involves the $\lim_{N\to\infty}\epsilon U_N(1-\epsilon^2\omega^2/2)$.

II. CHEBYSHEV POLYNOMIALS OF THE SECOND KIND

As discussed in the previous section, we define a real symmetric $N\times N$ matrix according to

$$T_N(x) = \begin{pmatrix} 2x & -1 & 0 & 0 & \\ -1 & 2x & -1 & 0 & \\ 0 & -1 & 2x & -1 & \\ 0 & 0 & -1 & 2x & \ddots \\ & & & \ddots & \ddots \end{pmatrix}. \tag{5}$$

Let

$$D_N(x) = \det T_N(x)\,. \tag{6}$$

We find that $D_N(x) = 2xD_{N-1}(x) - D_{N-2}(x)$, a result that follows easily upon expanding $\det T_N(x)$ along the first row. Letting $N-1\to N$, we record this result as

$$D_{N+1}(x) - 2xD_N(x) + D_{N-1}(x) = 0\,. \tag{7}$$

Moreover, Eqs. (5)–(7) imply that

$$D_N(1) = N+1\,. \tag{8}$$

Hence, by Eqs. (7) and (8) we deduce that the $D_N(x)$ are the Chebyshev polynomials of the second kind $U_N(x)$ (see Ref. 2) $D_N(x) = U_N(x)$.

The $U_N(x)$ may be defined as

$$U_N(x) = \sin(N+1)\theta/\sin\theta\,, \tag{9}$$

where $x = \cos\theta$ (see Ref. 2). For the problem at hand, $x = 1-\epsilon^2\omega^2/2$, so that $\theta \approx \epsilon\omega$. We see that

$$\lim_{N\to\infty}\epsilon U_N(1-\epsilon^2\omega^2/2) = \sin[\omega(t_b-t_a)]/\omega\,.$$

Thus we find that

$$\begin{aligned} G(x_b,t_b;x_a,t_a) &= [m\omega/2\pi i\sin\{\omega(t_b-t_a)\}]^{1/2} \\ &\quad\times\exp\{iS(x_b,t_b;x_a,t_a)\}\,. \end{aligned}$$

Although we have obtained nothing new, the reader may find some appeal in the directness and simplicity of this approach that utilizes the Chebyshev polynomials of the second kind.

[1] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley–Interscience, New York, 1981).
[2] *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw–Hill, New York, 1953), Vol. 2, pp. 183–187.

 0022-2488/86/122963-01$02.50

Parametrization of the linear zeros of $6j$ coefficients

J. J. Labarthe
Laboratoire Aimé Cotton, Centre National de la Recherche Scientifique II, Bâtiment 505, F-91405 Orsay Cedex, France

(Received 16 January 1986; accepted for publication 28 July 1986)

The linear zeros of $6j$ coefficients are fully parametrized apart from a multiplicative factor in terms of four integers.

I. INTRODUCTION

Zeros of $6j$ have been studied recently by means of Diophantine equations.[1-4] Racah's expression for the $6j$ is an alternating polynomial multiplied by a normalizing factor. Equating the polynomial to zero gives an equation to be solved in integers (Diophantine equation). The case that has been studied mainly is that of *linear zeros* (also called zeros of degree 1, or of weight 1), in which the polynomial is a sum of two terms. Some results have been obtained also in the case of zeros of degree 2 in which the polynomial is a sum of three terms.[3]

In this paper we present a complete parametrization of the linear zeros of the $6j$.

Rather than using the angular momenta, we use the decomposition of $6j$ in terms of closed diagrams.[5] The corresponding theory has been developed for coupling–recoupling coefficients ($3j,6j,9j,12j,\ldots$) in general. It allows one to obtain a formula of the same form as Racah's $3j$ and $6j$ formulas for any coupling–recoupling coefficients. Since this theory was presented in a rather abstract way, we specialize it in Sec. II for $6j$, in which case the closed diagrams are the same as the extremal elements.

II. DECOMPOSITION OF $6j$ IN TERMS OF EXTREMAL ELEMENTS

Let E denote the set of arrays $\left[\begin{smallmatrix} a & b & c\\ d & e & f\end{smallmatrix}\right]$ formed of integers or half-integers that satisfy the triangle conditions of the $6j\left\{\begin{smallmatrix} a & b & c\\ d & e & f\end{smallmatrix}\right\}$. If

$$x=\begin{bmatrix} a & b & c\\ d & e & f\end{bmatrix},\quad x'=\begin{bmatrix} a' & b' & c'\\ d' & e' & f'\end{bmatrix}$$

we define

$$x+x'=\begin{bmatrix} a+a' & b+b' & c+c'\\ d+d' & e+e' & f+f'\end{bmatrix},$$

$$\lambda x=\begin{bmatrix} \lambda a & \lambda b & \lambda c\\ \lambda d & \lambda e & \lambda f\end{bmatrix}.$$

It is easily seen that if $x\in E$, $x'\in E$, $\lambda\in N$, then $x+x'\in E$ and $\lambda x\in E$. In other words E is closed under addition and under multiplication by a non-negative integer.

Let us call an element $u\in E$ *reducible* if it can be written as a sum $u=x+y$ of nonzero elements $x, y\in E$. An *extremal element* is a nonzero element of E that is not reducible.

There are seven external elements in E, namely,

$$e_1=\begin{bmatrix} \frac12 & \frac12 & 0\\ \frac12 & \frac12 & 0\end{bmatrix},\quad e_2=\begin{bmatrix} 0 & \frac12 & \frac12\\ 0 & \frac12 & \frac12\end{bmatrix},\quad e_3=\begin{bmatrix} \frac12 & 0 & \frac12\\ \frac12 & 0 & \frac12\end{bmatrix},$$
$$e_4=\begin{bmatrix} \frac12 & \frac12 & 0\\ 0 & 0 & \frac12\end{bmatrix},\quad e_5=\begin{bmatrix} 0 & \frac12 & \frac12\\ \frac12 & 0 & 0\end{bmatrix},\quad e_6=\begin{bmatrix} \frac12 & 0 & \frac12\\ 0 & \frac12 & 0\end{bmatrix},$$
$$e_7=\begin{bmatrix} 0 & 0 & 0\\ \frac12 & \frac12 & \frac12\end{bmatrix}.$$

Every $x\in E$ can be decomposed over extremal elements

$$x=\sum_{i=1}^{7}\alpha_i e_i,\quad \alpha_i\in N, \tag{1}$$

but since the extremal elements are linked by the relation

$$e_1+e_2+e_3=e_4+e_5+e_6+e_7, \tag{2}$$

this decomposition is not unique, in general. However, the number of different decompositions is always finite.

Now we rewrite with our notation Racah's formula for the value of the $6j$ corresponding to the array $x\in E$:

$$\frac{1}{N}\sum\frac{(-)^{|\alpha|}(|\alpha|+1)!}{[\alpha!]}, \tag{3}$$

where the sum runs over all the different decompositions of x over extremal elements, $[\alpha!]=\alpha_1!\alpha_2!\cdots\alpha_7!$, $|\alpha|=\Sigma_{i=1}^{7}\alpha_i$, and N is a normalization factor, the explicit form of which is not useful for finding the zeros of the $6j$.

Let us mention briefly how the symmetries of the $6j$ (including Regge symmetries) come about in this picture. They correspond to permutations of extremal elements within (e_1,e_2,e_3) and (e_4,e_5,e_6,e_7). These 6×24 permutations leave Eqs. (2) and (3) unchanged.

One of these symmetries apart, the $6j$ for which sum (3) contains two terms have arrays with two possible decompositions:

$$\begin{bmatrix} (A+C+T-1)/2 & (A+B+S-1)/2 & (B+C+S+T-2)/2\\ (B+T)/2 & (C+S)/2 & (A+S+T-1)/2\end{bmatrix}$$
$$=e_1+Se_2+Te_3+(A-1)e_4+(B-1)e_5+(C-1)e_6 \tag{4a}$$
$$=(S-1)e_2+(T-1)e_3+Ae_4+Be_5+Ce_6+e_7, \tag{4b}$$

 0022-2488/86/122964-02$02.50

where A, B, C, S, and T are integers $\geqslant 1$.

Such a $6j$ has value zero if and only if

$$ABC = ST(A + B + C + S + T). \tag{5}$$

Remarks: (1) Various formulas other than Eq. (3) are known for the value of the $6j$, so the definition of a zero of degree k (when the polynomial part of the formula is a sum of $k+1$ terms) depends on the formula. However, for the formulas (22.1b,v,g,d) of Jucys and Bandzaitis[6] it can be checked that the number of terms of the polynomials is always greater than or equal to the number of terms in Racah's formula. So a zero of degree k for such a formula is a zero of lesser or equal degree for Racah's formula. Racah's formula is then simpler; for example, a linear zero for Racah's formula can appear as a zero of degree 2 for another formula. Let us notice, however, that there is always a symmetry of the $6j$ such that the degrees become equal.

(2) Equation (5) is the same as Eq. (4b) of Brudno and Louck[1]. They found too that the variables S,T,A,B, and C were the most convenient ones.

(3) The general coupling–recoupling coefficient is expressed by a formula[5] similar to Eq. (3) that can be used to study the zeros of these higher coupling–recoupling coefficients. We only examined the case of $9j$ of degree 1, but nothing really new results since these $9j$ reduce to $6j$ or $3j$ coefficients by formulas like Eq. (24.16) or (25.14) of Jucys and Bandzaitis.[6]

III. PARAMETRIZATION OF THE LINEAR ZEROS

Equation (5) is homogeneous so that from a given solution (A,B,C,S,T) we can generate a *ray* of solutions $(\lambda A,\lambda B,\lambda C,\lambda S,\lambda T)$ obtained by multiplication by

$$\lambda = p/\gcd(A,B,C,S,T), \tag{6}$$

where p is an integer and $\gcd(\cdots)$ designates the greatest common divisor. The general solution is obtained by finding one solution on each ray and then by multiplying by the factor λ.

Theorem: All solutions of Eq. (5) are obtained exactly once by multiplying the factor λ with the parametrized solution

$$\begin{aligned} A &= (ab - st)a, \\ B &= (ab - st)b, \\ C &= st(a + b + s + t), \\ S &= (ab - st)s, \\ T &= (ab - st)t, \end{aligned} \tag{7}$$

where the strictly positive integers a,b,s,t are relatively prime, and such that $ab > st$.

Proof: Let us start with any solution (A',B',C',S',T') of Eq. (5). Then we can write

$$A' = qa, \quad B' = qb, \quad S' = qs, \quad T' = qt, \tag{8}$$

where $q = \gcd(A',B',S',T')$, and a,b,s,t are relatively prime. Rewriting Eq. (5) as

$$(ab - st)C' = qst(a + b + s + t),$$

and multiplying the solution given by Eq. (8) by $(ab - st)/q$, we obtain another solution on the same ray. This solution is that given by Eq. (7). This proves that all solutions are obtained by multiplying (6) and (7). Conversely it is easily checked that different sets of a,b,s,t,λ give rise to solutions of Eq. (5), which are different because of the condition $\gcd(a,b,s,t) = 1$, thus proving the theorem. $\square$

IV. CONCLUDING REMARKS

(1) The simplicity of the parametrization has been reached at the expense of symmetry: Equation (7) is not symmetric in the permutations of A, B, and C. A symmetric parametrization seems much more difficult to obtain.[1] This can be compared with the famous Pythagorean equation $x^2 + y^2 = z^2$. A classical theorem[7] states that the general solution with integers x, y, z is given by multiplying by a factor the parametrization

$$x = r^2 - s^2, \quad y = 2rs, \quad z = r^2 + s^2,$$

where r and s are integers. This parametrization is not symmetric in x and y.

(2) From Brudno and Louck[1] we have that Eq. (5) for rational numbers is equivalent to a system of Diophantine equations

$$X + Y + Z = U + V + W,$$
$$X^3 + Y^3 + Z^3 = U^3 + V^3 + W^3.$$

The parametrization (7) gives then a full set of solutions of this system as (apart from a multiplicative factor)

$$\begin{aligned} X &= ab(a + b + 2s) + st(-s + t), \\ Y &= ab(a + b + 2t) + st(s - t), \\ Z &= ab(a - b) - st(2a + s + t), \\ U &= ab(a + b) - st(2a + 2b + s + t), \\ V &= ab(a - b) + st(2b + s + t), \\ W &= ab(a + b + 2s + 2t) - st(s + t). \end{aligned}$$

[1] S. Brudno and J. D. Louck, J. Math. Phys. **26**, 2092 (1985).
[2] K. Srinivasa Rao, J. Math. Phys. **26**, 2260 (1985).
[3] S. Brudno and J. D. Louck, J. Math. Phys. **26**, 1125 (1985).
[4] S. Brudno, J. Math. Phys. **26**, 434 (1985).
[5] J. J. Labarthe, J. Phys. A **8**, 1543 (1975).
[6] A. Jucys and A. Bandzaitis, *Teoriia Momenta Kolichestva Dvizheniia v Kvantovoi Mekhanike* (Mintis, Vilnius, 1965).
[7] L. J. Mordell, *Diophantine Equations* (Academic, New York, 1969).

Recurrence relations for two-center harmonic oscillator integrals

J. Morales, L. Sandoval, and A. Palma
Instituto Mexicano Del Petroleo, Investigacion Basica de Procesos, Apartado Postal 14-805, 07730 Mexico, D. F. Mexico

(Received 19 March 1986; accepted for publication 14 August 1986)

General recurrence relations for the calculation of two-center harmonic oscillator (HO) integrals are obtained by means of a hypervirial-like-theorem commutator algebra procedure, combined with a second quantization formalism. The method is based on a linear transformation between the creation and annihilation operators of two displaced HO with different frequencies. Ansbacher's recurrence relations for the calculation of Franck–Condon factors are obtained straightforwardly from the proposed general recurrence relations. The application to polynomial, exponential, and Gaussian operator integrals is shown and new recurrence relations are given. In all cases, the proposed recurrence relations reduce, as particular cases, to the corresponding formulas for the calculation of one-center integrals.

I. INTRODUCTION

Closed formulas for two-center matrix elements of quantum mechanical operators in the harmonic oscillator (HO) representation can be evaluated in a number of different analytical or algebraic ways. Whether these are obtained by direct integration with wave functions or indirectly by the use of the generating function,[1] the Hermite polynomials are explicitly or implicitly involved. As an example, Morales *et al.*[2] have proposed an algebraic approach based on the combined use of Cauchy's integral formula for a complex variable and the Baker–Campbell–Hausdorff theorem to determine closed formulas for the evaluation of matrix elements of polynomials operators in the HO representation; a more general closed formulation for arbitrary operators will follow in a forthcoming publication. However, despite the aesthetic advantage of the closed form expressions, in practice, they are sometime cumbersome to use making the recurrence relations desirable. In this aspect, we have proposed recently[3] an algebraic procedure, based on the hypervirial theorem and ladder operators, which is useful in the determination of general recurrence relations in the calculation of one-center HO integrals of arbitrary operators. The results thus obtained are given in terms of eigenenergies and potential parameters without the explicit use of the eigenfunctions. In the present work, the idea expressed in Ref. 3 is extended to the determination of the appropriate general recurrence relations for the calculation of two-center HO matrix elements; this is displayed in Sec. II. The subsequent sections are devoted to the application of the proposed general recurrence relations to the determination of the corresponding expressions to overlap, polynomials, exponential, and Gaussian operators. New recursion formulas are given and are shown to be very easy to handle.

II. GENERAL FORMULATION

In order to determine general recurrence relations for calculation of two-center HO integrals, let us consider two displaced HO's with different frequencies (ω_E,ω_G) and respective Hamiltonians:

$$\hat{H}_G = (\hat{a}_G^+ \hat{a}_G + \tfrac{1}{2})\alpha_G , \tag{2.1}$$

and

$$\hat{H}_E = (\hat{a}_E^+ \hat{a}_E + \tfrac{1}{2})\alpha_E , \tag{2.2}$$

where $\alpha_G = \hbar\omega_G$ and $\alpha_E = \hbar\omega_E$. Here G and E refer to the ground $G\langle\ |$ and excited $|\ \rangle E$ states, and $\hat{a}_G(\hat{a}_G^+)$ $[\hat{a}_E^+(\hat{a}_E)]$ are the creation (annihilation) operators with the properties

$$\hat{a}_E^+ |n\rangle_E = \sqrt{n+1}\, |n+1\rangle_E , \quad \hat{a}_E|n\rangle_E = \sqrt{n}|n-1\rangle_E , \tag{2.3}$$

$$\begin{aligned} {}_G\langle m|\hat{a}_G &= \sqrt{m+1}\ {}_G\langle m+1|, \\ {}_G\langle m|\hat{a}_G^+ &= \sqrt{m}\ {}_G\langle m-1| . \end{aligned} \tag{2.4}$$

In the most general case, these two HO's are centered at different equilibrium positions $(\hat{X}_E,\hat{X}_G)$ and have different force contants. However, the positions for both oscillators are related to each other by $\hat{X}_G - \hat{X}_E = l$. Consequently, for the derivation of recurrence relations useful for practical calculations, the spatial displacement and different frequencies can be treated simultaneously by using the linear transformation between the ladder operators $\{\hat{a}_G,\hat{a}_G^+,1\}$ for the ground state and those corresponding to the excited state $\{\hat{a}_E,\hat{a}_E^+,1\}$, this relation is conveniently given by the ordered expressions[4]

$$(1+\beta^2)\hat{a}_G = \sqrt{2}\beta^2\gamma + (1-\beta^2)\hat{a}_G^+ + 2\beta\hat{a}_E \tag{2.5}$$

and

$$(1+\beta^2)\hat{a}_E^+ = -\sqrt{2}\beta\gamma - (1-\beta^2)\hat{a}_E + 2\beta\,\hat{a}_G^+ , \tag{2.6}$$

where β and γ are the spectroscopic constants given by

$$\beta = (\mu_E\omega_E/\mu_G\omega_G)^{1/2} \text{ and } \gamma = (\mu_G\omega_G/\hbar)^{1/2}\, l ,$$

where $\mu_E(\mu_G)$ is the corresponding mass.

A. Right-hand recurrence relation ($n>m$)

Let $f(\hat{a}_E,\hat{a}_E^+) = f_E$ be a function that may be expanded in power series in $\hat{a}_E$ and $\hat{a}_E^+$ that satisfy the commutation relation $[\hat{a}_E,\hat{a}_E^+] = 1$; the choosing of $f(\hat{a}_G,\hat{a}_G^+)$ leads to

 0022-2488/86/122966-07$02.50

similar results. It can be shown[5] that

$$[\hat{a}_E, f(\hat{a}_E^+, \hat{a}_E)] = \frac{\partial f_E}{\partial \hat{a}_E^+}, \tag{2.7}$$

$$[\hat{a}_E^+, f(\hat{a}_E^+, \hat{a}_E)] = -\frac{\partial f_E}{\partial \hat{a}_E}. \tag{2.8}$$

An hypervirial-like-theorem commutator algebra[6] leads to

$$[\hat{H}_E, f(\hat{a}_E^+, \hat{a}_E)] = \alpha_E \left\{ \hat{a}_E^+ \frac{\partial f_E}{\partial \hat{a}_E^+} - \frac{\partial f_E}{\partial \hat{a}_E} \hat{a}_E \right\}, \tag{2.9a}$$

and

$$[\hat{H}_E, f(\hat{a}_E^+, \hat{a}_E)] = \alpha_E \hat{a}_E^+ \hat{a}_E f_E + f_E\{\tfrac{1}{2}\alpha_E - H_E\}. \tag{2.9b}$$

The normal ordered expression given in Eq. (2.6), combined with the commutation relation

$$\hat{a}_E \frac{\partial f_E}{\partial \hat{a}_E^+} = \frac{\partial^2 f_E}{\partial \hat{a}_E^+ \partial \hat{a}_E^+} + \frac{\partial f_E}{\partial \hat{a}_E^+} \hat{a}_E$$

and the properties specified in Eqs. (2.3) and (2.4), transforms Eq. (2.9a) into

$$\begin{aligned}
&[(1+\beta^2)/\alpha_E]\, {}_G\langle m|[H_E, f_E]|n\rangle_E \\
&\quad = -\sqrt{2}\beta\gamma\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E + 2\beta\sqrt{m}\, {}_G\langle m-1| \frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E - (1-\beta^2)\, {}_G\langle m| \frac{\partial^2 f_E}{\partial \hat{a}_E^+ \partial \hat{a}_E^+}|n\rangle_E \\
&\qquad - (1-\beta^2)\sqrt{n}\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E^+}|n-1\rangle_E - (1+\beta^2)\sqrt{n}\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E}|n-1\rangle_E .
\end{aligned} \tag{2.10}$$

In a similar way, the identity

$$\begin{aligned}
(1+\beta^2)\, \hat{a}_E^+ a_E f_E = {} & -\sqrt{2}\beta\gamma \frac{\partial f_E}{\partial \hat{a}_E^+} - \sqrt{2}\beta\gamma f_E \hat{a}_E + 2\beta\, \hat{a}_G^+ \frac{\partial f_E}{\partial \hat{a}_E^+} \\
& + 2\beta\, \hat{a}_G^+ f_E \hat{a}_E - (1-\beta^2) \frac{\partial^2 f_E}{\partial \hat{a}_E^+ \partial \hat{a}_E^+} - 2(1-\beta^2) \frac{\partial f_E}{\partial \hat{a}_E^+} \hat{a}_E - (1-\beta^2) f_E \hat{a}_E \hat{a}_E
\end{aligned}$$

transforms Eq. (2.9b) to

$$\begin{aligned}
[(1+\beta^2)/\alpha_E]\, {}_G\langle m|[H_E, f_E]|n\rangle_E = {} & -\sqrt{2}\beta\gamma\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E - \beta\gamma\sqrt{2n}\, {}_G\langle m|f_E|n-1\rangle_E \\
& + 2\beta\sqrt{m}\, {}_G\langle m-1| \frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E + 2\beta\sqrt{mn}\, {}_G\langle m-1|f_E|n-1\rangle_E \\
& - (1-\beta^2)\, {}_G\langle m| \frac{\partial^2 f_E}{\partial \hat{a}_E^+ \partial \hat{a}_E^+}|n\rangle_E - 2(1-\beta^2)\sqrt{n}\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E^+}|n-1\rangle_E \\
& - (1-\beta^2)\sqrt{n(n-1)}\, {}_G\langle m|f_E|n-2\rangle_E - (1+\beta^2) n\, {}_G\langle m|f_E|n\rangle_E .
\end{aligned} \tag{2.11}$$

The general right-hand recurrence relation for the calculation of HO integrals leads trivially, when matching Eq. (2.10) and Eq. (2.11), to

$$\begin{aligned}
{}_G\langle m|f(\hat{a}_E^+, \hat{a}_E)|n\rangle_E = {} & -\frac{\beta\gamma}{1+\beta^2}\sqrt{\frac{2}{n}}\, {}_G\langle m|f_E|n-1\rangle_E + \frac{2\beta}{1+\beta^2}\sqrt{\frac{m}{n}}\, {}_G\langle m-1|f_E|n-1\rangle_E \\
& - \frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{n-1}{n}}\, {}_G\langle m|f_E|n-2\rangle_E - \frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{1}{n}}\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E^+}|n-1\rangle_E \\
& + \frac{1}{\sqrt{n}}\, {}_G\langle m| \frac{\partial f_E}{\partial \hat{a}_E}|n-1\rangle_E, \quad m, m \neq 0, \quad n > m .
\end{aligned} \tag{2.12}$$

B. Left-hand recurrence relation ($m > n$)

The twin general recurrence relation of Eq. (2.12) is similarly obtained from

$$\frac{2\beta}{1+\beta^2}[H_G, f_E] = \hat{a}_G^+ \frac{\partial f_E}{\partial \hat{a}_E^+} - \frac{1-\beta^2}{1+\beta^2}\hat{a}_G^+ \frac{\partial f_E}{\partial \hat{a}_E} - \frac{\partial f_E}{\partial \hat{a}_E}\hat{a}_G + \frac{1-\beta^2}{1+\beta^2}\frac{\partial f_E}{\partial \hat{a}_E^+}\hat{a}_G, \tag{2.13}$$

and

$$[H_G, f_E] = (H_G - \alpha_G/2) f_E - \alpha_G f_E\, \hat{a}_G^+ \hat{a}_G . \tag{2.14}$$

In order to avoid the operator $\hat{a}_G$ at the right side of the f_E derivative, we can use, in Eq. (2.13), the identities

$$\frac{\partial f_E}{\partial \hat{a}_E}\hat{a}_G=\hat{a}_G\frac{\partial f_E}{\partial \hat{a}_E}+\frac{1-\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E}-\frac{1+\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E}, \tag{2.15}$$

and

$$\frac{\partial f_E}{\partial \hat{a}_E^+}\hat{a}_G=\hat{a}_G\frac{\partial f_E}{\partial \hat{a}_E^+}+\frac{1-\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E^+}-\frac{1+\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E^+}, \tag{2.16}$$

along with the ordered expression (2.5). Thus, the use of the commutator relations

$$\hat{a}_E\frac{\partial f_E}{\partial \hat{a}_E^+}=\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E^+}+\frac{\partial f_E}{\partial \hat{a}_E^+}\hat{a}_E, \tag{2.17}$$

$$\hat{a}_E\frac{\partial f_E}{\partial \hat{a}_E}=\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E}+\frac{\partial f_E}{\partial \hat{a}_E}\hat{a}_E, \tag{2.18}$$

with the properties specified in Eqs. (2.3) and (2.4), allow us to obtain

$$\begin{aligned}
&\frac{1}{\alpha_E}\,{}_G\langle m|[H_G,f_E]|n\rangle_E\\
&\quad=\frac{1+\beta^4}{\beta(1+\beta^2)}\sqrt{m}\,{}_G\langle m-1|\frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E-\frac{1-\beta^2}{\beta}\sqrt{m}\,{}_G\langle m-1|\frac{\partial f_E}{\partial \hat{a}_E}|n\rangle_E\\
&\qquad+\frac{\beta\gamma(1-\beta^2)}{\sqrt{2}(1+\beta^2)}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E-\frac{\beta\gamma}{\sqrt{2}}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E}|n\rangle_E+\left(\frac{1-\beta^2}{2\beta}\right)^2{}_G\langle m|\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E^+}|n\rangle_E\\
&\qquad-\frac{(1-\beta^2)^3}{4\beta^2(1+\beta^2)}\,{}_G\langle m|\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E^+}|n\rangle_E-\frac{1-\beta^4}{4\beta^2}\,{}_G\langle m|\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E}|n\rangle_E\\
&\qquad+\left(\frac{1-\beta^2}{2\beta}\right)^2{}_G\langle m|\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E}|n\rangle_E+\frac{1-\beta^2}{1+\beta^2}\sqrt{n}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E^+}|n-1\rangle_E-\sqrt{n}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E}|n-1\rangle_E.
\end{aligned} \tag{2.19}$$

In the same way, the identity

$$\begin{aligned}
f_E\hat{a}_G^+\hat{a}_G&=\frac{\sqrt{2}\beta^2\gamma}{1+\beta^2}\hat{a}_G^+f_E+\frac{\beta\gamma}{\sqrt{2}}\frac{\partial f_E}{\partial \hat{a}_E}-\frac{\beta\gamma}{\sqrt{2}}\frac{(1-\beta^2)}{(1+\beta^2)}\frac{\partial f_E}{\partial \hat{a}_E^+}\\
&\quad+\frac{1-\beta^2}{1+\beta^2}\hat{a}_G^+\hat{a}_G^+f_E+\frac{1-\beta^2}{2\beta}\hat{a}_E^+\frac{\partial f_E}{\partial \hat{a}_E}-\frac{(1-\beta^2)^2}{2\beta(1+\beta^2)}\hat{a}_G^+\frac{\partial f_E}{\partial \hat{a}_E^+}\\
&\quad+\frac{1-\beta^2}{2\beta}\frac{\partial f_E}{\partial \hat{a}_E}\hat{a}_G^+-\frac{(1-\beta^2)^2}{2\beta(1+\beta^2)}\frac{\partial f_E}{\partial \hat{a}_E^+}\hat{a}_G^++\frac{2\beta}{1+\beta^2}\hat{a}_G^+f_E\hat{a}_E+\frac{\partial f_E}{\partial \hat{a}_E}\hat{a}_E-\frac{1-\beta^2}{1+\beta^2}\frac{\partial f_E}{\partial \hat{a}_E^+}\hat{a}_E,
\end{aligned} \tag{2.20}$$

with the commutation relations

$$\frac{\partial f_E}{\partial \hat{a}_E}\hat{a}_G^+=\hat{a}_G^+\frac{\partial f_E}{\partial \hat{a}_E}+\frac{1+\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E}-\frac{1-\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E} \tag{2.21}$$

and

$$\frac{\partial f_E}{\partial \hat{a}_E^+}\hat{a}_G^+=\hat{a}_G^+\frac{\partial f_E}{\partial \hat{a}_E^+}+\frac{1+\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E^+}-\frac{1-\beta^2}{2\beta}\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E^+}, \tag{2.22}$$

is used in Eq. (2.14) along the properties of ladder operators in order to get

$$\begin{aligned}
&\frac{1}{\alpha_G}\,{}_G\langle m|[H_G,f_E]|n\rangle_E\\
&\quad=m\,{}_G\langle m|f_E|n\rangle_E-\frac{\beta^2\gamma}{1+\beta^2}\sqrt{2m}\,{}_G\langle m-1|f_E|n\rangle_E-\frac{\beta\gamma}{\sqrt{2}}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E}|n\rangle_E+\frac{\beta\gamma(1-\beta^2)}{\sqrt{2}(1+\beta^2)}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E\\
&\qquad-\frac{1-\beta^2}{1+\beta^2}\sqrt{m(m-1)}\,{}_G\langle m-2|f_E|n\rangle_E-\frac{1-\beta^2}{\beta}\sqrt{m}\,{}_G\langle m-1|\frac{\partial f_E}{\partial \hat{a}_E}|n\rangle_E+\frac{(1-\beta^2)^2}{\beta(1+\beta^2)}\sqrt{m}\,{}_G\langle m-1|\frac{\partial f_E}{\partial \hat{a}_E^+}|n\rangle_E\\
&\qquad-\frac{1-\beta^4}{4\beta^2}\,{}_G\langle m|\frac{\partial^2 f_E}{\partial \hat{a}_E\,\partial \hat{a}_E}|n\rangle_E-\frac{(1-\beta^2)^3}{4\beta^2(1+\beta^2)}\,{}_G\langle m|\frac{\partial^2 f_E}{\partial \hat{a}_E^+\,\partial \hat{a}_E^+}|n\rangle_E-\frac{2\beta}{1+\beta^2}\sqrt{mn}\,{}_G\langle m-1|f_E|n-1\rangle_E\\
&\qquad-\sqrt{n}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E}|n-1\rangle_E+\frac{1-\beta^2}{1+\beta^2}\sqrt{n}\,{}_G\langle m|\frac{\partial f_E}{\partial \hat{a}_E^+}|n-1\rangle_E
\end{aligned}$$

$$+\left(\frac{1-\beta^2}{2\beta}\right)^2 {}_G\langle m|\frac{\partial^2 f_E}{\partial\hat{a}_E^+\,\partial\hat{a}_E}|n\rangle_E+\left(\frac{1-\beta^2}{2\beta}\right)^2 {}_G\langle m|\frac{\partial^2 f_E}{\partial\hat{a}_E\,\partial\hat{a}_E^+}|n\rangle_E\,. \tag{2.23}$$

The general left-hand recurrence relation is likewise obtained from Eq. (2.19) and Eq. (2.23). It is given by

$$\begin{aligned}{}_G\langle m|f(\hat{a}_E^+,\hat{a}_E)|n\rangle_E &=\frac{\gamma\beta^2}{1+\beta^2}\sqrt{\frac{2}{m}}\,{}_G\langle m-1|f_E|n\rangle_E+\frac{2\beta}{1+\beta^2}\sqrt{\frac{n}{m}}\,{}_G\langle m-1|f_E|n-1\rangle_E\\ &\quad+\frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{m-1}{m}}\,{}_G\langle m-2|f_E|n\rangle_E+\frac{2\beta}{1+\beta^2}\sqrt{\frac{1}{m}}\,{}_G\langle m-1|\frac{\partial f_E}{\partial\hat{a}_E^+}|n\rangle_E\,,\quad n\neq 0,\quad m>n\,.\end{aligned}\tag{2.24}$$

Both Eqs. (2.12) and (2.24) are exact general recurrence relations for the calculation of two-center HO integrals of $f(\hat{a}_E,\hat{a}_E^+)$ arbitrary operators.

Before presenting some examples that show the usefulness of the proposed general recurrence relations, we want to point out that these equations contain, as a particular case, the generalized recurrence relation for the calculation of one-center HO integrals of $f(\hat{a},\hat{a}^+)$ arbitrary operators.[3] In fact, in the particular case of $\omega_E=\omega_G$ and $l=0$ we get $\beta=1$ and $\gamma=0$. Consequently, Eqs. (2.12) and (2.24) reduce to[7]

$$\begin{aligned}(m-n)\langle m|f(\hat{a}^+,\hat{a})|n\rangle &=\sqrt{m}\,\langle m-1|\frac{\partial f}{\partial\hat{a}^+}|n\rangle\\ &\quad-\sqrt{n}\,\langle m|\frac{\partial f}{\partial\hat{a}}|n-1\rangle\,.\end{aligned}\tag{2.25}$$

This last equation is identical to that obtained from the hypervirial theorem and the second quantization formalism.[3]

III. APPLICATIONS

In this section we present the most useful particular cases used in literature and new recurrence relations for the calculation of integrals of power, exponential, and Gaussian operators are shown.

A. Overlap integrals

These kind of integrals are particularly useful through the so-called Franck–Condon factors and are obtained straightforwardly when $f(\hat{a}_E,\hat{a}_E^+)=\text{const}$. In such a case,

$$\frac{\partial f_E}{\partial\hat{a}_E^+}=\frac{\partial f_E}{\partial\hat{a}_E}=0\,. \tag{3.1}$$

When applied to Eqs. (2.12) and (2.24), the following expressions are obtained:

$$\begin{aligned}{}_G\langle m|n\rangle_E &=-\frac{\beta\gamma}{1+\beta^2}\sqrt{\frac{2}{n}}\,{}_G\langle m|n-1\rangle_E\\ &\quad+\frac{2\beta}{1+\beta^2}\sqrt{\frac{m}{n}}\,{}_G\langle m-1|n-1\rangle_E\\ &\quad-\frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{n-1}{n}}\,{}_G\langle m|n-2\rangle_E\end{aligned}\tag{3.2}$$

and

$$\begin{aligned}{}_G\langle m|n\rangle_E &=\frac{\beta^2\gamma}{1+\beta^2}\sqrt{\frac{2}{m}}\,{}_G\langle m-1|n\rangle_E\\ &\quad+\frac{2\beta}{1+\beta^2}\sqrt{\frac{n}{m}}\,{}_G\langle m-1|n-1\rangle_E\\ &\quad+\frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{m-1}{m}}\,{}_G\langle m-2|n\rangle_E\,.\end{aligned}\tag{3.3}$$

These are the well-known recurrence relations given by Ansbacher[8] widely used in vibrational spectroscopy for the calculation of Franck–Condon factors. Their usefulness stems from the fact that any overlap can be calculated from the generator transition element

$${}_G\langle 0|0\rangle_E=\left(\frac{2\beta}{1+\beta^2}\right)^{1/2}\exp\left(-\frac{\beta^2\gamma^2}{2(1+\beta^2)}\right), \tag{3.4}$$

which has been also evaluated algebraically elsewhere.[4]

B. Integrals of power operators

Let

$$f(\hat{a}_E,\hat{a}_E^+)=\hat{X}_E^K=\hat{X}_E^{k+1}, \tag{3.5}$$

where the position operator $\hat{X}_E$ is defined in terms of ladder operators by

$$\hat{X}_E=(\alpha_E/\sqrt{2})(\hat{a}_E+\hat{a}_E^+)\,. \tag{3.6}$$

Its partial derivatives are given by

$$\frac{\partial f_E}{\partial\hat{a}_E}=\frac{\partial f_E}{\partial\hat{a}_E^+}=(k+1)\,\frac{\alpha_E}{\sqrt{2}}\,\hat{X}_E^k\,. \tag{3.7}$$

At this point it should be noted that, depending on the choice of $\hat{X}_E^{(k+1)}$ in terms of ladder operators, many recurrence relations can be obtained from the general right and left-hand recurrence relations. However, the most useful recursion formulas come from the identities

$$\text{(a)}\quad \hat{X}_E^{k+1}=\hat{X}_E^k\,\hat{X}_E\,, \tag{3.8a}$$

and

$$\text{(b)}\quad \hat{X}_E^{k+1}=\hat{X}_E\,\hat{X}_E^k\,. \tag{3.8b}$$

In case (a), the use of the definition (3.6) in the general right-hand recurrence specified by Eq. (2.12) leads to

$$n(n-1)(1+\beta^2)_G\langle m|\hat{x}_E^k|n\rangle_E$$
$$= [\beta^2(2k+1)-2n+3]\sqrt{n(n-1)}\ _G\langle m|\hat{x}_E^k|n-2\rangle_E$$
$$+2\beta(n-1)\sqrt{mn}\ _G\langle m-1|\hat{x}_E^k|n-1\rangle_E + 2\beta\sqrt{mn(n-1)(n-2)}\ _G\langle m-1|\hat{x}_E^k|n-3\rangle_E$$
$$-(1-\beta^2)\sqrt{n(n-1)(n-2)(n-3)}\ _G\langle m|\hat{x}_E^k|n-4\rangle_E - \beta\gamma(n-1)\sqrt{2n}\ _G\langle m|\hat{x}_E^k|n-1\rangle_E$$
$$-\beta\gamma\sqrt{2n(n-1)(n-2)}\ _G\langle m|\hat{x}_E^k|n-3\rangle_E\ . \tag{3.9}$$

Similarly, the corresponding second recurrence relation for the calculation of integrals of X_E^k, is obtained from the general left-hand recurrence relation, i.e.,

$$mn(1+\beta^2)\ _G\langle m|\hat{X}_E^k|n\rangle_E$$
$$= -(1+\beta^2)m\sqrt{n(n-1)}\ _G\langle m|\hat{X}_E^k|n-2\rangle_E + \beta^2\gamma n\sqrt{2m}\ _G\langle m-1|\hat{X}_E^k|n\rangle_E + \beta^2\gamma\sqrt{2nm(n-1)}$$
$$\times _G\langle m-1|\hat{X}_E^k|n-2\rangle_E + 2\beta(n+k)\sqrt{mn}\ _G\langle m-1|\hat{X}_E^k|n-1\rangle_E + 2\beta\sqrt{mn(n-1)(n-2)}$$
$$\times _G\langle m-1|\hat{X}_E^k|n-3\rangle_E + (1-\beta^2)n\sqrt{m(m-1)}\ _G\langle m-2|\hat{X}_E^k|n\rangle_E$$
$$+(1-\beta^2)\sqrt{m(m-1)n(n-1)}\ _G\langle m-2|\hat{X}_E^k|n-2\rangle_E\ ,\quad m\neq 0\ . \tag{3.10}$$

We can observe that in the case of $\omega_E = \omega_G$ and $l=0$, the above two equations reduce to

$$(m-n+1)\sqrt{n}\langle m|X^k|n\rangle = (k+1)\sqrt{m}\langle m-1|X^k|n-1\rangle - (m-n+k+2)\sqrt{n-1}\ \langle m|X^k|n-2\rangle\ ,$$
$$n+m\geqslant 2,\quad n\neq m+1\ , \tag{3.11}$$

that is, one of the recursion relations obtained from the hypervirial theorem and ladder operators for the calculation of $\hat{X}^k$ one-center HO matrix elements [Eq. (4.4) in Ref. 3].

Case (b) can be treated similarly through the identity

$$\hat{a}_E^+ + \hat{a}_E = \beta\hat{a}_G + \beta\,\hat{a}_G^+ - \sqrt{2}\beta\gamma\ , \tag{3.12}$$

in order to get the recurrence relations

$$mn(1+\beta^2)\ _G\langle m|\hat{X}_E^k|n\rangle_E$$
$$= -n(1+\beta^2)\sqrt{m(m-1)}\ _G\langle m-2|\hat{X}_E^k|n\rangle_E$$
$$+\gamma(1+\beta^2)n\sqrt{2m}\ _G\langle m-1|\hat{X}_E^k|n\rangle_E + 2\beta\sqrt{nm(m-1)(m-2)}\ _G\langle m-3|\hat{X}_E^k|n-1\rangle_E$$
$$+2\beta(\gamma^2+k+m)\sqrt{mn}\ _G\langle m-1|\hat{X}_E^k|n-1\rangle_E - 3\beta\gamma\sqrt{2nm(m-1)}\ _G\langle m-2|\hat{X}_E^k|m-1\rangle_E$$
$$-(1-\beta^2)\sqrt{m(m-1)n(n-1)}\ _G\langle m-2|\hat{X}_E^k|n-2\rangle_E - m(1-\beta^2)\sqrt{n(n-1)}\ _G\langle m|\hat{X}_E^k|n-2\rangle_E$$
$$+\gamma(1-\beta^2)\sqrt{2mn(n-1)}\ _G\langle m-1|\hat{X}_E^k|n-2\rangle_E - \beta\gamma m\sqrt{2n}\ _G\langle m|\hat{X}_E^k|n-1\rangle_E \tag{3.13}$$

and

$$m(m-1)(1+\beta^2)\ _G\langle m|\hat{X}_E^k|n\rangle_E$$
$$= [2k+1-\beta^2(2m+2\gamma^2-3)]\sqrt{m(m-1)}\ _G\langle m-2|\hat{X}_E^k|n\rangle_E$$
$$+\gamma(1+2\beta^2)(m-1)\sqrt{2m}\ _G\langle m-1|\hat{X}_E^k|n\rangle_E + \gamma(2\beta^2-1)\sqrt{2m(m-1)(m-2)}\ _G\langle m-3|\hat{X}_E^k|n\rangle_E$$
$$+2\beta\sqrt{nm(m-1)(m-2)}\ _G\langle m-3|\hat{X}_E^k|n-1\rangle_E + 2\beta(m-1)\sqrt{mn}\ _G\langle m-1|\hat{X}_E^k|n-1\rangle_E$$
$$-2\beta\gamma\sqrt{2nm(m-1)}\ _G\langle m-2|\hat{X}_E^k|n-1\rangle_E + (1-\beta^2)\sqrt{m(m-1)(m-2)(m-3)}\ _G\langle m-4|\hat{X}_E^k|n\rangle_E\ . \tag{3.14}$$

As before, the above equations reduce to the corresponding one-center recursion relation when $\beta=1$, $\gamma=0$, i.e.,

$$(n-m+1)\sqrt{m}\langle m|\hat{X}^k|n\rangle$$
$$= -(n-m+k+2)(\sqrt{m-1}\langle m-2|\hat{X}^k|n\rangle$$
$$+(k+1)\sqrt{n}\langle m-1|\hat{X}^k|n-1\rangle\ ,$$
$$n+m\geqslant 2,\quad n\neq m-1\ , \tag{3.15}$$

as reported in Ref. 3.

C. Integrals of exponential operators

Let

$$f(\hat{a}_E,\hat{a}_E^+) = \exp(-\rho\hat{X}_E)\ . \tag{3.16}$$

The use of

$$\frac{\partial f_E}{\partial\hat{a}_E} = \frac{\partial f_E}{\partial\hat{a}_E^+} = -\frac{\rho\alpha_E}{\sqrt{2}}\exp(-\rho\hat{X}_E)\ , \tag{3.17}$$

in the corresponding equations (2.12) and (2.24) leads straightforwardly to the useful recursion relations

$$
\begin{aligned}
&{}_G\langle m|\exp(-\rho\hat{X}_E)|n\rangle_E \\
&\quad=\frac{2\beta}{1+\beta^2}\sqrt{\frac{m}{n}}\,{}_G\langle m-1|\exp(-\rho\hat{X}_E)|n-1\rangle_E \\
&\qquad-\frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{n-1}{n}}\,{}_G\langle m|\exp(-\rho\hat{X}_E)|n-2\rangle_E \\
&\qquad-\frac{\beta(\beta\rho\alpha_E+\gamma)}{1+\beta^2}\sqrt{\frac{2}{n}} \\
&\qquad\times{}_G\langle m|\exp(-\rho\hat{X}_E)|n-1\rangle_E\,, \\
&\qquad n\neq 0,\quad n>m\,,
\end{aligned}
\tag{3.18}
$$

and

$$
\begin{aligned}
&{}_G\langle m|\exp(-\rho\hat{X}_E)|n\rangle_E \\
&\quad=\frac{\beta(\beta\gamma-\rho\alpha_E)}{1+\beta^2}\sqrt{\frac{2}{m}}\,{}_G\langle m-1|\exp(-\rho\hat{X}_E)|n\rangle_E \\
&\qquad+\frac{2\beta}{1+\beta^2}\sqrt{\frac{n}{m}}\,{}_G\langle m-1|\exp(-\rho\hat{X}_E)|n-1\rangle_E \\
&\qquad+\frac{1-\beta^2}{1+\beta^2}\sqrt{\frac{m-1}{m}}\,{}_G\langle m-2|\exp(-\rho\hat{X}_E)|n\rangle_E\,, \\
&\qquad m\neq 0,\quad m>n\,.
\end{aligned}
\tag{3.19}
$$

The generator matrix element is given by

$$
{}_G\langle 0|\exp(-\rho\hat{X}_E)|0\rangle_E=\exp\left[\frac{(\beta\rho\gamma+\beta^2\rho^2/2)}{(1+\beta^2)}\right]{}_G\langle 0|0\rangle_E\,.
\tag{3.20}
$$

The corresponding recurrence relation for one-center HO integrals is[3]

$$
\begin{aligned}
&(m-n)\langle m|\exp(-\rho\hat{X})|n\rangle \\
&\quad=\rho\sqrt{(n/2)}\,\langle m|\exp(-\rho\hat{X})|n-1\rangle \\
&\qquad-\rho\sqrt{(m/2)}\,\langle m-1|\exp(-\rho\hat{X})|n\rangle\,,
\end{aligned}
\tag{3.21}
$$

with generator

$$
\langle 0|\exp(-\rho\hat{X})|0\rangle=\exp(\tfrac{1}{4}\rho^2)\,.
\tag{3.22}
$$

D. Integrals of Gaussian operators

We consider finally

$$
f(\hat{a}_E,\hat{a}_E^{+})=\exp(-\rho\hat{X}_E^2)\,.
\tag{3.23}
$$

In this case, the normal ordered relation

$$
\frac{\partial f_E}{\partial\hat{a}_E}=\frac{\partial f_E}{\partial\hat{a}_E^{+}}=-\frac{2\beta\rho\alpha_E^2}{\eta}\hat{a}_G^{+}\exp(-\rho\hat{X}_E^2)+\frac{\sqrt{2}\beta\gamma\rho\alpha_E^2}{\eta}\exp(-\rho\hat{X}_E^2)-\frac{2\beta^2\rho\alpha_E^2}{\eta}\exp(-\rho\hat{X}_E^2)\hat{a}_E\,,
\tag{3.24}
$$

where $\eta=1+\beta^2+2\beta^2\alpha_E^2\rho$, is used, respectively in, Eq. (2.12) and Eq. (2.24) in order to get the appropriate recurrence relations:

$$
\begin{aligned}
{}_G\langle m|\exp(-\rho\hat{X}_E^2)|n\rangle_E&=\frac{2\beta}{\eta}\sqrt{\frac{m}{n}}\,{}_G\langle m-1|\exp(-\rho\hat{X}_E^2)|n-1\rangle_E \\
&\quad-\frac{\eta-2\beta^2}{\eta}\sqrt{\frac{n-1}{n}}\,{}_G\langle m|\exp(-\rho\hat{X}_E^2)|n-2\rangle_E \\
&\quad-\frac{\beta\gamma}{\eta}\sqrt{\frac{2}{n}}\,{}_G\langle m|\exp(-\rho\hat{X}_E^2)|n-1\rangle_E\,,\quad n\neq 0,\quad n>m\,,
\end{aligned}
\tag{3.25}
$$

and

$$
\begin{aligned}
{}_G\langle m|\exp(-\rho\hat{X}_E^2)|n\rangle_E&=\frac{\beta^2\gamma(1+2\rho\alpha_E^2)}{\eta}\sqrt{\frac{2}{m}}\,{}_G\langle m-1|\exp(-\rho\hat{X}_E^2)|n\rangle_E \\
&\quad+\frac{2\beta}{\eta}\sqrt{\frac{n}{m}}\,{}_G\langle m-1|\exp(-\rho\hat{X}_E^2)|n-1\rangle_E \\
&\quad-\frac{\eta-2}{\eta}\sqrt{\frac{m-1}{m}}\,{}_G\langle m-2|\exp(-\rho\hat{X}_E^2)|n\rangle_E\,,\quad m\neq 0,\quad m>n\,,
\end{aligned}
\tag{3.26}
$$

with generator given by

$$
\begin{aligned}
&{}_G\langle 0|\exp(-\rho\hat{X}_E^2)|0\rangle_E \\
&\quad=(1+\xi)^{-1/2}\exp\left(-\frac{1}{2}\frac{\xi}{(1+\xi)}\right){}_G\langle 0|0\rangle_E\,,
\end{aligned}
\tag{3.27}
$$

where

$$\xi=2\rho\beta^2/(1+\beta^2)\,.$$

As before, the particular case of the corresponding recurrence relation for the evaluation of one-center integrals arises from Eqs. (3.25) and (3.26) for $\beta=1$ and $\gamma=0$. It is written

$$
\begin{aligned}
&(m-n)\langle m|\exp(-\rho\hat{X}^2)|n\rangle \\
&\quad=[\rho/(1+\rho)]\sqrt{n(n-1)}\,\langle m|\exp(-\rho\hat{X}^2)|n-2\rangle
\end{aligned}
$$

$$- [\rho/(1+\rho)]\sqrt{m(m-1)}$$
$$\times \langle m-2|\exp(-\rho \hat{X}^2)|n\rangle . \quad (3.28)$$

This last equation can be transformed, by the use of the identity[3]

$$[1/(1+\rho)]\exp(-\rho \hat{X}^2)\hat{a}\hat{a}$$
$$= \exp(-\rho \hat{X}^2)\hat{a}^+\hat{a} + \exp(-\rho \hat{X}^2)\hat{a}\hat{a}$$
$$- [1/(1+\rho)]\hat{a}^+ \exp(-\rho \hat{X}^2)\hat{a} , \quad (3.29)$$

into the useful recurrence relation

$$(m - n(1+\rho))\langle m|\exp(-\rho \hat{X}^2)|n\rangle$$
$$= \rho\sqrt{n(n-1)}\langle m|\exp(-\rho \hat{X}^2)|n-2\rangle$$
$$- [\rho/(1+\rho)]\sqrt{mn}\langle m-1|\exp(-\rho \hat{X}^2)|n-1\rangle$$
$$- [\rho/(1+\rho)]\sqrt{m(m-1)}$$
$$\times \langle m-2|\exp(-\rho \hat{X}^2)|n\rangle , \quad (3.30)$$

from where it is directly recognized that

$$\langle m|\exp(-\rho \hat{X}^2)|n\rangle = 0 \Leftrightarrow n+m = \text{odd} . \quad (3.31)$$

It should be noted that the use of the recurrence relation specified by Eq. (3.30) needs only the knowledge of the element

$$\langle 0|\exp(-\rho \hat{X}^2)|0\rangle = (1+\rho)^{-1/2} , \quad (3.32)$$

avoiding the Chan–Stelman diamond rule.[9]

IV. DISCUSSION

We have obtained general recurrence relations for the calculation of two-center HO integrals of arbitrary operators, by means of a method that proposes the combined use of an hypervirial-like-theorem commutator algebra and second quantization formalism. Comparatively, the proposed formulas are by far easier to handle than the corresponding ones given previously by Wong.[1] For example, the particular case of overlap integrals is straightforwardly deduced from our recursion formulas and the respective Ansbacher recurrence relations for the calculation of Franck–Condon factors are rederived. The proposed general recurrence relations have been also applied to the particular case of power, exponential, and Gaussian functions and new recurrence relations are given. In addition, the recursion formulas reduce, in all cases, to the appropriate recurrence relations for the calculation of one-center HO integrals. The algebraic procedure shown here can be extended to other potentials.

[1]C. W. Wong, Nucl. Phys. A **147**, 563 (1970).
[2]J. Morales, A. Palma, and M. Berrondo, Int. J. Quantum Chem. **S18**, 57 (1984).
[3]J. Morales, A. Palma, and L. Sandoval, Int. J. Quantum Chem. **29**, 211 (1986).
[4]A. Palma and J. Morales, Int. J. Quantum Chem. **S17**, 393 (1983).
[5]W. H. Louissell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1963).
[6]J. O. Hirschfelder, J. Chem. Phys. **33**, 1462 (1960).

[7]In all cases where we consider the particular recurrence relations for the calculation of one-center integrals, it should be noted that we use natural units $\hbar = \mu = \omega = 1$.

[8]F. Ansbacher, Z. Naturforsch. Teil A **14**, 889 (1959).
[9]S. I. Chan and D. Stelman, J. Chem. Phys. **39**, 545 (1963).

3+1 Regge calculus with conserved momentum and Hamiltonian constraints

John L. Friedman and Ian Jack
Physics Department, University of Wisconsin, Milwaukee, Wisconsin 53201

(Received 20 May 1986; accepted for publication 11 July 1986)

The Einstein action is evaluated for space-times whose three-metrics on a family of spacelike hypersurfaces are piecewise flat. The 3 + 1 action of Lund and Regge [F. Lund and T. Regge (private communication)], recently generalized by Piran and Williams [T. Piran and R. M. Williams, Phys. Rev. D **33**, 1622 (1986)], is recovered in this way. A natural interpretation of the momentum constraint is obtained for simplicial initial data sets; and, by incorporating a nonzero shift vector and a nonconstant lapse, one finds a formalism in which the constraints are preserved by the time evolution. (In contrast to the continuum case, the constraints are *not* conserved if the lapse and shift are chosen *a priori.*) A consistent Hamiltonian formalism is readily obtained by the standard (Bergmann–Dirac) procedure or, alternatively, by algebraically solving the constraint equations for the lapse and shift on each three-simplex. Explicit solutions to the classical equations are found for spaces built from congruent simplices. In this special case, the action is that of a free relativistic particle moving in a curved space-time with indefinite metric and a conformal timelike Killing vector. For general space-times, if one *a priori* sets the shift vector to zero, the action has the form of a sum of such free-particle actions, but one for which the different particles interact by having coordinates in common.

I. INTRODUCTION

Despite the formal elegance of the four-dimensional Regge calculus,[1] it has not yet been used in numerical calculations of explicit classical space-times. This is due partly to the fact that its initial value formulation[2] is unfamiliar, as is the required machinery of four-dimensional polytopes. Several calculations have been done, however, using 3 + 1 versions, in which space-time is a Cartesian product of a simplicial space and a continuous or discrete time axis.[3–12] An action and associated Hamiltonian for a 3 + 1 theory (with time continuous) was given for homogeneous, isotropic spaces by Lund and Regge[13] in unpublished notes; and a generalization to arbitrary space-times has recently been obtained by Piran and Williams.[14] Their formalism does not, however, include a discrete version of the "momentum constraint" of the continuum theory, the analog in gravity of the Gauss constraint of electromagnetism. For homogeneous space-times, this constraint is automatically satisfied, but in the general case it is needed to complete the theory. Moreover, as Piran and Williams observe, the Hamiltonian constraint is not conserved by the time evolution of the system, at least not when the lapse is chosen *a priori*.

In the present paper, we obtain a theory that incorporates a natural version of the momentum constraint, by evaluating the continuum action on a space-time foliated by *piecewise-flat* three-manifolds. That is, the space-time metric is continuous everywhere and smooth only in world tubes that represent the history of three-simplices. For one particular slicing of space-time, the induced metric on each three-simplex is flat, determined in the usual way by the edge lengths, which thus become the dynamical variables of the theory. An arbitrariness remains in how one specifies the lapse and the shift vector, or, equivalently, the time–time and time–space components of the metric. With zero shift we recover the Lund–Regge–Piran–Williams action in the case of constant lapse. (We were initially unaware of their work and independently derived an action that agrees with theirs for zero shift.) When both shift and lapse are present and piecewise constant (on cells), one obtains discrete forms of the Hamiltonian and momentum constraints.

In contrast to the continuum theory, the lapse and shift are not "pure gauge." In fact, no gauge freedom remains in the discrete theory: a generic four-geometry of the form described above has a unique description in terms of edge lengths, lapse, and shift.

The reason is that in specifying the *class* of four-geometries one fixes a foliation of the four-manifold—a set of $t=$ const surfaces—together with a set of world tubes, the histories of the flat three-simplices. In general, a metric in the allowed class will be mapped out of that class by the infinitesimal diffeomorphism generated by any allowed lapse and shift: One fails either to map world tubes to world tubes or to map the $t=$ const surfaces to each other.

Instead of specifying the lapse and shift *a priori*, one must solve the constraint equations for them. That is, given an initial data set, the constraint equations, together with the equations of motion for the edge lengths, uniquely determine the time-evolved lapse and shift. The constraints are then automatically preserved by the time evolution of the system.

Regarded as a model for continuum gravity, the 3 + 1 theory is a substantial improvement over the usual discretization of the field equations by finite difference equations, where the constraints are only conserved to lowest order in the lattice spacing. The 3 + 1 Regge equations form a consistent finite-dimensional Lagrangian system, with a corresponding well-defined Hamiltonian formalism.

The Hamiltonian corresponding to our discrete Lagrangian is clumsy, however, because it involves the inverse of a band-diagonal matrix (the supermetric), which has nonvanishing contributions from each edge in the complex. One can avoid this loss of locality by directly evaluating the continuum Hamiltonian for a simplicial metric and a piecewise-constant conjugate momentum. Again the Hamiltonian and momentum constraints are preserved and second class, although their structure differs from that in the inequivalent Lagrangian theory. When the shift is set to zero, our Hamiltonian corrects that given by Piran and Williams[12] (see Sec. III).

From the standpoint of quantum gravity, the $3+1$ Regge theory is complementary to the four-dimensional functional integral approach considered recently by a number of authors in the context of Regge calculus.[15–30] It provides a set of minisuperspace theories whose structure is quite close to that of the full theory. In particular, one has again a vanishing Hamiltonian, and the associated supermetric has a conformal Killing vector that corresponds to uniform scaling of the edge lengths (to changing the three-metric by a conformal factor).

The plan of the paper is as follows. In Sec. II we introduce a class of space-times having piecewise-flat spacelike hypersurfaces. We briefly review the description of simplicial geometry in terms of affine coordinates. A simplicial covariant derivative is introduced and used to calculate the extrinsic curvature of the spacelike hypersurfaces. Next, in Sec. III, we compute the Einstein action for space-times in the class introduced in Sec. II. We obtain in this way a finite-dimensional dynamical system whose configuration space consists of the squared edge lengths of each cell (three-simplex), together with piecewise-constant lapse function and shift vector. By specifying the shift on cells, one obtains a natural simplicial form of the momentum constraint. The equations of motion are obtained and a discussion of the constraints is given: In the generic case they form a system of linear equations that can be solved for the lapse and shift on each cell. A Hamiltonian formalism is then presented, and the iterative Bergmann–Dirac procedure is shown to terminate, implying the existence of a well-defined Hamiltonian theory, although in contrast to the continuum case, the constraints are not first class.

Section IV applies the formalism to the simple case where all three-simplices are congruent. A theorem is proved to the effect that for any tiling of S^3 or R^3 by congruent cells, the dynamical equations can be cast into a form identical to that for the continuum theory. Two spherical spaces, the quaternion space and the family of lens spaces $L(p,1)$, are presented to illustrate the $3+1$ theory and to emphasize the ease with which spatial topologies can be handled in the $3+1$ theory. The family of lens spaces provides an example of a partial approach to the continuum: as $p\to\infty$, the sequence of spaces becomes smooth in two of the three spatial directions. Finally, in Sec. V, we briefly consider the corresponding quantum theory for the simple minisuperspace where all cells are congruent. The constraints can then be solved explicitly, and the resulting Lagrangian describes a free relativistic particle moving in a six-dimensional curved space-time whose (super) metric has signature $-+++++$. In the general case, where the cells are not congruent, the Lagrangian has the form of a sum of free-particle Lagrangians, but one in which the particles have coordinates in common.

II. SPACE-TIMES WITH PIECEWISE-FLAT SPACELIKE SLICES

A. Simplicial three-geometry

By a three-geometry we mean a three-manifold M together with a positive-definite metric g_{ab}. A three-geometry will be called simplicial if the metric is piecewise flat: that is, if there is a decomposition of M into a set of three-simplices (tetrahedra) that intersect only at their boundaries, and for which g_{ab} is flat in the interior of each simplex. A smooth three-geometry can be approximated by a sequence of simplicial geometries on a fixed manifold M.

We will use the following index conventions:

space-time indices: $\alpha,\beta,\gamma,\delta$,
spatial indices: a,b,c,d,
cells of complex: C,D,
edges of complex: ι, κ, λ,
vertices of complex: i,j,k,l.

Space-time will have signature $-+++$.

Our calculations are simplified by using an affine chart on each flat three-simplex.[31] Affine coordinates respect the tetrahedral symmetry at the cost of introducing a superfluous coordinate. Taking the origin to be the barycenter of the simplex, one chooses as basis $\{\mathbf{e}_i\}$ the *four* vectors that connect the origin to each of the four vertices. The basis vectors are then related by

$$\sum \mathbf{e}_i = 0\,. \tag{2.1}$$

The affine components of a vector are defined by

$$\mathbf{v} = v^i\mathbf{e}_i, \quad \sum v^i = 0\,, \tag{2.2}$$

and, in particular, any point x of the simplex has affine coordinates x^i, where $x^i\mathbf{e}_i$ is the vector from the origin to x. Defining the covariant basis $\{\boldsymbol{\omega}^i\}$ dual to $\{\mathbf{e}_i\}$ by

$$\boldsymbol{\omega}^i\otimes\mathbf{e}_i = 1, \quad \sum\boldsymbol{\omega}^i = 0\,, \tag{2.3}$$

one finds that the affine components of the unit tensor (the Kronecker delta) have the form

$$\tilde{\delta}^i_j = \begin{cases} \frac{3}{4}, & i=j \\ -\frac{1}{4}, & i\neq j\,. \end{cases} \tag{2.4}$$

Then $\tilde{\delta}^i_j$ projects onto the affine basis in the sense that if $\mathbf{v} = \hat{v}^i\mathbf{e}_i$, then $\mathbf{v}$ has affine components $v^i = \tilde{\delta}^i_j\hat{v}^j$.

The components of the flat metric can be regarded as the projection to the affine basis of s_{ij}, the squared length of the edge l_{ij} joining vertex i to vertex j:

$$g_{ij} = -\tfrac{1}{2}\tilde{\delta}^k_i\,\tilde{\delta}^l_j s_{kl}\,. \tag{2.5}$$

The affine components of the contravariant metric can be written as follows in terms of the volume V of the tetrahedron and the areas of its faces: Diagonal components have the form

$$g^{ii} = A_i^2/9V^2 \quad (\text{no sum})\,, \tag{2.6}$$

where A_i denotes the area of the face opposite the ith vertex; and off-diagonal components have the form

$$g^{ij} = -\frac{1}{V^2}\frac{\partial V^2}{\partial s_{ij}}. \qquad (2.7)$$

Equation (2.7) can be regarded as a form of the familiar relation

$$g^{ab} = \frac{1}{g}\frac{\partial g}{\partial g_{ab}}, \qquad (2.8)$$

expressing the contravariant metric as a derivative of the density g, whose value in any basis is the determinant of the covariant metric components. If the squared volume is regarded as a density proportional to g, then Eq. (2.7) follows from (2.8) and the equation

$$\frac{\partial g_{ij}}{\partial s_{kl}} = -\tilde{\delta}^k_{(i}\,\tilde{\delta}^l_{j)}\,. \qquad (2.9)$$

[To be more precise, we are treating $\{s_{ij}\}$ as a set of six independent variables, and Eq. (2.9) is used here in the form $\partial g_{ab}/\partial s_{kl} = -\omega^k_{(a}\,\omega^l_{b)}\cdot$.]

A similar identity relates the second derivatives of the density g to the supermetric G^{abcd}

$$G^{abcd} = -\frac{1}{g}\frac{\partial^2 g}{\partial g_{ab}\,\partial g_{cd}} = g^{a(c}g^{d)b} - g^{ab}g^{cd}. \qquad (2.10)$$

The affine components of the supermetric are then given in terms of the squared volume by

$$G^{ijkl} = -\frac{1}{V^2}\frac{\partial^2 V^2}{\partial s_{ij}\,\partial s_{kl}}. \qquad (2.11)$$

Equation (2.5) relating the affine components of the metric to its projections along the edges can be generalized to an arbitrary symmetric covariant tensor σ_{ab}: the component of σ associated with the ιth edge (l^a), joining vertices i and j,

$$\sigma_\iota \equiv \hat{\sigma}_{ij} = \sigma_{ab}l^a l^b \qquad (2.12)$$

is related to the affine components by

$$\sigma_{ij} = -\tfrac{1}{2}\tilde{\delta}^k_i\,\tilde{\delta}^l_j\hat{\sigma}_{kl}\,. \qquad (2.13)$$

Then for an arbitrary contravariant tensor τ^{ab} the inner product of τ and σ involves only the off-diagonal affine components of τ:

$$\tau^{ab}\sigma_{ab} = -\tfrac{1}{2}\tau^{ij}\hat{\sigma}_{ij}\,, \qquad (2.14)$$

where $\hat{\sigma}_{ij}$ is taken to vanish for $i=j$. In particular, we will use the identity

$$G^{abcd}\sigma_{ab}\sigma_{cd} = -\frac{1}{V^2}\frac{\partial^2 V^2}{\partial s_\iota\,\partial s_\kappa}\sigma_\iota\sigma_\kappa\,, \qquad (2.15)$$

which follows from Eqs. (2.11), (2.12), and (2.14).

If $\mathbf{n}$ is the unit normal to the face (with area A) opposite the ith vertex, Eq. (2.6) implies

$$\mathbf{n} = (3V/A)\boldsymbol{\omega}^i \qquad (2.16)$$

or

$$n_j = (3V/A)\tilde{\delta}^i_j = (g^{ii})^{-1/2}\tilde{\delta}^i_j \quad \text{(no sum)}. \qquad (2.17)$$

Finally, it will be convenient to introduce an explicit labeling for the vertices and edges of a tetrahedron in Fig. 1 below. Then the tetrahedron's volume is

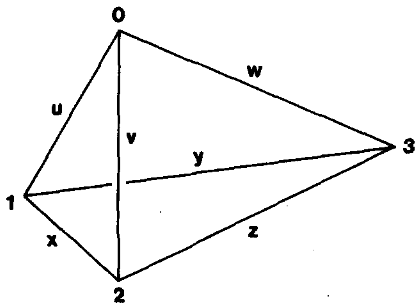


FIG. 1. A tetrahedron labeled in accordance with the conventions of the text. Letters assigned to each edge denote its squared length.

$$\begin{aligned}144\,V^2 = {} & -wx^2 - vy^2 - uz^2 - w^2x - v^2y - u^2z - xyz \\ & + xyv + xyw + xzu + xzw + yzu + yzv \\ & - xuv + xuw + xvw + yuv \\ & - yuw + yvw + zuv + zuw - zvw \end{aligned} \qquad (2.18)$$

and the area of a face with squared edges x,y,z has the form

$$\begin{aligned}A_0^2 := {} & A(\Delta 123)^2 \\ = {} & \tfrac{1}{16}[-x^2 - y^2 - z^2 + 2(xy + xz + yz)]\,. \end{aligned} \qquad (2.19)$$

Diagonal components of the metric are given by

$$g_{00} = \tfrac{1}{16}[-x - y - z + 3(u + v + w)] \qquad (2.20)$$

and off-diagonal components by

$$g_{01} = \tfrac{1}{16}[-5u - z + v + w + x + y]\,. \qquad (2.21)$$

(The remaining components are implied by the tetrahedral symmetry.) The contravariant metric has diagonal components of the form

$$g^{00} = (A_0/3V)^2 \qquad (2.22)$$

and off-diagonal components of the form

$$\begin{aligned}g^{01} = {} & -(1/144V^2)[-z^2 - 2uz + z(v + w + x + y) \\ & + (w - v)(x - y)]\,. \end{aligned} \qquad (2.23)$$

B. Space-times with simplicial slices

We turn now to the approximation of a smooth space-time by one in which the three-geometries on a family of spacelike hypersurfaces are simplicial. Consider a space-time $\mathscr{N} \approx M \times R$ with smooth metric $g_{\alpha\beta}$. Let T_t be a family of embeddings of M into $\mathscr{N}$, for which the images $M_t = T_t(M)$ are a sequence of spacelike hypersurfaces that foliate $\mathscr{N}$. Let t^α be the vector field tangent to the congruence of timelike curves $t \to T_t(x)$. (Each curve is the orbit in $\mathscr{N}$ of a point x in M.) One decomposes t^α into vectors perpendicular and parallel to M_t in the manner

$$t^\alpha = Nn^\alpha + N^\alpha\,, \qquad (2.24)$$

where

$$N = (-\nabla_\alpha t\nabla^\alpha t)^{1/2} \qquad (2.25)$$

is the lapse function,

$$n_\alpha = -N^{-1}\nabla_\alpha t \qquad (2.26)$$

is the future pointing unit normal to M_t, and N^α, defined by Eq. (2.24), is the shift vector. On $\mathscr{N}$, specifying a space-time metric is equivalent to specifying the lapse, shift, and "three-metric"

$$^3g_{\alpha\beta} = {}^4g_{\alpha\beta} + n_\alpha n_\beta \,. \tag{2.27}$$

Equivalently, if one uses the embedding T_t to pull the lapse, shift, and three-metric back to M, the four-metric can be described as a family of lapse functions N, shift vectors N^a and positive-definite three-metrics g_{ab} on M, parametrized by t: That is,

$$g_{ab} = T_t{}^*({}^3g)_{ab} \tag{2.28}$$

and $N^a(t)$ is the unique vector on M for which $T_{t*}(N^a) = N^\alpha$.

Given a simplicial decomposition of M, the embeddings T_t provide a simplicial decomposition of each spacelike hypersurface M_t. As noted at the beginning of Sec. II A, one can approximate the three-metric g_{ab} on M_t by assigning to each edge a length equal to the geodesic distance between its vertices. The resulting piecewise-flat metric on M_t then has in each simplex the form (2.5); the history of the spatial geometry is described by a set of time-dependent edge lengths. And an approximating space-time geometry may be completed by specifying on each cell values of the lapse and shift that agree with those of the smooth metric at the cell's barycenter.

The $3+1$ Regge theory we construct will then be the theory obtained by evaluating the Einstein action on space-times having the simplicial character described above: space-times foliated by simplicial slices, such that on each cell the lapse and shift are constant, and the three-metric is flat, determined by its edge lengths in accordance with Eq. (2.5). Note that the simplicial decomposition of M and the smooth family T_t of embeddings of M into $\mathscr{N}$ are given *before* one has chosen a metric. Consequently, although the time–time and time–space components of the metric are only piecewise-continuous, cell boundaries trace out smooth histories on $\mathscr{N}$.

C. Extrinsic curvature

In the standard $3+1$ formalism for the continuum theory, time derivatives of the three-metric on a hypersurface M_t enter the action via the extrinsic curvature of M_t. If n_α is its unit normal, the extrinsic curvature of M_t can be expressed in terms of the Lie derivative of ${}^3g_{\alpha\beta}$ along n_α,

$$K_{\alpha\beta} = -\tfrac{1}{2}\,{}^3g_\alpha^\gamma\,{}^3g_\beta^\delta\,\mathscr{L}_{\mathbf{n}}\,{}^3g_{\gamma\delta} \,, \tag{2.29a}$$

if as before, one identifies tensors orthogonal to n_α with tensors on M_t. Equivalently, for each t one can regard the extrinsic curvature as the pullback to M of $K_{\alpha\beta}$,

$$K_{ab} = (T_t^*K)_{ab} \,. \tag{2.29b}$$

From (2.28) we have

$$\partial_t\, g_{ab}(t) = (T_t^*\,\mathscr{L}_{\mathbf{n}} g)_{ab} \,, \tag{2.30}$$

whence, by (2.24) and (2.29a), we recover the familiar expression

$$K_{ab} = -(1/2N)\partial_t\, g_{ab} + (1/N)\nabla_{(a}N_{b)} \,. \tag{2.31}$$

Then, from Eq. (2.5), the affine components of K_{ab} have the form

$$K_{ij} = (1/4N)\tilde{\delta}_i{}^k\,\tilde{\delta}_k{}^l\dot{s}_{kl} + (1/N)\nabla_{(i}N_{j)} \,. \tag{2.32}$$

We will see in Sec. III B that the simplicial form of the momentum constraint leads one to specify a piecewise-constant shift vector. Evaluating the action will have the effect of approximating the covariant derivative $\nabla_a N_b$ as a finite difference, which we can obtain heuristically as follows. Denote by $*C$ the star of a cell C, the union of C together with the four cells that adjoin it. The difference $N^D{}_a - N^C{}_a$ between the values of N_a on two adjacent cells, C and D, is well defined because the metric is flat on $*C$: One knows how to parallel transport a vector from D to C. The problem of finding the gradient $\nabla_a N_b$ is thus reduced to approximating the gradient of a scalar.

Let f be constant on cells. Then ∇f is a distribution with support on the faces, and its average over a cell C is given by

$$\Delta^C{}_a f := \frac{1}{V_C}\int_C dV\,\nabla_a f. \tag{2.33}$$

With half of each δ function contributing to the integral (regarded as extending halfway across the boundary of C), we have

$$\int dV\,\nabla_a f = \frac{1}{2}\sum_{D\in *C} A_a^{CD}(f_D - f_C)$$

$$= \frac{1}{2}\sum_{D\in *C} A_a^{CD} f_D \,, \tag{2.34}$$

where f_C denotes the value of f on C, A_a^{CD} is the outward pointing normal to face $C\cap D$ with magnitude equal to the face's area, and Stokes' theorem in the form $\Sigma_D A_a^{CD} = 0$ has been used. Then

$$\Delta^C{}_a f = \frac{1}{2V_C}\sum_{D\in *C} A_a^{CD} f_D \,, \tag{2.35}$$

and for the representation of the gradient on C of any tensor that is piecewise constant on $*C$ we have the identical form

$$\Delta^C{}_a v_b = \frac{1}{2V_c}\sum_{D\in *C} A_a^{CD} v_{Db} \,. \tag{2.36a}$$

The analogous formula for the gradient of a tensor density ν_a of weight $\frac{1}{2}$ is

$$\Delta^C{}_a \nu_b = \frac{1}{2}\sum_{D\in *C} V_D^{-1} A_a^{CD} \nu_{Db} \,. \tag{2.36b}$$

The finite difference given here allows a discrete version of integration by parts for an integral over M. When σ^{ab} and τ_b are piecewise-constant tensors, the relation

$$\int dV\,\sigma^{ab}\nabla_a\tau_b = -\int dV\,\tau_b\nabla_a\sigma^{ab}$$

becomes

$$\sum_C V_C\sigma_C{}^{ab}\Delta^C{}_a\tau_b = -\sum_C V_C\tau_b^C\Delta^C{}_a\sigma^{ab} \,. \tag{2.37}$$

Proof:

$$\sum_C V_C\sigma_C{}^{ab}\Delta^C{}_a\tau_b$$

$$= \sum_{\substack{C\\ D\in *C}} \sigma_C{}^{ab}\frac{1}{2}A^{CD}_a\tau_{Db} = -\sum_{\substack{D\\ C\in *D}} \sigma_C{}^{ab}\frac{1}{2}A^{DC}_a\tau_{Db}$$

$$= -\sum_D V_D \Delta^D{}_a \sigma^{ab} \tau_{Db} ,$$

where we have used the identity

$$\sum_{\substack{C \\ D\in^*C}} = \sum_{\substack{D \\ C\in^*D}} \tag{2.38}$$

and the relation

$$A_a^{CD} = -A_a^{DC} . \tag{2.39}$$

The form for the extrinsic curvature corresponding to the discrete representation (2.36a) of the gradient is

$$K_{ab}^C = -\frac{1}{2N_C}(\dot{g}_{ab} - 2\Delta^C_{(a}N_{b)}) . \tag{2.40}$$

In practice, because the affine components of $N^D{}_a$ will be given with respect to D's affine chart, we need to relate the affine bases of adjacent cells. Let C and D have vertices 1–3 in common and denote the fourth vertex of each cell by O_C and O_D, respectively. Then, using for $i,j = 1,2,3$,

$$\mathbf{e}_i^C - \mathbf{e}_j^C = \mathbf{e}_i^D - \mathbf{e}_j^D \tag{2.41}$$

together with Eq. (2.16), we find, for all i,

$$\omega^i_{Da}(\delta^a_b - n^a n_b) = \omega^i_{Ca}(\delta^a_b - n^a n_b) , \tag{2.42}$$

where $\mathbf{n}$ is the unit normal to the common face $C \cap D$ (because $\mathbf{n}$ occurs quadratically, the relation is symmetric in C and D—independent of which unit normal is chosen). From Eq. (2.42), the components of a vector v^a along the affine bases of C and D are related by

$$v^{i_D} = v^{i_C} + (n^{i_D} - n^{i_C})\mathbf{n}\cdot\mathbf{v} , \tag{2.43}$$

where the components of $\mathbf{n}$ are given by Eq. (2.17). The corresponding relation between covariant components is then

$$v_{i_D} = v_{i_C} + n_{i_D} \sum_j (n^{j_C} - n^{j_D}) v_{j_C} . \tag{2.44}$$

III. ACTION AND FIELD EQUATIONS

A. 3+1 action and constraint equations

The Einstein action can be written in the form

$$I = \int dt\, L = \int dt\, dV (p^{ab}\dot{g}_{ab} - N\mathscr{H} - N^a \mathscr{H}_a) , \tag{3.1}$$

where the three-metric g_{ab}, its conjugate momentum tensor p^{ab}, the lapse N, and the shift N_a are regarded as independent variables; here

$$\mathscr{H} = -{}^3R + p^{ab}p_{ab} - \tfrac{1}{2}p^2 \tag{3.2}$$

and

$$\mathscr{H}^a = -2\nabla_b p^{ab} . \tag{3.3}$$

For a piecewise-flat metric with affine components given in terms of the edge lengths [as in Eq. (2.5)] by

$$g_{ij}^C = -\tfrac{1}{2}\tilde{\delta}_i^k \tilde{\delta}_j^l s_{kl} ,$$

the curvature is a distribution with support on the edges. The integral of the scalar curvature over a spacelike hypersurface is unambiguous,[2] but the potential term involves a somewhat arbitrary average value, N_ι, of the lapse on cells bounding the edge l_ι [see Eq. (3.7) below]:

$$\int dV N\, {}^3R = 2\sum_\iota N_\iota \theta_\iota l_\iota . \tag{3.4}$$

Here θ_ι is the deficit angle of edge l_ι, expressed in terms of edge lengths of cells that share l_ι in the following manner: By definition,

$$\theta = 2\pi - \sum_C \phi_C , \tag{3.5}$$

where ϕ_C is the dihedral angle between the two faces of cell C that have edge l in common. Then if A and A' are the areas of the two faces and V the volume of C, ϕ_C is given by

$$\sin\phi_C = \tfrac{3}{2} lV/AA' . \tag{3.6}$$

Finally, the volume of a cell C is given in terms of its squared edge lengths by Eq. (2.18), and the area of a triangle with squared edge lengths x,y,z is similarly given by Heron's equation, (2.19).

There is, however, some arbitrariness in how one chooses to discretize the lapse and shift—to represent them by a finite set of variables. One choice, in analogy with the metric, is to make them constant on each cell. Another, adopted by Piran and Williams, and which we initially used as well, is to specify the lapse N on each vertex; it can then be extended uniquely to the interior of each cell by demanding linearity in the affine coordinates. The analogous procedure for the shift vector fails, however, because the three-space does not have a differentiable structure at the vertices or edges (to define a vector at a vertex, one must give a separate value for each cell). In addition, as we shall see below, the momentum constraint has a natural form in the 3 + 1 formalism, and to use that form one must specify the shift on cells, not vertices. One should then specify the lapse on each cell as well to maintain lapse and shift as parallel and perpendicular components of a single vector t^α—and this is what we shall do. The choice has another advantage: if N is not constant on each cell, the action involves integrated averages of N and N^{-1}; so the form of the constraints and of the equations of motion is somewhat simplified by a piecewise-constant lapse. We resolve the remaining freedom in assigning an average value of the lapse to each edge in Eq. (3.4) by writing an angle-weighted average,

$$N_\iota \theta_\iota := \sum N_C\, \theta_{\iota C}, \tag{3.7a}$$

where $\theta_{\iota C}$ assigns to cell C part of the deficit angle θ_ι,

$$\theta_{\iota C} = 2\pi/c_\iota - \phi_C , \tag{3.7b}$$

with c_ι the number of cells sharing edge l_ι. On each face $C \cap D$, the shift will be assigned the value

$$N^a_{CD} = \tfrac{1}{2}(N^a_C + N^a_D) \tag{3.8}$$

(but any linear combination of the form $\alpha N^a_C + \beta N^a_D$, with $\alpha + \beta = 1$, will give the same action). Then, from Eqs. (3.2) and (3.4) the super-Hamiltonian is

$$\int dV N\mathscr{H} = -2\sum_\iota \theta_\iota N_\iota l_\iota + \sum_C N_C V_C \Big(p^{ab}_C p_{Cab} - \frac{1}{2}p_C^2\Big) , \tag{3.9}$$

where the metric on cell C is used to contract the indices of p^{ab}_C.

The momentum constraint is obtained in the continuum theory as the variational derivative of the action with respect to the shift vector. It has the form

$$\nabla_b p^{ab} = 0\,, \tag{3.10}$$

where p^{ab} is given in terms of the extrinsic curvature K_{ab} by

$$p^{ab} = -(K^{ab} - g^{ab} K_c{}^c)\,. \tag{3.11}$$

(The symbol π will be reserved for the associated tensor density, $\pi^{ab} = \sqrt{g} p^{ab}$.)

Because the metric is continuous across each face, the star $*C$ of any cell C is flat, and *one can define constant vectors on* $*C$. (If C is a cell, its star is the union of C and the four cells that share a face with C.) Dotted into a constant vector v^a, the momentum constraint has the form of Gauss' law

$$\nabla_b (p^{ab} v_b) = 0\,. \tag{3.12}$$

We can thus obtain one (vector) constraint for each cell C by requiring, for each constant vector v^a, that the flux of $p^{ab} v_b$ through the boundary of C vanishes:

$$\int_{\partial C} p^{ab} v_a\, dS_b = 0\,. \tag{3.13a}$$

The extrinsic curvature must be evaluated on the outer boundary in order to obtain a constraint that is not trivially satisfied by a $\mathbf{p}$ constant on each cell. Equation (3.13a) then becomes

$$\sum_{D\in *C} p_D^{ab} v_a A_{\ b}^{CD} = 0\,, \tag{3.13b}$$

or, in the notation of Eq. (2.36a),

$$\Delta_b^C p^{ab} = \frac{1}{2V_C} \sum_{D\in *C} p_D^{ab} A_{\ b}^{CD} = 0 \tag{3.14}$$

(the flatness of $*C$ is required to make the sum well defined).

This form of the momentum constraint shares with the continuum version the feature that it is automatically satisfied when p_{ab} is proportional to g_{ab}. Here the constraint holds for cell C when p_{ab} is proportional to g_{ab} on $*C$. This is the case for Regge models of homogeneous, isotropic spacetimes, and as a result (see Sec. IV A below) the discrete equations are identical in form to their continuum counterparts.

We shall now verify that by specifying the shift N_a on cells, with the convention (3.8), we reproduce the momentum constraint in the form (3.14) as the variational derivative of the action with respect to N_a^C. For p^{ab} piecewise constant, $D_b p^{ab}$ is a δ function distribution with support on the faces, and we have

$$\begin{aligned} -\int dV N_a \mathscr{H}^a &= 2\int dV N_a D_b p^{ab} = \sum_{\substack{C\\ D\in *C}} N_{Ca} (p_D^{ab} - p_C^{ab}) A_{\ b}^{CD} \\ &= \sum_{\substack{C\\ D\in *C}} N_{Ca} p_D^{ab} A_{\ b}^{CD}\,, \end{aligned} \tag{3.15}$$

or, from Eq. (2.36a),

$$-\int dV N_a \mathscr{H}^a = 2\sum_C V_C N_{Ca} \Delta_b^C p^{ab}\,. \tag{3.16}$$

The momentum constraint therefore takes the desired form (3.14)

$$0 = \frac{1}{V_C} \frac{\delta L}{\delta N_{Ca}} = 2\Delta_b^C p^{ab}\,. \tag{3.17}$$

The remaining term in the action has the form

$$\int dV p^{ab} \dot{g}_{ab} = -\frac{1}{2} \sum_C p_C^{ij} \dot{s}_{ij}\,, \tag{3.18}$$

where Eq. (2.5) has been used. Finally, combining Eqs. (3.9), (3.16), and (3.18), we have

$$\begin{aligned} L = \sum_C V_C \Big[p_C^{ab} \dot{g}_{ab}^C - N_C \Big(p_C^{ab} p_{ab}^C - \frac{1}{2} p_C^2 \Big) + 2N_{Ca} \Delta_b^C p^{ab} \Big] \\ + \sum_\iota 2N_\iota \theta_\iota l_\iota\,. \end{aligned} \tag{3.19}$$

By treating p^{ab} as an independent variable one avoids the divergent action that would arise from terms quadratic in the distribution $\nabla_a N_b$. Now, however, we can eliminate p^{ab}, obtaining an action that involves only the edge lengths, lapse, and shift. Since the Lagrangian (3.19) is already independent of $\dot{p}^{ab}$, its variation with respect to p^{ab} provides a constraint

$$\begin{aligned} 0 = \frac{1}{V_C} \frac{\partial L}{\partial p_C^{ab}} = -2N_C \Big(p_{ab}^C - \frac{1}{2} g_{ab}^C p^C \Big) \\ + \dot{g}_{ab}^C - 2\Delta_{(a}^C N_{b)}\,. \end{aligned} \tag{3.20}$$

[Use Eq. (2.37) to obtain the term involving the shift.] Equivalently,

$$K_{ab}^C = -(1/2N_C)(\dot{g}_{ab}^C - 2\Delta_{(a}^C N_{b)})\,, \tag{3.21}$$

in agreement with the expression for the extrinsic curvature expected from Eq. (2.40). With the constraint (3.20) used to express p_{ab} in terms of the remaining variables, the Lagrangian (3.19) becomes

$$\sum_\iota 2N_\iota \theta_\iota l_\iota + \sum_C N_C V_C (K_{ab}^C K^{Cab} - (K_C)^2)\,, \tag{3.22}$$

where K_{ab}^C is given by Eq. (3.21) and g_{ab}^C by Eq. (2.5).

An elegant form of the kinetic term was noticed by Lund and Regge[13] in their treatment of the 3 + 1 action for a complex of congruent cells. We recover this form by writing, in accordance with Eq. (2.15),

$$\begin{aligned} &K^C{}_{ab} K^{Cab} - (K^C)^2 \\ &= G_C^{abcd} K^C{}_{ab} K^C{}_{cd} = -\frac{1}{V_C{}^2} \frac{\partial^2 V_C{}^2}{\partial s_\iota \partial s_\kappa} K_\iota^C K_\kappa^C\,, \end{aligned} \tag{3.23}$$

where K_ι^C is the projection of $\mathbf{K}$ along the ιth edge l_ι^a:

$$K_\iota^C := K_{ab}^C l_\iota^a l_\iota^b = -(1/2N_C)(\dot{s}_\iota - 2\Delta^C N_\iota) \tag{3.24}$$

and

$$\Delta^C N_\iota := \Delta^C{}_a N_b l_\iota^a l_\iota^b\,. \tag{3.25}$$

Then from (3.22) we have

$$L = \sum_\iota 2N_\iota \theta_\iota l_\iota - \sum_C \frac{N_C}{V_C} \frac{\partial^2 V_C^2}{\partial s_\iota\, \partial s_\kappa} K_\iota^C K_\kappa^C\,. \tag{3.26}$$

Varying the action with respect to the lapse gives a discrete version of the Hamiltonian constraint, namely

$$0=\frac{\partial L}{\partial N_C}=\sum_{\iota\in C}2\theta_{\iota C}l_\iota+\frac{1}{V_C}\frac{\partial^2 V_C^2}{\partial s_\iota\,\partial s_\kappa}K^C_\iota K^C_\kappa \tag{3.27}$$

[in performing the variation, note that by Eq. (3.24), K^{ab} is proportional to N^{-1}].

The constraint equations (3.14) and (3.27) give a set of $4n_{\text{cell}}$ linear algebraic equations for the $4n_{\text{cell}}$ values of the shift and the squared lapse. They can therefore be solved explicitly to give an unconstrained action, when the determinant of the system is nonzero. We expect this to be generically true (that the system is nondegenerate), but have only verified it for a few simple complexes. For large complexes it may not be feasible to solve the constraints analytically, but the fact that it is in principle possible to do so guarantees that the constraints are preserved by the time evolution of the system. There remains the question of whether the system of constraints plus field equations will generically have a well-defined time evolution. Because only ordinary differential equations are involved, a *finite* nonsingular evolution is essentially guaranteed. As in the continuum theory, the *global* evolution can be singular. Here, in addition to curvature singularities, the edge lengths may fail to satisfy the triangle inequalities that allow them to bound flat cells, and coefficients of the terms with second time derivatives in the equations of motion may vanish.

B. Equations of motion

By varying the action with respect to the edge lengths one obtains the equations governing their time evolution. [We will regard as independent variables in the Lagrangian (3.22) the edge lengths, the covariant components of the shift vector in each cell along the affine basis of that cell, and the value of the lapse on each cell.]

The momentum conjugate to edge $l_\iota\equiv l_{ij}$ joining vertices i and j is given by

$$\pi^{ij}=\frac{\partial L}{\partial \dot s_{ij}}=-\sum_{C\ni l_{ij}}V_C p^{ij}_C\,. \tag{3.28}$$

Here $\mathbf{p}_C$ is a tensor [defined by Eqs. (3.11) and (2.40)], while V_C and $\boldsymbol{\pi}$ can be regarded as densities of weight 1.

Because of the form (3.7) of N_ι, the variation of the potential term has its usual simplicity,

$$\frac{\partial}{\partial s_\iota}2\sum_\kappa N_\kappa\theta_\kappa l_\kappa=2\sum_{C,\kappa}N_C\theta_{\kappa C}\frac{\partial l_\kappa}{\partial s_\iota}=\frac{N_\iota\theta_\iota}{l_\iota}, \tag{3.29}$$

arising from the simplicial form of Stokes' theorem.

To find $\partial T/\partial s_{ij}$, we use the identity

$$\frac{\partial g^{kl}}{\partial s_{ij}}=g^{i(k}g^{l)j}, \tag{3.30}$$

which follows from Eq. (2.9). Equations (2.7) and (3.30) imply the relation

$$\frac{\partial}{\partial s_{ij}}(VG^{klmn})K_{kl}K_{mn}$$
$$=2V(g^{k(i}G^{j)lmn}-\tfrac14 g^{ij}G^{klmn})K_{kl}K_{mn}\,. \tag{3.31}$$

Now the tensor K also depends on edge lengths when the shift is nonzero because the difference between the shift vector on adjacent cells C and D involves the parallel transport of n^D_a from D to C described by Eq. (2.44). (In the continuum theory, of course, the covariant derivative of the shift vector depends analogously on the metric.) We have

$$\frac{\partial}{\partial s_{ij}}K^C_{kl}=\frac{1}{N_C}\frac{\partial}{\partial s_{ij}}\Delta^C_{(k}N_{l)}\,, \tag{3.32}$$

and, from (2.44),

$$\frac{\partial}{\partial s_{ij}}(N^D_{k_C}-N^C_{k_C})=\sum_m\frac{\partial}{\partial s_{ij}}\left[n_{k_C}(n^{m_D}-n^{m_C})\right]N^D_{m_D}\,. \tag{3.33}$$

Equation (3.30) implies

$$\frac{\partial}{\partial s_{ij}}n_k=-\frac12 n_k n^i n^j, \tag{3.34}$$

since $n_k=\tilde\delta^j_k/\sqrt{g^{jj}}$ for D opposite the jth vertex of C. Then using Eqs. (3.30) and (3.34), a few lines of computation yield

$$\frac{\partial}{\partial s_{ij}}(N^D_{k_C}-N^C_{k_C})=n_{k_C}\gamma_{CD}{}^{ijm}N^D_{m_D}\,, \tag{3.35}$$

where

$$\gamma_{CD}{}^{ijm}=g^{m_D(i_D}n^{j_D)}-g^{m_C(i_C}n^{j_C)}$$
$$+n^{m_C}n^{i_C}n^{j_C}-\tfrac12 n^{m_D}(n^{i_C}n^{j_C}+n^{i_D}n^{j_D})\,. \tag{3.36}$$

Finally,

$$\frac{\partial}{\partial s_{ij}}\Delta^C_k N_l=\sum_{D\in^* C}\frac{A_{CD}}{2V_C}n_k n_l\gamma^{ijm}_{CD}N^D_m\,. \tag{3.37}$$

From Eqs. (3.29), (3.31), (3.32), and (3.37), the equation of motion for the edge $l_\iota=l_{ij}$ has the form

$$0=\mathscr{G}^{ij}:=\frac{d}{dt}\left(\frac{\partial L}{\partial \dot s_{ij}}\right)-\frac{\partial L}{\partial s_{ij}}$$
$$=\dot\pi^{ij}+\frac{N_\iota\theta_\iota}{l_\iota}-\sum_{C\ni l_\iota}2N_C V_C\Big[K^{ik}_C K_C{}^j{}_k-K_C K_C{}^{ij}$$
$$-\frac14 g_C{}^{ij}(K_C{}^{kl}K_{Ckl}-K^2_C)\Big]$$
$$-\sum_{\substack{C\\D\in^* C}}A_{CD}p_C{}^{kl}n_k n_l\gamma^{ijm}_{CD}N^D_m\,. \tag{3.38}$$

In contrast to the continuum theory, the lapse and shift cannot be chosen arbitrarily. There is in general no gauge freedom available to simplify the equations of motion, and in return for exactly conserved constraints one must accept a system of equations complicated by nonzero shift and nonconstant lapse.

If one regards the 3 + 1 equations as simply a tool for the numerical approximation of classical space-times, it may be more efficient to pick a lapse and shift to simplify the equations. The constraints would then be imposed only as initial conditions on the three-metric g_{ab} and its first time derivative. Although they would no longer be exactly conserved in the limit of small time intervals, the error after a finite time might be no larger than the numerical error for the full set of equations.

C. Hamiltonian formalism

In constructing a Hamiltonian formalism from the action of the last section a difficulty is encountered. Although the continuum action of Eq. (3.1) is already essentially in Hamiltonian form, with $\pi^{ab} = \sqrt{g}p^{ab}$ as the momentum conjugate to g_{ab}, the corresponding simplicial action given by Eq. (3.19) is deceptive. The kinetic term in the Hamiltonian,

$$T = \sum_C N_C V_C^{-1}\left(\pi_C{}^{ab}\pi_{Cab} - \frac{1}{2}\pi_C^2\right) \tag{3.39}$$

is nearly identical in form to its continuum counterpart; however, the momentum π^ι conjugate to the squared edge length $s_\iota = s_{ij}$ is not $\pi_C{}^{ij}$ but the sum

$$\pi^\iota = -\sum_C \pi_C{}^{ij}. \tag{3.40}$$

To write the Hamiltonian explicitly in terms of π^ι, one must solve Eq. (3.40) for $\pi_C{}^{ij}$, which is equivalent to inverting a band-diagonal matrix with e columns, where e is the number of edges in the complex: the locality of the continuum theory is lost. If, like $\dot{s}_{ij} \equiv \dot{s}_\iota$, the component $\pi_C{}^{ij}$ were really a function specified on the edge l_ι, having the same value π^ι for each cell C containing l_ι, then one could write

$$\pi^\iota = -c_\iota \pi_C{}^{ij} \tag{3.41}$$

with c_ι the number of cells sharing edge ι, and the difficulty would disappear. Unfortunately, a constant tensor σ^{ab} has contravariant affine components that differ from cell to cell, and it is the contravariant affine components of π^{ab} that comprise the momentum conjugate to s_{ij}. As one approaches the continuum, each tensor becomes essentially constant over many cells, and the contravariant affine components π^{ij} must therefore change from cell to cell while the covariant projections $\hat{\pi}_{ij}$ of π along the edges are cell independent.

This seems to us a genuine drawback of the theory, but there is a fairly natural way out. One can require that in the continuum limit cells in any small region become congruent, as is the case for a number of repetitive complexes (for example, the decomposition of R^3 into cubes subdivided as in Fig. 1). For two adjacent, congruent cells C and D, corresponding edges (and their time derivatives) are equal, and assigning adjacent cells the same value of π^{ij} is then consistent with the condition that in the continuum limit $\hat{\pi}_{Cij} = \hat{\pi}_{Dij}$. We will assume that the simplicial decomposition used has this property. We then use Eq. (3.41) to assign to each cell the value of $\pi_C{}^{ij}$ obtained from the single momentum π^ι conjugate to s_ι.

The kinetic term (3.39) now has the form

$$T = \sum_C N_C \widetilde{G}^{\,C}_{abcd}\pi_C{}^{ab}\pi_C{}^{cd} = \left(\sum_C N_C G^C{}_{\iota\kappa}\right)\pi^\iota\pi^\kappa, \tag{3.42}$$

where

$$\widetilde{G}^{\,C}{}_{abcd} = \frac{1}{2V_C}(g^C{}_{ac}g^C{}_{bd} + g^C{}_{ad}g^C{}_{bc} - g^C{}_{ab}g^C{}_{cd}) \tag{3.43}$$

and, with ι and κ labeling, respectively, edges l_{ij} and l_{kl},

$$G^C{}_{\iota\kappa} = \frac{1}{c_\iota c_\kappa}\widehat{G}_{ijkl} = \frac{1}{2V_C c_\iota c_\kappa}(s_{ik}s_{jl} + s_{il}s_{jk} - s_{ij}s_{kl}). \tag{3.44}$$

From Eq. (3.16), the term in the action involving the shift vector has the form

$$-2\sum_C N_{Ca}\Delta^C{}_b\pi^{ab} = 2\sum_C \Delta^C{}_a N_b \pi_C{}^{ab} \tag{3.45}$$

with $\Delta^C{}_a$ defined for tensor densities by Eq. (2.36b). Denoting the projection of $\Delta^C{}_a N_b$ along edge l_ι of C by

$$\Delta^C N_\iota = \Delta^C{}_a N_b l^a_\iota l^b_\iota, \tag{3.46}$$

we can write the Hamiltonian in the form

$$H = -\sum 2N_\iota\theta_\iota l_\iota + \sum_C N_C G^C{}_{\iota\kappa}\pi^\iota\pi^\kappa + 2\sum_{C,\iota}\frac{1}{c_\iota}\Delta^C N_\iota \pi^\iota \tag{3.47}$$

with a corresponding action

$$I = \int dt(\dot{\pi}^\iota s_\iota - H). \tag{3.48}$$

This Hamiltonian formalism appears not quite to agree with that recently given by Piran and Williams for the case of zero shift. The supermetric G of Eq. (3.44) includes factors c_ι that count the number of cells shared by each edge and these do not appear in the Piran–Williams version.

The equations of motion obtained from the Hamiltonian (3.47) have the form

$$\dot{s}_\iota = \frac{\partial H}{\partial \pi^\iota} = 2\sum_C N_C G^C{}_{\iota\kappa}\pi^\kappa + \frac{2}{c_\iota}\sum_C \Delta^C N_\iota \tag{3.49}$$

and

$$\dot{\pi}^\iota = -\frac{\partial H}{\partial s_\iota} = \frac{N_\iota\theta_\iota}{l_\iota} - \sum_C N_C\left(\frac{\partial}{\partial s_\iota}G^C{}_{\kappa\lambda}\right)\pi^\kappa\pi^\lambda - 2\sum_{C,\lambda}\frac{\partial}{\partial s_\iota}(\Delta^C N_\lambda)\frac{\pi^\lambda}{c_\lambda}, \tag{3.50}$$

where $(\partial/\partial s_\iota)(\Delta^C N_\lambda)$ is given by Eq. (3.37). The Hamiltonian constraint corresponding to cell C is

$$0 = \frac{\partial H}{\partial N_C} = -\sum_{l_\iota \in C} 2\theta_{\iota C} l_\iota + G^C{}_{\iota\kappa}\pi^\iota\pi^\kappa \tag{3.51}$$

and the momentum constraint is

$$0 = \frac{\partial H}{\partial N^C_a} = -2\Delta^C_b\pi^{ab}. \tag{3.52}$$

In the case of congruent cells and vanishing shift, the present action (3.48) is identical to that obtained from the Lagrangian (3.22) of Sec. III B above.

In the Lagrangian formalism of the last section, the constraints were preserved by the time evolution not because of a symmetry of the equations (as in the continuum theory) but because the constraints formed a linear system that could be solved to give the lapse and shift on each cell. But here, because the Hamiltonian is linear in lapse and shift, they do not appear in the constraints. Instead, the $4n_{\text{cell}}$ constraint equations restrict the initial values of the canonical variables—the edge lengths and their conjugate momenta. Because of the lack of general covariance, these constraints do not commute with the Hamiltonian, and the commutators form a new set of $4n_{\text{cell}}$ constraints that must be solved. The new constraints, however, are linear in the lapse and shift. In general they determine the initial values of the lapse and shift and provide one additional restriction on the canonical variables.

In the Bergmann–Dirac[32] terminology the constraints have the following structure in the Hamiltonian theory of this section. The primary constraints are that the momenta p_N and p_{Na} associated with lapse and shift vanish; and the momentum and Hamiltonian constraints, Eqs. (3.51) and (3.52), are the corresponding secondary constraints. Commutators of these secondary constraints with the Hamiltonian (3.47) are not linear combinations of constraints (they do not vanish "weakly" and therefore constitute additional constraints) *but the commutators do depend on the lapse and shift*. Consequently they do not commute with the primary constraints, and the Dirac procedure terminates: The commutators do not generate further secondary constraints.

What has been lost in discretizing the theory is the character of the constraints: The loss of gauge invariance means that they are no longer first class.

IV. SPACES WITH CONGRUENT CELLS

A. Equations of motion

The simplest space-times of the Regge calculus are those with congruent cells, for which the lapse is constant and the shift vanishes. Several authors have constructed models of this kind, in which a homogeneous, isotropic space-time is approximated by a simplicial complex with a dust source of constant density comoving with the lattice.[4–11] We show here that regardless of the simplicial decomposition, the equations of motion can always be cast in the form of the continuum equations

$$2\frac{\ddot a}{a}+\frac{\dot a^2}{a^2}+\frac{k}{a^2}=0\,, \tag{4.1}$$

$$3\frac{\dot a^2}{a^2}+3\frac{k}{a^2}=8\pi\rho\,. \tag{4.2}$$

A more precise phrasing of the result is as follows.

Theorem: Consider any tiling of S^3 or R^3 by congruent simplices. Let $g_{ab}(t)$ satisfy the $3+1$ Regge equations with zero shift and constant lapse, and for which the source is dust of constant density ρ in a frame comoving with the lattice. Then

$$g_{ab}(t)=e^{2\alpha}g_{ab}(0)\,, \tag{4.3}$$

where

$$\dot\alpha=\dot a/a\,, \tag{4.4}$$

for $a(t)$ and $\rho(t)$ satisfying Eqs. (4.1) and (4.2).

A similar result holds for the non-simply-connected space-times obtained by identification from S^3 and R^3. That is, if G is any finite group that acts freely and transitively on S^3 (R^3), the spherical space S^3/G (hyperbolic space R^3/G) is locally isometric to S^3 (R^3) and in the continuum theory, its time evolution is again a homogeneous, isotropic space-time satisfying the same field equations. Any tiling of a spherical or hyperbolic space by congruent cells induces a tiling of the covering space (S^3 or R^3), and the corresponding space-times, evolved by the $3+1$ Regge equations, are again locally isometric. Thus we have the following corollary.

Corollary: The above theorem holds with S^3 and R^3 replaced by S^3/G and R^3/G.

To prove the theorem, we first need the equations of motion for the edge lengths. As noted in Sec. IIIA, the momentum constraint is automatically satisfied when the conjugate momentum p_{ab} is proportional to the metric, and from Eq. (4.3), this will be the case if the shift is set to zero. When the lapse is constant, a constant rescaling of time, $t\to t/N$ gives the equations in their form for unit lapse—a remnant of the continuum gauge freedom. Then for a metric with time dependence given by Eq. (4.3), the kinetic term in the Lagrangian (3.22) is simply

$$T_C=-6V\dot\alpha^2 \tag{4.5}$$

and for the equation of motion associated with the edge $l_\iota\equiv l_{ij}$, we obtain

$$0=\frac{d}{dt}\left(\frac{\partial L}{\partial\dot s_\iota}\right)-\frac{\partial L}{\partial s_\iota}=(2\ddot\alpha+3\dot\alpha^2)c_\iota Vg^{ij}-\frac{\theta_\iota}{l_\iota}\,. \tag{4.6}$$

The Hamiltonian constraint (3.27) for a single cell C has the form

$$\frac{1}{V}\sum_{\iota\in C}\frac{1}{c_\iota}\theta_\iota l_\iota+3\dot\alpha^2=8\pi\rho\,, \tag{4.7}$$

where c_ι is the number of cells sharing the edge l_ι and θ_ι is the deficit angle of that edge.

Proof of Theorem: The dynamical equations (4.6) imply that for each edge $l_\iota\equiv l_{ij}$ the length $a(t)$ defined by

$$a^2(t):=-c_\iota Vg^{ij}l_\iota/\theta_\iota \tag{4.8}$$

has the same value, independent of the choice of edge. The metric's time dependence (4.3) implies corresponding time dependence for the edge lengths, volume and contravariant metric, namely

$$l\propto e^{\alpha},\quad V\propto e^{3\alpha},\quad g^{ab}\propto e^{-2\alpha}\,. \tag{4.9}$$

Then

$$\dot a/a=\dot\alpha\,, \tag{4.10}$$

and the equations of motion take the form

$$0=2\frac{d}{dt}\left(\frac{\dot a}{a}\right)+3\left(\frac{\dot a}{a}\right)^2-\frac{\theta_\iota}{c_\iota Vg^{ij}l_\iota}=2\frac{\ddot a}{a}+\frac{\dot a^2}{a^2}+\frac{1}{a^2}\,, \tag{4.11}$$

in agreement with Eq. (4.1).

Similarly, using Eq. (4.8) to replace θ in the first term of the constraint (4.7), we have

$$\sum_\iota\frac{1}{c_\iota}\theta_\iota l_\iota=\frac{1}{a^2}\sum_\iota 2s_\iota\frac{\partial V}{\partial s_\iota}=\frac{3V}{a^2}\,,$$

and Eq. (4.7) becomes

$$3\dot a^2/a^2+3/a^2=8\pi\rho\,. \tag{4.12}$$

Thus with $a(t)$ chosen in accordance with (4.8), the equations of motion and the Hamiltonian constraint assume their continuum forms, Eqs. (4.1) and (4.2), respectively, as claimed.

While the time dependence of the metric is given by the continuum equations, its form on the initial hypersurface is determined by Eq. (4.8), which can be solved to give the six independent edge lengths in terms of $a(t)$.

B. Examples of spherical spaces

A virtue of 3 + 1 Regge calculus is the ease with which one can treat space-times whose spatial topology is nontrivial. In this section we present explicit solutions to the 3 + 1 equations on compact spaces with topologies of the form S^3/H, where H is a subgroup of SU(2) acting on $S^3 \approx \mathrm{SU}(2)$ by left multiplication. We look first at the quaternion space S^3/Q, where Q is the eight-element quaternion subgroup of SU(2), $\{\pm 1, \pm i\sigma_x, \pm i\sigma_y, \pm i\sigma_z\}$. We then consider the lens spaces $L(p,1) = S^3/Z_p$. Regarded as a sequence of tilings of S^3, they have the property that as $p \to \infty$, the sequence approaches the continuum limit in one spatial direction: piecewise-flat curves parallel to an equator become smooth, although the poles remain singular.

Given a finite (order n) subgroup H of SO(4) that acts freely on S^3, one can partition S^3 into n congruent cells (not tetrahedra) intersecting only at their boundaries, such that the orbit of any cell is the set of all n cells. Then S^3/H can be constructed from a cell by identifying pairs of faces, and a tiling of any cell by m tetrahedra provides a tiling of S^3 by nm tetrahedra.

In particular, the quaternion space S^3/Q can be obtained by identifying opposite faces of a cube after a rotation by 90°, as shown in Fig. 2(a).[33] Figure 2(b) illustrates a simplicial decomposition of S^3/Q chosen to respect the identification of Fig. 2(a). Within the Regge framework, the topology of the space S^3/Q is described simply by stating which edges are to be identified; its geometry is then fixed as usual by its edge lengths, where identified edges are of course assigned the same length.

For the smooth metric on the quaternion space induced by that on S^3, the symmetries of the cube are isometries. In the simplicial case, one expects edge lengths to share the cube's symmetry: that is, a solution with congruent tetrahedra to the 3 + 1 equations (4.1), (4.2), and (4.8) should satisfy

$$l_{02} = l_{03}, \quad l_{12} = l_{13}\,.$$

As noted above, the $G^{pq} = 0$ equations, (4.6), imply that the quantity

$$q_\iota := c_\iota V g^{ij} l_\iota / \theta_\iota \tag{4.13}$$

has the same value for each edge l_{ij}. From Eqs. (3.5) and (3.6), together with the fact that edges l_{01} and l_{12} each belong to four cells, we have

$$q_{01} = \frac{4Vg^{01}l_{01}}{2\pi - 4\sin^{-1}[(3/2)(l_{01}V/A_2A_3)]} \tag{4.14}$$

and

$$q_{12} = \frac{4Vg^{12}l_{12}}{2\pi - 4\sin^{-1}[(3/2)(l_{12}V/A_0A_3)]}, \tag{4.15}$$

where A_i is the area of the face opposite vertex i. Similarly, since edges l_{02} and l_{23} each belong to six cells, we have

$$q_{02} = \frac{6Vg^{02}l_{02}}{2\pi - 6\sin^{-1}[(3/2)(l_{02}V/A_1A_3)]}, \tag{4.16}$$

$$q_{23} = \frac{6Vg^{23}l_{23}}{2\pi - 6\sin^{-1}[(3/2)(l_{23}V/A_0A_1)]}. \tag{4.17}$$

The l_{03} and l_{13} equations are identical to those for l_{02} and l_{12} by the cubic symmetry. The form of the four equations (4.14)–(4.17) suggests an additional symmetry

$$l_{23} = l_{02} = l_{03} \equiv x, \quad l_{01} = l_{12} = l_{13} \equiv \sqrt{\alpha}x\,. \tag{4.18}$$

When these relations hold, we have

$$A_0 = A_2 = A_3\,, \tag{4.19}$$

and the equality of the q_{ij} is reduced to the single relation

$$q_{01}/q_{02} = 1\,. \tag{4.20}$$

That is, from (4.18) there are only two independent edge lengths. Their ratio α is time independent and determined by Eq. (4.20); their magnitude is then given in terms of the length $a(t)$ by any component of Eq. (4.8). From Eq. (2.23) for g^{01} (and the analogous equation for g^{02}), we find

$$g^{02}/g^{01} = 2\alpha - 1\,, \tag{4.21}$$

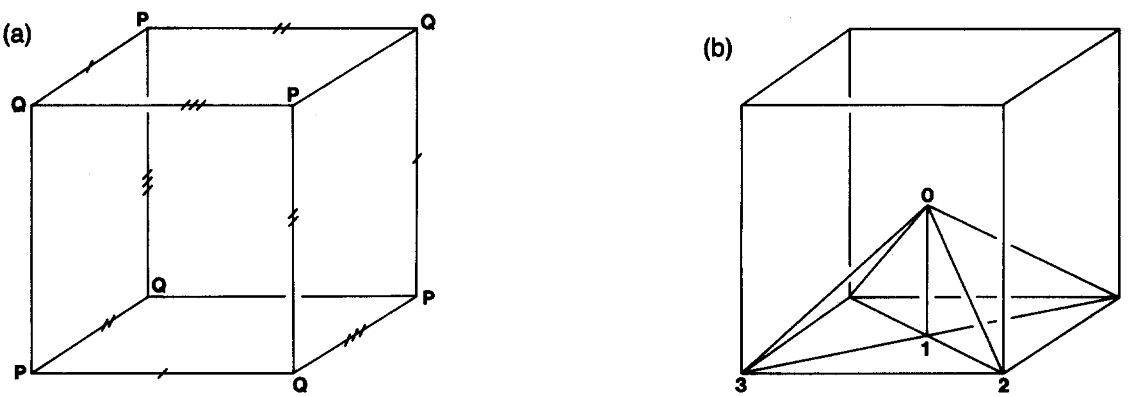


FIG. 2. (a) The quaternion space S^3/Q is constructed by identifying opposite faces of a cube after a relative rotation by 90°. Edges labeled by the same number of hash marks (and vertices labeled by the same numeral) are thereby identified. (b) Six of the 24 tetrahedron in our simplicial decomposition of S^3/Q. The remaining tetrahedra are implied by cubic symmetry.

while Eqs. (2.18) and (2.19) imply

$$A_0 = (x/4)(4\alpha - 1)^{1/2}, \quad A_1 = (\sqrt{3/4})x\,, \tag{4.22}$$

and

$$V = \tfrac{1}{12}x^{3/2}(3\alpha - 1)^{1/2}\,. \tag{4.23}$$

When length, area, volume, and components of the contravariant metric are expressed in terms of α and x by Eqs. (4.18) and (4.21)–(4.23), Eq. (4.20) becomes

$$\frac{2\pi - 4\sin^{-1}[2\sqrt{3\alpha^2 - \alpha}/(4\alpha - 1)]}{2\pi - 6\sin^{-1}[2\sqrt{3\alpha - 1}/\sqrt{3(4\alpha - 1)}]} \cdot \frac{3(2\alpha - 1)}{2\sqrt{\alpha}} = 1\,. \tag{4.24}$$

There is a unique real solution for α, namely

$$\alpha = 0.5641\,, \tag{4.25a}$$

and Eqs. (3.5) and (3.6) then give as the independent deficit angles

$$\theta_{01} = 23.4°, \quad \theta_{02} = 6.02°\,. \tag{4.25b}$$

To compare the simplicial geometry with a continuum solution, one can choose as the comparison smooth spacetime the Friedmann solution with the same time dependence, so that the difference is entirely in the spatial geometry. That is, by the theorem of Sec. IV A, there is a smooth solution with the same density $\rho(t)$; and for that solution the radius $a(t)$ of the three-sphere ($2\pi^2 a^3$ is the volume V) is the simplicial length $a(t)$ of Eq. (4.8). The continuum metric restricted to the spherical space S^3/Q makes each edge into a geodesic. Here l_{01} and l_{02} are, respectively, $\frac{1}{6}$ and $\frac{1}{8}$ the circumference of S^3, and their squared ratio is

$$\alpha_{\text{smooth}} = 0.5625\,, \tag{4.26}$$

only 0.3% larger than its value for the simplicial manifold. The volumes do not agree to the same accuracy,

$$V_{\text{smooth}}/V_{\text{simplicial}} = 1.07\,, \tag{4.27}$$

but from the Hamiltonian constraint (4.7), the average curvatures are identical

$$R_{\text{smooth}} = \overline{R}_{\text{simplicial}} \equiv \frac{1}{V}\int dV R\,, \tag{4.28}$$

when one compares solutions with the same density.

This accuracy is consistent with that obtained in previous work, where S^3 is approximated by one of the three complexes of equilateral tetrahedra.[6,11,14] These have $c_E = 3,4$, or 5 cells sharing an edge, and a total number of cells $N = 5$, 16, or 600, respectively. Because each cell is equilateral, the consistency conditions implied by Eq. (4.8) are automatically satisfied. The three-geometry is determined by the single deficit angle $\theta = 2\pi - c_E \cos^{-1}(\frac{1}{3})$, and Eq. (4.8) then implies

$$a^2 = (\sqrt{2}/24)(c_E/\theta)l^2\,, \tag{4.29}$$

where l is the edge length. Again the smooth Friedmann universe with the same density has radius $a(t)$ and a scalar curvature identical with the average simplicial scalar curvature. The difference between the simplicial and smooth three-geometry is measured by

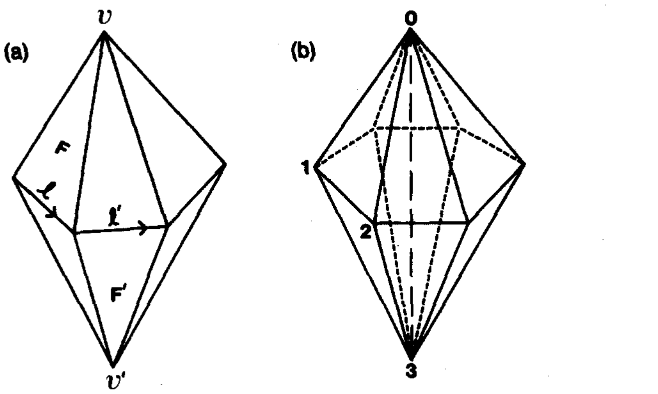


FIG. 3. (a) The lens space $L(6,1)$ is constructed from a 12-sided polyhedron by identifying each upper face with a corresponding lower face as shown. (b) Adding the vertical edge l_{03} decomposes the polyhedron into six congruent simplices. The vertices of one of these are labeled.

$$\frac{V_{\text{smooth}}}{V_{\text{simplicial}}} = \frac{3^{1/2}2^{1/4}}{\pi^2} N\left(\frac{\theta}{c_E}\right)^{3/2} = 1.67,1.32,1.030\,, \tag{4.30}$$

respectively, for the equilateral complexes with 5, 16, and 600 cells. If one views the quaternion space above as a decomposition of the three-sphere into $24 \times 8 = 192$ cells, its accuracy, measured by Eq. (4.27), is, as expected, between that of the 16-cell and 600-cell solutions.

Let us turn now to the family of lens spaces $L(p,1)$ obtained by identifying the upper and lower hemispheres of the surface of a ball after a relative rotation by $2\pi/p$. Figure 3(a) illustrates the identification and Fig. 3(b) shows the simplicial decomposition we have used. Edges l_{01}, l_{02}, l_{13}, and l_{23} each share four tetrahedral cells, while edges l_{03} and l_{12} each share p cells. As in the previous example, we consider a space-time filled with dust of constant density. If we set

$$l_{01} = l_{02} = l_{13} = l_{23} \tag{4.31a}$$

and

$$l_{03} = l_{12}\,, \tag{4.31b}$$

the equations are symmetric under the simultaneous replacements

$$0 \leftrightarrow 1, \quad 2 \leftrightarrow 3\,. \tag{4.32}$$

Consistency of the G^{ij} equations (4.6) then determines the ratio

$$\alpha = l_{01}/l_{03} \tag{4.33}$$

via the equation

$$q_{01}/q_{03} = 1\,, \tag{4.34}$$

where q_{01} has the same form (4.14) it assumed for the quaternion space, while q_{03} is given by

$$q_{03} = \frac{pVg^{03}l_{03}}{2\pi - p\sin^{-1}(\frac{3}{2}l_{03}V/A_1A_2)}\,. \tag{4.35}$$

Proceeding as in the previous example, we obtain from Eq. (4.34) the condition

$$\frac{\theta_{03}}{\theta_{01}} = \frac{2\pi - p\sin^{-1}[(\alpha - \frac{1}{2})/(\alpha - \frac{1}{4})]^{1/2}}{2\pi - 4\sin^{-1}[\alpha(\alpha - \frac{1}{2})/(\alpha - \frac{1}{4})]^{1/2}} = p\,\frac{\alpha - \frac{3}{4}}{\alpha^{1/2}}\,. \tag{4.36}$$

We are interested here in the limiting case where the number of cells becomes infinite. Then the lengths $l_{03} = l_{12}$ shrink to zero (while the other lengths remain finite), and one expects the space to become smooth along the l_{03} and l_{12} directions: the deficit angle θ_{01} associated with the finite edges should vanish in this quasicontinuum limit. As $p \to \infty$, $\alpha \to \infty$, and the angles θ_{01} and θ_{03} assume the asymptotic form

$$\theta_{01} = \alpha^{-1} + O(\alpha^{-2}) , \tag{4.37a}$$

$$\theta_{03} = 2\pi - p\alpha^{-1/2} + O(\alpha^{-3/2}) , \tag{4.37b}$$

Eq. (4.36) then yields the asymptotic solution

$$\alpha = p^2/\pi^2 + O(p) \tag{4.38}$$

and for large p, $\theta_{03} \to \pi$, while, as expected, $\theta_{01} \to 0$.

C. The momentum constraint for congruent cells

From Eq. (3.13), the momentum constraint for a cell C has the form

$$\sum_{D \in {}^*C} p_D^{ab} v_a A_{\ b}^{CD} = 0 , \tag{4.39}$$

where v^a is any constant vector on $^*C$. If the cells of $^*C$ are congruent and the momentum tensor p^{ab} respects the congruence $C \to D$, we have

$$p_C{}^{i_C j_C} = p_D{}^{i_D j_D} \equiv p^{ij} , \tag{4.40}$$

where i_D is the vertex of D corresponding to vertex i_C of C. Equation (4.39) may then be written in the following form:

$$\sum_k \frac{p^{kk}}{A_{(k)}^2} v^k = 0 , \tag{4.41}$$

where the indices refer to vertices of cell C and $A_{(k)}$ is the area of the face opposite vertex k of C.

Proof: If the face $C \cap D$ is opposite vertex k (k_D of D and k_C of C), then

$$A_{j_C}^{CD} = -A_{j_D}^{CD} = -3V\tilde{\delta}_j^k$$

and from Eq. (2.44) we obtain the relation

$$\begin{aligned} v_{i_D} &= v_{i_C} + n_{i_D} \sum_j (n^{j_C} - n^{j_D}) v_{j_C} \\ &= v_{i_C} - \frac{18V^2}{A_{(k)}{}^2} \tilde{\delta}_i^k v^k \quad \text{(no sum)} \end{aligned}$$

between the components of $\mathbf{v}$ along the affine bases of C and D. The constraint then takes the form

$$0 = \sum_D p_D{}^{i_D j_D} v_{i_D} A_{j_D}^{CD} = (-3V)(-18V^2) \sum_k \frac{p^{kk}}{A_{(k)}{}^2} v^k , \tag{4.42}$$

where we have used the relation $\Sigma_k p^{ik} = 0$.

Equivalently, since $\Sigma_k v^k = 0$, and (4.42) holds for all $\mathbf{v}$, the momentum constraint is equivalent to the three relations

$$0 = \frac{p^{11}}{A_1{}^2} - \frac{p^{00}}{A_0{}^2} = \frac{p^{22}}{A_2{}^2} - \frac{p^{00}}{A_0{}^2} = \frac{p^{33}}{A_3{}^2} - \frac{p^{00}}{A_0{}^2} . \tag{4.43}$$

The assumption of congruent cells is similar to assuming homogeneity in the continuum theory. In the latter case, additional symmetries, implying isotropy, must be imposed to ensure uniform expansion. Here, although there is no exact isotropy, a similar result holds: A complex built from congruent cells will generically have six independent edge lengths—all edge lengths of one cell can be freely chosen. By setting pairs of edges equal, one restricts the geometry, and if enough edges are fixed in this way the momentum constraint will determine the expansion rate of every edge in terms of a single parameter. When this is done, *regardless of which edges are set equal, the momentum constraint always implies uniform expansion*—the same logarithmic time derivative $d \log l/dt$ for each edge.

For example, suppose in Fig. 2, one sets

$$l_{01} = l_{23} \equiv u, \quad l_{02} = l_{03} \equiv v, \quad l_{12} = l_{13} \equiv y . \tag{4.44}$$

For congruent cells (and zero shift) the conjugate momentum has the form

$$p^{ij} = (V/8)(g^{ik}g^{jl} - g^{ij}g^{kl})\dot{s}_{kl} . \tag{4.45}$$

Then using Eq. (2.7), we find two independent momentum constraints,

$$\frac{p^{11}}{A_1{}^2} - \frac{p^{00}}{A_0{}^2} = 0 \Rightarrow (v-y)\dot{u} + (4y-u)\dot{v} - (4v-u)\dot{y} = 0 \tag{4.46}$$

and

$$\begin{aligned} \frac{p^{22}}{A_2{}^2} - \frac{p^{00}}{A_0{}^2} = 0 \Rightarrow (y-v)(u^2 + yu - vu - 2y^2 + 2vy)\frac{\dot{u}}{u} \\ - (u^2 + 2v^2 - 2y^2 + yu - 5uv)\dot{y} \\ + (-u + v - y)\dot{v} = 0 . \end{aligned} \tag{4.47}$$

Eliminating $\dot{v}$ between Eqs. (4.46) and (4.47), we obtain

$$\frac{\dot{y}}{y} = \frac{\dot{u}}{u} . \tag{4.48}$$

Equations (4.48) and (4.47) then imply uniform expansion

$$\frac{\dot{v}}{v} = \frac{\dot{y}}{y} = \frac{\dot{u}}{u} . \tag{4.49}$$

We have proved the same result on a case by case basis for the various other possible distinct ways of equating enough edge lengths that the momentum constraint determines the fractional expansion of each independent length. Although in every case the constraint implies uniform expansion, this feature is not apparent from its original form (4.41), and we are aware of no simple, general proof.

V. COMMENTS ON MINISUPERSPACE QUANTIZATION

As emphasized by Kuchar,[34] the Einstein field equations are closely analogous to the equations of motion for a relativistic particle moving in a curved (infinite-dimensional) space with indefinite metric and a time-dependent potential. The analogy is drawn by parametrizing the particle's time—introducing an arbitrary time coordinate τ—and noting that

$$N = \frac{d\tau}{dt}$$

plays the role of the lapse. In the 3 + 1 Regge theory, the analogy is exact for the case of congruent cells. When one deparametrizes the theory by solving the Hamiltonian constraint for the lapse,[35] the action is that of a relativistic particle moving in a six-dimensional curved space-time (signature $- + + + + +$), and the particle is *free*. As in the continuum case, there is a timelike hypersurface-orthogonal conformal Killing vector.

When all cells are congruent the shift can be set to zero, and from Eqs. (3.24) and (3.27) the Lagrangian takes the form

$$L = 2N \sum_\iota \frac{\theta_\iota l_\iota}{c_\iota} + \frac{1}{N} G^{\iota\kappa} \dot{s}_\iota \dot{s}_\kappa , \tag{5.1}$$

where we have removed an overall factor, dividing by the total number of edges. Here c_ι is as usual the number of cells sharing the ιth edge, the sum is over the six edges of a single cell, and the metric $G^{\iota\kappa}$ is given by

$$G^{\iota\kappa} = -\frac{1}{4V} \frac{\partial^2 V^2}{\partial s_\iota \, \partial s_\kappa} . \tag{5.2}$$

The single Hamiltonian constraint,

$$2 \sum_\iota \frac{\theta_\iota l_\iota}{c_\iota} - \frac{1}{N^2} G^{\iota\kappa} \dot{s}_\iota \dot{s}_\kappa , \tag{5.3}$$

has the immediate solution

$$N = \left[\frac{\frac{1}{2} G^{\iota\kappa} \dot{s}_\iota \dot{s}_\kappa}{(\Sigma_\lambda \, \theta_\lambda l_\lambda / c_\lambda)} \right]^{1/2} . \tag{5.4}$$

Using Eq. (5.4) to eliminate the lapse N, we obtain the unconstrained Lagrangian

$$L = [\widetilde{G}^{\iota\kappa} \dot{s}_\iota \dot{s}_\kappa]^{1/2} , \tag{5.5}$$

where

$$\widetilde{G}^{\iota\kappa} = 8 \sum_\lambda \frac{\theta_\lambda l_\lambda}{c_\lambda} G^{\iota\kappa} = -2 \left(\sum_\lambda \frac{\theta_\lambda l_\lambda}{c_\lambda} \right) \left(\frac{1}{V} \frac{\partial^2 V^2}{\partial s_\iota \, \partial s_\kappa} \right) . \tag{5.6}$$

Because the free-particle action is invariant under a reparametrization of time, the Hamiltonian constraint reappears in the form

$$\widetilde{G}_{\iota\kappa} \pi^\iota \pi^\kappa - 1 = 0 \tag{5.7a}$$

or

$$G_{\iota\kappa} \pi^\iota \pi^\kappa - 2 \sum \frac{\theta_\iota l_\iota}{c_\iota} = 0 , \tag{5.7b}$$

equivalent to Eq. (3.47) specialized to congruent cells. The signature of the metric G can be found by looking at the matrix

$$V^{\iota\kappa} = -\frac{\partial^2 V^2}{\partial s_\iota \, \partial s_\kappa} \tag{5.8}$$

in the special case when all lengths are equal. The eigenvalues of (5.8) have signs $- + + + + +$, and its determinant is proportional to V^4

$$\det \| V^{\iota\kappa} \| = -2^{-10} 3^{-8} V^4 . \tag{5.9}$$

Since the determinant is negative definite, the signature will not change under changes in the edge lengths, as long as the triangle inequalities are obeyed. Thus when the triangle inequalities are obeyed (and space is not exactly flat), $\widetilde{G}$ has signature $- + + + + +$ as well.

Because the squared volume is a polynomial cubic in the squared edge lengths, under a uniform expansion

$$s_\iota \to \lambda s_\iota , \tag{5.10}$$

the metric $\widetilde{G}$ changes by a conformal factor,

$$\psi_\lambda \widetilde{G}^{\iota\kappa}(s) = \lambda^{-2} \widetilde{G}^{\iota\kappa}(s) . \tag{5.11}$$

The vector generating this expansion,

$$\xi_\iota(s) = s_\iota , \tag{5.12}$$

is thus a conformal Killing vector, with

$$\mathscr{L}_\xi \widetilde{G}_{\iota\kappa} = s_\lambda \frac{\partial}{\partial s_\lambda} \widetilde{G}_{\iota\kappa} - \widetilde{G}_{\lambda\kappa} \frac{\partial}{\partial s_\lambda} s_\iota - \widetilde{G}_{\iota\lambda} \frac{\partial}{\partial s_\lambda} s_\kappa = -2 \widetilde{G}_{\iota\kappa} .$$

It is timelike, because V^2 cubic in the squared edge lengths implies

$$-V^{\iota\kappa} s_\iota s_\kappa = -6V^2 . \tag{5.13}$$

In attempting to quantize this minisuperspace theory, one faces two of the principal difficulties encountered in the canonical approach to the continuum theory: The curvature is not bounded below, and the metric $\widetilde{G}$ has a conformal symmetry, but no true Killing vector. If one tries to exploit the fact that the Lagrangian (5.5) describes a free particle with unit rest mass, moving in a curved space-time, one encounters the "third quantization" problem[34]: because the "particle" is not massless (equivalently, because the supermetric and the curvature have different conformal weights) a Klein–Gordon quantization will involve either an indefinite inner product or particle production, and one does not know how to interpret the creation of "particles" on superspace.

An alternative, Schrödinger-like, quantization (like that used by Blyth and Isham[36]) is suggested by the fact that the conformal Killing vector is hypersurface orthogonal. If one chooses as a timelike variable the volume V, the five remaining degrees of freedom are the independent edge lengths of a unit-volume tetrahedron, and the restriction of the metric $G_{\iota\kappa}$ to this five-dimensional space is positive definite. Edge lengths S^ι of a unit-volume cell are defined by

$$S^\iota = V^{-2/3} s^\iota . \tag{5.14}$$

Because only five lengths are independent, the conjugate momenta

$$\Pi^\iota = \frac{\partial L}{\partial S_\iota} \tag{5.15}$$

are related by

$$S_\iota \Pi^\iota = 0 . \tag{5.16}$$

Then the Hamiltonian constraint (5.7) has the form

$$2 \sum_\iota \frac{\theta_\iota l_\iota}{c_\iota} + \frac{3}{8} V (\Pi^V)^2 - \frac{1}{4} V^{-4/3} G_{\iota\kappa} \Pi^\iota \Pi^\kappa = 0 \tag{5.17}$$

or

$$\Pi^V = H \equiv \left[\frac{2}{3} V^{-7/3} G_{\iota\kappa} \Pi^\iota \Pi^\kappa - \frac{16}{3} V^{-1} \sum \frac{\theta_\iota l_\iota}{c_\iota} \right]^{1/2} . \tag{5.18}$$

This suggests the Schrödinger equation

$$\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial V}, \qquad (5.19)$$

where $G_{\iota\kappa}\Pi^\iota\Pi^\kappa$ is replaced by the covariant Laplacian of the metric $G_{\iota\kappa}$ restricted to the five-space.

However, the integrated curvature at fixed volume ranges from $-\infty$ to ∞, even if the triangle inequalities are imposed: Tetrahedra that are either nearly flat or are long and thin can have arbitrarily large integrated curvatures $(\Sigma(\theta l/c))$ for bounded volume. Thus the operator $\hat{H}$ will not in general be Hermitian. The Schrödinger equation (5.19) does make sense on the subspace of state vectors ψ with support on edge lengths for which $\Sigma(\theta l/c)$ is nonpositive.

For the complementary subspace of state vectors with support on edge lengths for which the integrated curvature is positive, one can obtain a Hermitian Hamiltonian by identifying Π^V as the timelike coordinate and solving the constraint (5.5) for V. But the formalism is clumsy, because the equation is a fifth-order polynomial in V, and the Hamiltonian is thus not available in closed form.

Finally, we mention the general situation where the cells are not congruent. Here the momentum constraint is not automatically satisfied, but we shall set the lapse to zero by fiat. The Hamiltonian constraint on each cell can then be solved as above,

$$N_C = \frac{\frac{1}{2}G^{\iota\kappa}_C\dot{s}_\iota\dot{s}_\kappa}{\Sigma_{\lambda\in C}\theta_\lambda l_\lambda/c_\lambda}, \qquad (5.20)$$

and the Lagrangian becomes a sum of free-particle Lagrangians of the form (5.5)

$$L = \sum_C [\tilde{G}^{\iota\kappa}_C\dot{s}_\iota\dot{s}_\kappa]^{1/2}, \qquad (5.21)$$

in which each coordinate s occurs as a coordinate of more than one "particle." The vector ξ_ι of Eq. (5.11) is here a conformal Killing vector of each metric $\tilde{G}_C$.

ACKNOWLEDGMENTS

We would like to thank L. Parker for helpful discussions and R. Williams and T. Piran both for discussions and for providing us with a copy of the Lund–Regge notes. J. F. is also indebted for conversations to many of the participants at the Santa Barbara Workshop on Discrete Gravity, including J. B. Hartle, K. Kuchar, T. Regge, and R. Sorkin.

This work was supported in part by National Science Foundation Grant No. PHY-8303597.

[1]T. Regge, Nuovo Cimento **19**, 558 (1961).
[2]R. Sorkin, Phys. Rev. D **12**, 385 (1975).
[3]J. W. Wheeler, "Geometrodynamics and the issue of the final state," in *Relativity, Groups, and Topology*, edited by C. de Witt and B. de Witt (Gordon and Breach, New York, 1964).
[4]C.-Y. Wong, J. Math. Phys. **12**, 70 (1971).
[5]P. A. Collins and R. M. Williams, Phys. Rev. D **5**, 1908 (1972).
[6]P. A. Collins and R. M. Williams, Phys. Rev. D **7**, 965 (1973).
[7]P. A. Collins and R. M. Williams, Phys. Rev. D **10**, 3537 (1974).
[8]S. M. Lewis, Phys. Rev. D **25**, 306 (1982).
[9]J. D. Porter, Ph.D. thesis, Magdalen College, Oxford, 1982.
[10]L. C. Brewin, Ph.D. thesis, Monash University, Clayton, Victoria, Australia, 1983.
[11]R. M. Williams, Gen. Relativ. Gravit. **17**, 559 (1985).
[12]T. Piran and R. Williams, Phys. Lett. B **163**, 331 (1985).
[13]F. Lund and T. Regge (private communication).
[14]T. Piran and R. M. Williams, Phys. Rev. D **33**, 1622 (1986).
[15]M. Rocek and R. M. Williams, Phys. Lett. B **104**, 31 (1981).
[16]M. Rocek and R. M. Williams, in *Quantum Structure of Space and Time*, edited by M. J. Duff and C. J. Isham (Cambridge U. P., Cambridge, 1982).
[17]J. Fröhlich, "Regge calculus and discretized gravitational functional integrals," unpublished, 1982.
[18]N. P. Warner, Proc. R. Soc. London Ser. A **383**, 359 (1982).
[19]M. Rocek and R. M. Williams, Z. Phys. C **21**, 371 (1984).
[20]G. Feinberg, R. Friedberg, T. D. Lee, and H. C. Ren, "Lattice gravity near the continuum limit," preprint, 1985.
[21]H. W. Hamber and R. M. Williams, Nucl. Phys. B **248**, 392 (1984).
[22]B. Berg, Phys. Rev. Lett. **55**, 904 (1985).
[23]H. W. Hamber and R. M. Williams, Phys. Lett. B **157**, 368 (1985).
[24]H. W. Hamber and R. M. Williams, "Two-dimensional simplicial quantum gravity," preprint, 1985.
[25]J. B. Hartle, J. Math. Phys. **26**, 804 (1985).
[26]J. B. Hartle, J. Math. Phys. **27**, 287 (1986).
[27]A. Jevicki and M. Ninomiya, "Functional formulation of Regge gravity," preprint, Brown University, 1985.
[28]M. Martellini and A. Marzuoli, "Cone subdivision procedure for Regge lattices: A renormalization group approach to quantum gravity," Abstract for Fourth Marcel Grossman Meeting, 1985.
[29]H. Romer and M. Zahringer, "Functional integration and the diffeomorphism group in Euclidean lattice quantum gravity," preprint, Universität Freiburg, 1985.
[30]R. M. Williams, Class. Quantum Grav. (in press).
[31]R. Sorkin, J. Math. Phys. **16**, 2432 (1975); **19**, 1800 (1978).
[32]P. A. M. Dirac, *Lectures in Quantum Mechanics* (Academic, London, 1967).
[33]J. A. Wolf, *Spaces of Constant Curvature* (McGraw–Hill, New York, 1967).
[34]K. Kuchar, "Canonical methods of quantization," in *Quantum Gravity 2*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Oxford U.P., Oxford, 1981).
[35]R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1865 (1962).
[36]W. F. Blyth and C. J. Isham, Phys. Rev. D **11**, 768 (1975).

Kinematic and dynamic properties of conformal Killing vectors in anisotropic fluids

R. Maartens and D. P. Mason
Department of Applied Mathematics, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2001, South Africa
M. Tsamparlis
Department of Physics, Section Astronomy–Astrophysics–Mechanics, University of Athens, Athens GR15771 Zografou, Greece

(Received 4 April 1986; accepted for publication 28 July 1986)

An example from a perfect fluid FRW space-time is presented to show that a conformal Killing vector (CKV) need not map fluid flow lines into fluid flow lines. Kinematic properties of the Lie derivative along a CKV of timelike and spacelike unit vectors are derived and applied to the fluid unit four-velocity vector. Dynamic properties of special conformal Killing vectors (SCKV) in a fluid with anisotropic pressure and vanishing energy flux are obtained using Einstein's field equations. It is shown that a SCKV maps both fluid flow lines and integral curves of n^a into themselves, where n^a is the unit spacelike vector of anisotropy. The relation between the anisotropic pressure components and the energy density is considered. By means of an example from a radiationlike viscous fluid FRW space-time it is shown that the dynamic results depend crucially on the vanishing of the energy flux vector. The extension of the dynamic results to a fluid with arbitrary stress tensor and zero energy flux vector is examined.

I. INTRODUCTION

The study of conformal Killing vectors (CKV) in fluid space-times has recently attracted some interest.[1–3] In this paper we will examine kinematic and dynamic properties of anisotropic fluids that admit a conformal motion. Unlike the kinematic properties, which are largely independent of the physical nature of the fluid, the dynamic properties depend on the kind of fluid through Einstein's field equations. We consider in detail fluids with anisotropic pressure and vanishing energy flux. The stress tensor has a dynamically defined preferred direction with symmetry in the orthogonal two-plane. They include as a special case perfect fluids in which the pressure is isotropic. We also examine the generalization of the results to anisotropic fluids with arbitrary stress tensor.

Suppose a fluid space-time admits a CKV ξ^a:

$$\mathscr{L}_\xi g_{ab} = 2\psi g_{ab}\,, \tag{1.1}$$

where $\mathscr{L}_\xi$ stands for the Lie derivative along ξ^a and $\psi(x^c)$ is the conformal factor. We denote by u^a the unit four-velocity vector of the fluid. In Sec. II we will establish the kinematic result

$$\mathscr{L}_\xi u^a = -\psi u^a + v^a\,, \tag{1.2}$$

where v^a is some vector orthogonal to u^a. Now in general v^a will be nonzero: ξ^a will not in general map integral curves of u^a into integral curves of u^a or equivalently ξ^a will not map fluid flow lines into fluid flow lines. This can be demonstrated by the following example. Consider a perfect fluid Friedmann–Robertson–Walker (FRW) space-time with $k=0$, which in conformal coordinates has metric

$$ds^2 = R^2(\eta)(-d\eta^2 + dx^2 + dy^2 + dz^2) \tag{1.3}$$

and fluid four-velocity

$$\mathbf{u} = R^{-1}\partial_\eta\,. \tag{1.4}$$

The vector

$$\boldsymbol{\xi} = \eta\,\partial_x + x\,\partial_\eta \tag{1.5}$$

is a CKV (see Ref. 3), and

$$\mathscr{L}_\xi \mathbf{u} = [\boldsymbol{\xi},\mathbf{u}] = x(R^{-1})_{,\eta}\,\partial_\eta - R^{-1}\,\partial_x \neq -\psi\mathbf{u}\,. \tag{1.6}$$

This counterexample shows that the statement in the literature[1] that the identity

$$\mathscr{L}_\xi u^a = -\psi u^a \tag{1.7}$$

follows from symmetry for a CKV ξ^a is not valid in general. Herrera *et al.*[1] based their kinematic derivation of (1.7) on the equation

$$\mathscr{L}_\xi\, dx^a = 0\,. \tag{1.8}$$

Now there exists a close relationship between (1.7) and the projection of (1.8) onto the rest space of u^a. For, since $dx^a = u^a\,d\tau$, where τ denotes proper time, we have

$$\mathscr{L}_\xi\, dx^a = u^a\,\mathscr{L}_\xi\, d\tau + d\tau\,\mathscr{L}_\xi u^a\,. \tag{1.9}$$

But $(d\tau)^2 = -ds^2 = -g_{ab}\,dx^a\,dx^b$ and therefore using also (1.1),

$$\mathscr{L}_\xi\, d\tau = \psi\, d\tau - u_a\,\mathscr{L}_\xi\, dx^a\,. \tag{1.10}$$

Substituting from (1.10) into (1.9) gives

$$h^a_b\,\mathscr{L}_\xi\, dx^b = (\mathscr{L}_\xi\, u^a + \psi u^a)\,d\tau\,; \tag{1.11}$$

thus

$$\mathscr{L}_\xi\, u^a = -\psi u^a \Leftrightarrow h^a_b\,\mathscr{L}_\xi\, dx^b = 0\,. \tag{1.12a,b}$$

Since (1.12a) is not necessarily satisfied by a CKV ξ^a then neither is (1.12b) and therefore the stronger condition (1.8) need not hold for a CKV.

The spacelike vector v^a in (1.2) may be expressed in terms of the fluid vorticity tensor ω^{ab} and this leads to kinematic conditions for (1.7) to be satisfied. As a particular case of a general kinematic expression for v^a established in Sec. II we find that when $\xi_a u^a = 0$,

$$v^a = 2\omega^{ab}\xi_b \,. \tag{1.13}$$

If the fluid is irrotational then (1.7) is satisfied by any CKV ξ^a orthogonal to u^a while in a rotational fluid, (1.7) is satisfied by a CKV ξ^a orthogonal to u^a if and only if ξ^a is parallel to the local fluid vorticity vector ω^a.

By introducing dynamics through the Einstein field equations we show that (1.7) may be satisfied in an anisotropic fluid if the energy flux vector q^a vanishes and ξ^a is a special conformal Killing vector (SCKV). A SCKV is defined by (1.1) and the condition $\psi_{;ab} = 0$ and includes homothetic motions as a particular case. We also show with the aid of Einstein's field equations that the unit spacelike vector of anisotropy n^a in a fluid with anisotropic pressure satisfies

$$\mathscr{L}_\xi\, n^a = -\psi n^a \,, \tag{1.14}$$

if ξ^a is a SCKV. Equation (1.14) is not true in general for a CKV.

Cahill and Taub[4] showed that spherically symmetric perfect fluid solutions admitting a homothetic motion represent the relativistic generalization of the self-similar solutions of classical hydrodynamics—and are thus important for the study of explosions and shock waves. [In fact Cahill and Taub pointed to the conditional nature of (1.7), when they stated that it followed from the transformation property of the Einstein tensor and from the perfect fluid form of the field source (see p. 7, Ref. 4).] Cahill and Taub used the physically motivated self-similar transformations to arrive at a homothetic KV that is isotropic (invariant under the rotational KV's), but neither orthogonal nor parallel to u^a. In a recent series of papers,[5] Herrera and Ponce de León have found exact spherically symmetric solutions including anisotropic fluids, by assuming the existence of an isotropic CKV ξ^a orthogonal to u^a. No motivation is given for the condition $\xi^a u_a = 0$, but it is clearly the simplest case allowing for a proper CKV and leading to nonstatic exact solutions; in particular, (1.7) holds because $\xi^a u_a = 0$ and the fluid is irrotational. Geometrically, their assumptions amount to assuming the existence of an isotropic intrinsic CKV in the surfaces $t = \text{const}$ (see Sec. II). This forces the metric to be isotropically conformal to a spatially homogeneous metric (a special Kantowski–Sachs metric).

The paper may be outlined as follows. In Sec. II kinematic results for the Lie derivative along a CKV ξ^a of the timelike and spacelike unit vectors u^a and n^a are derived. The special case in which $\xi_a u^a = 0$ is considered and the geometrical and physical interpretation of (1.7) is discussed. The fluid energy-momentum tensor is considered in Sec. III. For a fluid with anisotropic pressure the pressure parallel to the preferred direction n^a, $p_\parallel$, is different from the pressure perpendicular to n^a, $p_\perp$; when $p_\parallel = p_\perp$ and the energy flux vector q^a vanishes the fluid reduces to a perfect fluid. Dynamic results for a fluid with anisotropic pressure and $q^a = 0$ are derived in Sec. IV. The Lie derivative of Einstein's field equations is decomposed with respect to u^a and n^a. Our approach is different from that of Herrara *et al.*[1] who used (1.7) and (1.14) to evaluate the Lie derivative of the field equations; we use the Lie derivative of the field equations to prove that (1.7) and (1.14) are satisfied in this fluid if ξ^a is a SCKV. In Sec. V we generalize to include the cosmological constant and a possible nonzero magnetic field two equations of state obtained by Herrera *et al.*[12] relating $p_\parallel$, $p_\perp$, and the total energy density μ, which hold when ξ^a is a SCKV orthogonal to u^a. We also consider the case not discussed by Herrera *et al.*[1] in which ξ^a is parallel to u^a. In Sec. VI a counterexample is given to demonstrate that the dynamic results depend critically on the vanishing of the energy flux vector q^a. The extension of the dynamic results to a fluid with $q^a = 0$ but with arbitrary stress tensor is considered. Finally concluding remarks are made in Sec. VII.

The notation and conventions of Ellis[6] for relativistic fluid dynamics are followed throughout.

II. LIE DERIVATIVES: KINEMATIC RESULTS

In this section we restrict discussion to kinematic results. In subsequent sections we introduce dynamics through Einstein's field equations and we examine how the kinematic results derived here may be extended. Suppose that ξ^a is a CKV satisfying (1.1); ξ^a may be either timelike, null, or spacelike.

Consider first any unit vector X^a, which may be either timelike or spacelike. Then $X_a X^a = \epsilon$, where $\epsilon = -1$ if X^a is timelike and $\epsilon = +1$ if X^a is spacelike. We prove that

$$\mathscr{L}_\xi\, X^a = -\psi X^a + Y^a \,, \tag{2.1}$$

$$\mathscr{L}_\xi\, X_a = +\psi X_a + Y_a \,, \tag{2.2}$$

for some vector Y^a satisfying $X_a Y^a = 0$. In order to establish (2.1) we observe that we can always write

$$\mathscr{L}_\xi\, X^a = \alpha X^a + Y^a \,, \tag{2.3}$$

for some scalar α and vector Y^a satisfying $X_a Y^a = 0$. To obtain α we contract (2.3) with X_a:

$$\alpha = \epsilon^{-1} X_a\, \mathscr{L}_\xi\, X^a \,. \tag{2.4}$$

But since $X_a X^a = \epsilon$ we have

$$X_a\, \mathscr{L}_\xi\, X^a + X^a\, \mathscr{L}_\xi\, X_a = 0 \,, \tag{2.5}$$

and hence by writing $\mathscr{L}_\xi\, X_a = \mathscr{L}_\xi (g_{at} X^t)$ and using (1.1) we find that

$$X_a\, \mathscr{L}_\xi\, X^a = -\psi\epsilon \,. \tag{2.6}$$

Thus $\alpha = -\psi$, which establishes (2.1). Equation (2.2) follows directly from (2.1) with the aid of (1.1):

$$\mathscr{L}_\xi\, X_a = 2\psi X_a + g_{at}\, \mathscr{L}_\xi\, X^t \,. \tag{2.7}$$

We now apply the results (2.1) and (2.2) to the fluid unit four-velocity vector u^a and to any spacelike unit vector n^a: $u_a u^a = -1$, $n_a n^a = +1$. Later n^a will be identified with the unit vector of anisotropy in the energy-momentum tensor for a fluid with anisotropic pressure. We have

$$\mathscr{L}_\xi\, u^a = -\psi u^a + v^a \,, \tag{2.8}$$

$$\mathscr{L}_\xi\, u_a = \psi u_a + v_a \,, \tag{2.9}$$

where $u_a v^a = 0$ and

$$\mathscr{L}_\xi n^a = -\psi n^a + m^a , \quad (2.10)$$

$$\mathscr{L}_\xi n_a = \psi n_a + m_a , \quad (2.11)$$

where $n_a m^a = 0$.

Identities (2.10) and (2.11) hold whether or not $n_a u^a = 0$. It is easily verified that if $n_a u^a = 0$ then

$$v_a n^a + m_a u^a = 0 . \quad (2.12)$$

For, since $n_a u^a = 0$ we have

$$n_a \mathscr{L}_\xi u^a + u^a \mathscr{L}_\xi n_a = 0 , \quad (2.13)$$

and (2.12) is derived by substituting from (2.8) and (2.11) into (2.13).

The spacelike vector v^a can be expressed in terms of the vorticity tensor ω^{ab}. For we can always write $\xi^a = \alpha u^a + \beta^a$, where $\beta_a u^a = 0$ and $\alpha = -\xi_a u^a$, and therefore

$$\mathscr{L}_\xi u_a = \dot\alpha u_a + \alpha \dot u_a - \alpha_{,b} h^b_a + 2\beta^b u_{[a;b]} , \quad (2.14)$$

where an overhead dot denotes covariant differentiation along a fluid particle world line.[6] But since

$$u_{a;b} = \theta_{ab} + \omega_{ab} - \dot u_a u_b , \quad (2.15)$$

where θ_{ab} is the rate of expansion tensor and $\dot u_a$ is the four-acceleration vector, Eq. (2.14) becomes

$$\mathscr{L}_\xi u_a = (\dot\alpha + \dot u_b \beta^b) u_a + \alpha \dot u_a - \alpha_{,b} h^b_a + 2\omega_{ab}\beta^b , \quad (2.16)$$

which may be rewritten in terms of ξ^a instead of β^a as

$$\mathscr{L}_\xi u_a = (\dot\alpha + \dot u_b \xi^b) u_a + \alpha \dot u_a - \alpha_{,b} h^b_a + 2\omega_{ab}\xi^b . \quad (2.17)$$

It follows directly from (2.9) and (2.17) that

$$\psi = \dot\alpha + \dot u_b \xi^b , \quad (2.18)$$

$$v_a = 2\omega_{ab}\xi^b + \alpha(\dot u_a + (\log \alpha^{-1})_{,b} h^b_a) . \quad (2.19)$$

Consider first the special case in which $\xi_a u^a = 0 = -\alpha$. Then (2.18) and (2.19) reduce to

$$\psi = \dot u_b \xi^b , \quad (2.20a)$$

$$v_a = 2\omega_{ab}\xi^b . \quad (2.20b)$$

From (2.20a) it follows that a CKV ξ^a orthogonal to u^a is necessarily a Killing vector (KV) if ξ^a is also orthogonal to $\dot u^a$ or if the flow is geodesic. From (2.20b) we have

$$\xi_a u^a = 0 \Rightarrow \mathscr{L}_\xi u^a = -\psi u^a + 2\omega^{ab}\xi_b , \quad (2.21)$$

and therfore if $\xi_a u^a = 0$ then

$$\mathscr{L}_\xi u^a = -\psi u^a \Leftrightarrow \omega^{ab}\xi_b = 0 . \quad (2.22)$$

We note also that if $\xi_a u^a = 0$ and the vorticity $\omega \neq 0$ then

$$\omega^{ab}\xi_b = 0 \Leftrightarrow \xi^a \text{ is parallel to } \omega^a . \quad (2.23)$$

If the fluid is irrotational then (1.7) is satisfied by any CKV orthogonal to u^a. If the fluid is rotational then (1.7) is satisfied by a CKV orthogonal to u^a if and only if ξ^a is parallel to ω^a. Herrera *et al.*[1] proved that if $\xi_a u^a = 0$ then

$$\mathscr{L}_\xi u_a = \psi u_a \Rightarrow \omega_{ab}\xi^b = 0 , \quad (2.24)$$

but their conclusion that $\omega_{ab}\xi^b = 0$ for all CKV ξ^a orthogonal to u^a is not valid because $\mathscr{L}_\xi u_a = \psi u_a$ is not necessarily satisfied by a CKV orthogonal to u^a.

Second, if ξ^a is parallel to u^a, then (1.7) is clearly satisfied. Note that in this case (2.19) shows that α^{-1} is an acceleration potential, while (2.15) contracted with σ^{ab} shows that $\sigma_{ab} = 0$:

$$\xi^a = \alpha u^a \Rightarrow \dot u_a = -(\log \alpha^{-1})_{,b} h^b_a , \quad \sigma_{ab} = 0 . \quad (2.25)$$

We conclude this section by considering two geometrical interpretations and a physical interpretation of (1.7).

Consider first the Lie derivative of the fluid projection tensor $h_{ab} = g_{ab} + u_a u_b$. A simple calculation based on (1.1) and (2.9) gives

$$\mathscr{L}_\xi h_{ab} = 2\psi h_{ab} + 2u_{(a} v_{b)} , \quad (2.26)$$

and therefore since $v_a u^a = 0$,

$$\mathscr{L}_\xi u^a = -\psi u^a \Leftrightarrow \mathscr{L}_\xi h_{ab} = 2\psi h_{ab} . \quad (2.27a,b)$$

Hence (2.27a) is satisfied if and only if ξ^a is a conformal motion of the fluid projection tensor h_{ab}. Now, in an irrotational fluid the rest spaces orthogonal to u^a at each point form global spacelike hypersurfaces orthogonal to u^a with intrinsic metric tensor h_{ab}. If further ξ^a is orthogonal to u^a then ξ^a lies in these hypersurfaces and when $\omega = 0$ we know from (2.21) that (2.27a) is satisfied. Thus (2.27b) also holds and ξ^a must be an *intrinsic* CKV of the hypersurfaces.

An alternative geometrical interpretation of (1.7) is that ξ^a maps integral curves of u^a into integral curves of u^a. When $\psi \neq 0$, the mapping involves a rescaling of u^a by a change of parameter, but the family of integral curves of u^a as a whole is mapped into itself. Thus ξ^a is a dynamical symmetry of the fluid flow. One consequence of this dynamical symmetry property is that new constants of the fluid motion may be generated from existing constants. For suppose that f is a constant of the fluid motion; then

$$\mathbf{u} f \equiv f_{,a} u^a \equiv \dot f = 0 . \quad (2.28)$$

It follows that $\boldsymbol{\xi} f \equiv f_{,a}\xi^a$ is also a constant of the fluid motion if (1.7) is satisfied:

$$\mathbf{u}(\boldsymbol{\xi} f) = [\mathbf{u},\boldsymbol{\xi}] f + \boldsymbol{\xi}(\mathbf{u} f) = \psi \mathbf{u} f + \boldsymbol{\xi}(\mathbf{u} f) = 0 . \quad (2.29)$$

A physical interpretation of (1.7) exists in terms of material curves in the fluid. A material curve in a fluid is a curve that consists at all times of the same fluid particles and therefore it moves with the fluid as the fluid evolves; it is sometimes said to be "frozen-in" to the fluid. The integral curves of ξ^a are therefore material curves if (1.7) is satisfied.

An important special case of material curves occurs when $\xi_a u^a = 0$. If $\omega \neq 0$ and (1.7) is satisfied then the CKV ξ^a must be parallel to ω^a. The integral curves of ξ^a are therefore vortex lines that will be material curves in the fluid. This is a consequence of a symmetry property of the flow and not of the physical nature of the fluid. It can be shown conversely that if ξ^a is a CKV orthogonal to u^a and if the integral curves of ξ^a are material curves then they must be vortex lines if $\omega \neq 0$. For, the vector $\omega^{ab}\xi_b$ is orthognal to both u^a and ξ^a and it therefore follows from (2.21) that $\omega^{ab}\xi_b = 0$ since otherwise the integral curves of ξ^a would not move with the fluid. Thus ξ^a must be parallel to ω^a.

III. ENERGY-MOMENTUM TENSOR

Before applying Einstein's field equations in Sec. IV we consider here the fluid energy-momentum tensor.

For a fluid with anisotropic pressure and vanishing energy flux the energy-momentum tensor may be written in the form

$$T^{ab} = \mu u^a u^b + p_{\parallel} n^a n^b + p_{\perp} p^{ab}, \tag{3.1}$$

where μ is the total energy density of the fluid measured by an observer with four-velocity u^a; n^a is a unit spacelike vector orthogonal to u^a,

$$n_a n^a = +1, \quad n_a u^a = 0; \tag{3.2}$$

p^{ab} is the projection tensor onto the two-plane orthogonal to u^a and n^a,

$$p^{ab} = g^{ab} + u^a u^b - n^a n^b; \tag{3.3}$$

and $p_{\parallel}$ and $p_{\perp}$ denote the pressure parallel and perpendicular to n^a, respectively. When $p_{\parallel} = p_{\perp}$, (3.1) reduces to the energy-momentum tensor for a perfect fluid. In a local comoving inertial system in which

$$u^a = \delta^a_0, \quad n^a = \delta^a_1, \quad g^{ab} = \mathrm{diag}[-1,1,1,1], \tag{3.4}$$

(3.1) becomes

$$T^{ab} = \mathrm{diag}[\mu, p_{\parallel}, p_{\perp}, p_{\perp}]. \tag{3.5}$$

Fluids with anisotropic pressure have been studied extensively in the recent literature.[1,5,7–12] We comment here briefly on two examples of physical systems described by the energy-momentum tensor (3.1).

First an energy-momentum tensor of the form (3.1) is derived for a fluid consisting of two perfect fluid components.[8,11] If the two perfect fluids are decoupled from each other or interact weakly then anisotropic pressure may be important in the time evolution of the fluid. In astrophysics this situation could exist in neutron stars.[11]

Second, a strong magnetic field in a plasma in which the particle collision density is low can cause the pressure along and perpendicular to the magnetic field lines to be unequal[13,14]; if the collision density is high an isotropic pressure distribution would quickly evolve. For this system the total energy-momentum tensor is

$$T^{ab} = \mu u^a u^b + p_{\parallel} n^a n^b + p_{\perp} p^{ab} + T^{ab}_{\mathrm{EM}}, \tag{3.6}$$

where $n^a = H^a/H$, H^a is the local magnetic field measured by u^a, and T^{ab}_{EM} is the electromagnetic energy-momentum tensor, which we will take to be the Minkowski tensor. If the local electric field E^a vanishes then the Minkowski tensor for a pure magnetic field is[15]

$$T^{ab}_{\mathrm{EM}} = \frac{\lambda}{2} H^2 u^a u^b - \frac{\lambda}{2} H^2 n^a n^b + \frac{\lambda}{2} H^2 p^{ab}, \tag{3.7}$$

where λ is the magnetic permeability. On substituting from (3.7) into (3.6) we obtain

$$T^{ab} = \bar{\mu} u^a u^b + \bar{p}_{\parallel} n^a n^b + \bar{p}_{\perp} p^{ab}, \tag{3.8}$$

where

$$\bar{\mu} = \mu + (\lambda/2) H^2, \tag{3.9}$$

$$\bar{p}_{\parallel} = p_{\parallel} - (\lambda/2) H^2, \tag{3.10}$$

$$\bar{p}_{\perp} = p_{\perp} + (\lambda/2) H^2. \tag{3.11}$$

From (3.10) we see that the magnetic field contributes a tensile stress along the field line. The total energy-momentum tensor (3.8) takes the form (3.1) for a fluid with anisotropic pressure.

The most general expression for the energy-momentum tensor of a fluid is

$$T^{ab} = \mu u^a u^b + p h^{ab} + 2q^{(a} u^{b)} + \pi^{ab}, \tag{3.12}$$

where p is the isotropic pressure, q^a is the energy flux vector relative to u^a, and π^{ab} is the tracefree stress tensor or anisotropic pressure tensor: $q_a u^a = 0$, $\pi_{ab} u^b = 0$, $\pi^a{}_a = 0$. The energy-momentum tensor (3.1) can be written in the form (3.12) with $q^a = 0$ and

$$p = \tfrac{1}{3}(2p_{\perp} + p_{\parallel}), \tag{3.13}$$

$$\pi^{ab} = (p_{\perp} - p_{\parallel})(\tfrac{1}{3} h^{ab} - n^a n^b). \tag{3.14}$$

A stress tensor of the form (3.14) is also obtained for a collisionfree gas in locally rotationally symmetric Bianchi spacetimes, where n^a is the local preferred direction.[16]

IV. LIE DERIVATIVES: DYNAMIC RESULTS

The Lie derivative along a CKV ξ^a of Einstein's field equations has been evaluated by Herrera *et al.*[1] If the cosmological constant Λ does not vanish then[17,18]

$$\mathscr{L}_\xi T_{ab} = 2(\Box\psi + \Lambda\psi) g_{ab} - 2\psi_{;ab}, \tag{4.1}$$

where

$$\Box\psi = g^{ab}\psi_{;ab}. \tag{4.2}$$

Because the energy-momentum tensor occurs in (4.1) the dynamic results depend on the kind of fluid considered. In this section we consider in detail a fluid with anisotropic pressure and vanishing energy flux vector described by energy-momentum tensor (3.1). A magnetic field can be included by replacing μ, $p_{\parallel}$, and $p_{\perp}$ by $\bar{\mu}$, $\bar{p}_{\parallel}$, and $\bar{p}_{\perp}$ defined by (3.9)–(3.11). In Sec. VI we examine the extension of the results to a fluid with a general energy-momentum tensor.

Suppose that T_{ab} is given by (3.1). With the aid of (2.9) for $\mathscr{L}_\xi u_a$ and (2.11) for $\mathscr{L}_\xi n_a$ a direct calculation yields

$$\begin{aligned} \mathscr{L}_\xi T_{ab} = {} & (\mathscr{L}_\xi \mu + 2\psi\mu) u_a u_b + (\mathscr{L}_\xi p_{\parallel} + 2\psi p_{\parallel}) n_a n_b \\ & + (\mathscr{L}_\xi p_{\perp} + 2\psi p_{\perp}) p_{ab} + 2(\mu + p_{\perp}) u_{(a} v_{b)} \\ & + 2(p_{\parallel} - p_{\perp}) n_{(a} m_{b)}, \end{aligned} \tag{4.3}$$

which, when substituted into (4.1), gives

$$\begin{aligned} 2\Box\psi & (p_{ab} - u_a u_b + n_a n_b) - 2\psi_{;ab} \\ & = [\mathscr{L}_\xi \mu + 2\psi(\mu + \Lambda)] u_a u_b \\ & \quad + [\mathscr{L}_\xi p_{\parallel} + 2\psi(p_{\parallel} - \Lambda)] n_a n_b \\ & \quad + [\mathscr{L}_\xi p_{\perp} + 2\psi(p_{\perp} - \Lambda)] p_{ab} \\ & \quad + 2(\mu + p_{\perp}) u_{(a} v_{b)} + 2(p_{\parallel} - p_{\perp}) n_{(a} m_{b)}. \end{aligned} \tag{4.4}$$

By contracting (4.4) in turn with the tensors $u^a u^b$, $u^a n^b$, $u^a p^{bc}$, $n^a n^b$, $n^a p^{bc}$, p^{ab}, and $p^{ac} p^{bd} - \frac{1}{2} p^{ab} p^{cd}$ the following seven equations are derived:

$$u^a u^b: \quad \mathscr{L}_\xi \mu + 2\psi(\mu + \Lambda) = -2(\Box\psi + \psi_{;ab} u^a u^b), \tag{4.5}$$

$$u^a n^b: \quad (\mu + p_{\parallel}) n^b v_b = 2\psi_{;ab} u^a n^b, \tag{4.6}$$

$$u^a p^{bc}: \quad (\mu + p_{\perp}) p^{bc} v_b = 2\psi_{;ab} u^a p^{bc}, \tag{4.7}$$

$$n^a n^b: \quad \mathscr{L}_\xi p_{\parallel} + 2\psi(p_{\parallel} - \Lambda) = 2(\Box\psi - \psi_{;ab} n^a n^b), \tag{4.8}$$

$$n^a p^{bc}: \quad (p_{\perp} - p_{\parallel}) p^{bc} m_b = 2\psi_{;ab} n^a p^{bc}, \tag{4.9}$$

$$p^{ab}: \quad \mathscr{L}_\xi p_\perp + 2\psi(p_\perp - \Lambda) = 2\Box\psi - \psi_{;ab}\, p^{ab}, \tag{4.10}$$

$$p^{ac}p^{bd} - \tfrac{1}{2}p^{ab}p^{cd}: \quad \psi_{;ab}\, p^{ac}p^{bd} - \tfrac{1}{2}(\psi_{;ab}\, p^{ab})p^{cd} = 0. \tag{4.11}$$

Since $n_a u^a = 0$, (2.12) applies and was used in the derivation of (4.6). Equations (4.5)–(4.11) are valid for any CKV ξ^a.

We now restrict consideration to special conformal Killing vectors (SCKV) that satisfy the condition $\psi_{;ab} = 0$. For a SCKV, (4.5)–(4.11) reduce to

$$\mathscr{L}_\xi \mu + 2\psi(\mu + \Lambda) = 0, \tag{4.12}$$

$$\mathscr{L}_\xi p_\parallel + 2\psi(p_\parallel - \Lambda) = 0, \tag{4.13}$$

$$\mathscr{L}_\xi p_\perp + 2\psi(p_\perp - \Lambda) = 0, \tag{4.14}$$

$$(\mu + p_\parallel)n^b v_b = 0, \tag{4.15}$$

$$(\mu + p_\perp)p^{ab} v_b = 0, \tag{4.16}$$

$$(p_\perp - p_\parallel)\, p^{ab} m_b = 0. \tag{4.17}$$

Equations (4.12)–(4.14) were derived by Herrera *et al.*[1] for $\Lambda = 0$.

Consider now (4.15) and (4.16) and assume that

$$\mu + p_\parallel \neq 0 \quad \text{and} \quad \mu + p_\perp \neq 0. \tag{4.18}$$

In the presence of a pure magnetic field the conditions corresponding to (4.18) are, using (3.9)–(3.11),

$$\bar{\mu} + \bar{p}_\parallel = \mu + p_\parallel \neq 0 \quad \text{and} \quad \bar{\mu} + \bar{p}_\perp = \mu + p_\perp + \lambda H^2 \neq 0. \tag{4.19}$$

If conditions (4.18) are satisfied then from (4.15) and (4.16),

$$n^b v_b = 0, \quad p^{ab} v_b = 0. \tag{4.20}$$

Thus v^b can at most be parallel to u^b and since $u^b v_b = 0$ we conclude that $v^b \equiv 0$. Equations (2.8) and (2.9) reduce to

$$\mathscr{L}_\xi u^a = -\psi u^a, \tag{4.21}$$

$$\mathscr{L}_\xi u_a = \psi u_a. \tag{4.22}$$

There remains only (4.17) to consider. If $p_\parallel = p_\perp$ then the anisotropic fluid reduces to a perfect fluid and the energy-momentum tensor does not depend on n^a. We therefore suppose that

$$p_\perp - p_\parallel \neq 0. \tag{4.23}$$

If a pure magnetic field is present then the corresponding condition is, from (3.10) and (3.11),

$$\bar{p}_\perp - \bar{p}_\parallel = p_\perp - p_\parallel + \lambda H^2 \neq 0. \tag{4.24}$$

Now, when (4.23) is satisfied, (4.17) gives

$$p^{ab} m_b = 0. \tag{4.25}$$

We also have $n^b m_b = 0$ and further from (2.12)

$$u^b m_b = -n^b v_b. \tag{4.26}$$

But if $\mu + p_\parallel \neq 0$ it follows from (4.15) that $n^b v_b = 0$ and therefore $u^b m_b = 0$ by (4.26). Thus since $p^{ab} m_b = 0$, $n^b m_b = 0$, and $u^b m_b = 0$ we conclude that $m_b \equiv 0$ and (2.10) and (2.11) reduce to

$$\mathscr{L}_\xi n^a = -\psi n^a, \tag{4.27}$$

$$\mathscr{L}_\xi n_a = \psi n_a. \tag{4.28}$$

Herrera *et al.*[1] used (4.22) and (4.28) to evaluate $\mathscr{L}_\xi T_{ab}$ and then derived the identities (4.12)–(4.14) (with $\Lambda = 0$). We have shown that (4.12)–(4.14) do not require (4.22) and (4.28) for their derivation. Equations (4.22) and (4.28) cannot be obtained in general from kinematic considerations alone. We have shown here that they are a consequence of Einstein's field equations for the particular case when ξ^a is a SCKV and the energy-momentum tensor has the form (3.1). The derivation of (4.22) required also the reasonable assumptions contained in (4.18), and (4.28) required $p_\parallel \neq p_\perp$ and $\mu + p_\parallel \neq 0$.

The foregoing results apply to any SCKV ξ^a. Consider now a SCKV ξ^a orthogonal to u^a. Then from (2.20b) the kinematic result $v^a = 2\omega^{ab}\xi_b$ holds and since $v^a = 0$ in a fluid with energy-momentum tensor (3.1) it follows that ξ^a must be parallel to ω^a if $\omega \neq 0$: any SCKV orthogonal to u^a admitted by a rotational fluid space-time with energy-momentum tensor (3.1) must be parallel to ω^a. The integral curves of ξ^a are vortex lines and since (4.21) is satisfied the vortex lines are material lines in the fluid.

V. EQUATIONS OF STATE

In a fluid with anisotropic pressure (and vanishing energy flux vector) the special conformal Killing vector property implies through the Einstein field equations relations between $p_\parallel$, $p_\perp$, and μ. These have been investigated by Herrera *et al.*[1] for a SCKV parallel to n^a and orthogonal to both n^a and u^a. We generalize their results to include the cosmological constant Λ and a nonzero magnetic field. We also consider the case of a SCKV parallel to u^a, which was not discussed by Herrera *et al.*[1]

If ξ^a is a CKV satisfying (1.1) then

$$(R^{ab}\xi_b)_{;a} = -3\Box\psi. \tag{5.1}$$

This result, which is purely kinematic, may be derived by first showing with the aid of (1.1) and the identity

$$R^{ab}{}_{;a} = \tfrac{1}{2} R_{,a}\, g^{ab} \tag{5.2}$$

that

$$(R^{ab}\xi_b)_{;a} = \tfrac{1}{2}(\mathscr{L}_\xi R + 2\psi R). \tag{5.3}$$

Equations (5.1) follow on noting that[18]

$$\mathscr{L}_\xi R = -2\psi R - 6\Box\psi. \tag{5.4}$$

Dynamics is introduced through Einstein's field equations[6]

$$R^{ab} = T^{ab} - \tfrac{1}{2} T g^{ab} + \Lambda g^{ab}, \tag{5.5}$$

which, when substituted into (5.1), give

$$[(T^{ab} - \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b]_{;a} = -3\Box\psi. \tag{5.6}$$

For a SCKV, $\Box\psi = 0$, and (5.6) reduces to

$$[(T^{ab} - \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b]_{;a} = 0. \tag{5.7}$$

The conservation law (5.7) is the basis of the subsequent analysis. For a fluid with energy-momentum tensor (3.8),

$$\begin{aligned}(T^{ab} &- \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b \\ &= \tfrac{1}{2}(\bar{\mu} + 2\bar{p}_\perp + \bar{p}_\parallel - 2\Lambda)u^a(u^b\xi_b) \\ &\quad + \tfrac{1}{2}(\bar{\mu} - \bar{p}_\parallel + 2\Lambda)p^{ab}\xi_b \\ &\quad + \tfrac{1}{2}(\bar{\mu} - 2\bar{p}_\perp + \bar{p}_\parallel + 2\Lambda)n^a(n^b\xi_b).\end{aligned} \tag{5.8}$$

Before studying a fluid with anisotropic pressure, consider a perfect fluid for which $\bar{p}_\parallel = \bar{p}_\perp = p$. Suppose that ξ^a

is a SCKV and consider first the case $\xi_a u^a = 0$. (We have seen that if $\omega \neq 0$, ξ^a must be parallel to ω^a.) Then from (5.8),

$$(T^{ab} - \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b = \tfrac{1}{2}(\mu - p + 2\Lambda)\xi^a, \quad (5.9)$$

which, when substituted into (5.7), gives

$$\mathscr{L}_\xi \mu - \mathscr{L}_\xi p + (\mu - p + 2\Lambda)\xi^a{}_{;a} = 0. \quad (5.10)$$

But from (1.1),

$$\xi^a{}_{;a} = 4\psi, \quad (5.11)$$

and using also (4.12) and (4.13) [or (4.14)], which hold for a SCKV, (5.10) becomes

$$\psi(\mu - p + 2\Lambda) = 0. \quad (5.12)$$

Thus

$$\text{either } \psi = 0 \quad \text{or} \quad p = \mu + 2\Lambda. \quad (5.13)$$

In a perfect fluid space-time a SCKV orthogonal to u^a is necessarily a KV unless $p = \mu + 2\Lambda$. Second, if ξ^a is parallel to u^a a similar calculation yields

$$\text{either } \psi = 0 \quad \text{or} \quad \mu + 3p - 2\Lambda = 0. \quad (5.14)$$

A SCKV parallel to u^a in a perfect fluid space-time is necessarily a KV unless $\mu + 3p - 2\Lambda = 0$. Equations (5.13) and (5.14) generalize to a SCKV result due to McIntosh[19,20] for a homothetic vector and $\Lambda = 0$. The work of Collins[21] on shearfree perfect fluid solutions of Einstein's equations with vanishing magnetic Weyl tensor includes an example of McIntosh's result for $\xi_a u^a = 0$ adjusted for $\Lambda \neq 0$.

Consider now the extension of these results to a fluid with anisotropic pressure and vanishing energy flux vector. The case $\xi^a u_a = 0$ depends on the relative orientation of ξ^a and n^a. Two special cases can be considered: ξ^a parallel to n^a and ξ^a orthogonal to n^a (and u^a). Both cases were discussed by Herrera *et al.*[1] and in each case, if $\omega \neq 0$, ξ^a must be parallel to ω^a.

First, suppose that ξ^a is parallel to n^a: $\xi^a = \xi n^a$. Then since $u^b n_b = 0$ and $p^{ab} n_b = 0$, (5.8) reduces to

$$(T^{ab} - \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b = \tfrac{1}{2}(\bar{\mu} - 2\bar{p}_\perp + \bar{p}_\parallel + 2\Lambda)\xi^a, \quad (5.15)$$

which, when substituted into (5.7), gives

$$\mathscr{L}_\xi \bar{\mu} - 2\mathscr{L}_\xi \bar{p}_\perp + \mathscr{L}_\xi \bar{p}_\parallel + (\bar{\mu} - 2\bar{p}_\perp + \bar{p}_\parallel + 2\Lambda)\xi^a{}_{;a} = 0. \quad (5.16)$$

Using (4.12)–(4.14), which remain valid if μ, $p_\parallel$, and $p_\perp$ are replaced by $\bar{\mu}$, $\bar{p}_\parallel$, and $\bar{p}_\perp$, and (5.11), (5.16) simplifies to

$$\psi(\bar{\mu} - 2\bar{p}_\perp + \bar{p}_\parallel + 2\Lambda) = 0, \quad (5.17)$$

and hence with the aid of (3.9)–(3.11) we find that

$$\text{either } \psi = 0 \quad \text{or} \quad 2p_\perp - p_\parallel = \mu - \lambda H^2 + 2\Lambda. \quad (5.18)$$

Second, suppose that ξ^a is orthogonal to n^a and u^a. Then $p^{ab}\xi_b = \xi^a$ and (5.8) becomes

$$(T^{ab} - \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b = \tfrac{1}{2}(\bar{\mu} - \bar{p}_\parallel + 2\Lambda)\xi^a. \quad (5.19)$$

On substituting (5.19) into (5.7) and proceeding as before we obtain

$$\psi(\bar{\mu} - \bar{p}_\parallel + 2\Lambda) = 0, \quad (5.20)$$

and hence

$$\text{either } \psi = 0 \quad \text{or} \quad p_\parallel = \mu + \lambda H^2 + 2\Lambda. \quad (5.21)$$

No condition is placed on $p_\perp$ by a SCKV orthogonal to n^a. When $H = 0 = \Lambda$, (5.18) and (5.21) reduce to the results of Herrera *et al.*[1] and, further, when $p_\parallel = p_\perp = p$, (5.13) is regained.

Finally, suppose that ξ^a is parallel to u^a. Equation (5.8) reduces to

$$(T^{ab} - \tfrac{1}{2}(T - 2\Lambda)g^{ab})\xi_b = -\tfrac{1}{2}(\bar{\mu} + 2\bar{p}_\perp + \bar{p}_\parallel - 2\Lambda)\xi^a, \quad (5.22)$$

and proceeding as previously we find that

$$\psi(\bar{\mu} + 2\bar{p}_\perp + \bar{p}_\parallel - 2\Lambda) = 0; \quad (5.23)$$

thus

$$\text{either } \psi = 0 \quad \text{or} \quad \mu + 2p_\perp + p_\parallel + \lambda H^2 - 2\Lambda = 0. \quad (5.24)$$

When $\Lambda = 0$, the second alternative in (5.24) is in general nonphysical. Equation (5.24) reduces to (5.14) when $H = 0$ and $p_\parallel = p_\perp = p$.

VI. GENERAL ENERGY-MOMENTUM TENSOR

The dynamic results of Secs. IV and V are based on the energy-momentum tensor (3.1). There are two directions in which (3.1) can be generalized: first by including a nonzero energy flux vector q^a and second by considering a general stress tensor π^{ab} instead of the special form (3.14).

We first show by means of a counterexample that when $q^a \neq 0$ the relation $\mathscr{L}_\xi u^a = -\psi u^a$ need no longer hold for a SCKV ξ^a. Consider the radiationlike viscous fluid solution in $k = 0$ FRW space-time found by Coley and Tupper.[22] The metric is given by (1.3) expressed in spherical polar coordinates (r,θ,ϕ) and with $R(\eta) = \tfrac{1}{2}\eta$. The fluid four-velocity vector is

$$\mathbf{u} = (2/\eta)(\cosh\Phi\,\partial_\eta + \sinh\Phi\,\partial_r), \quad (6.1)$$

where $\Phi(\eta,r)$ is the hyperbolic tilt angle. Consider the translation KV

$$\boldsymbol{\xi} = \partial_x = \sin\theta\cos\phi\,\partial_r + (1/r)(\cos\theta\cos\phi\,\partial_\theta - \csc\theta\sin\phi\,\partial_\phi), \quad (6.2)$$

which is a SCKV with $\psi = 0$. Clearly,

$$\mathscr{L}_\xi \mathbf{u} = [\boldsymbol{\xi},\mathbf{u}] \neq -\psi\mathbf{u} = \mathbf{0}. \quad (6.3)$$

In this counterexample, the stress tensor π^{ab} is of the form (3.14). For,

$$\pi^{ab} = -\lambda\sigma^{ab}, \quad (6.4)$$

where λ is the coefficient of shear viscosity and σ^{ab} is the fluid shear tensor, and since u^a is isotropic (invariant under the rotational Killing vectors), σ^{ab} must be of the form

$$\sigma^{ab} = \sqrt{3}\,\sigma(\eta,r)(\tfrac{1}{3}h^{ab} - n^a n^b), \quad (6.5)$$

where $\mathbf{n} = (2/\eta)\partial_r$ is the unit radial vector. This counterexample demonstrates the difficulty of extending results of Secs. IV and V to fluids with $q^a \neq 0$ even if the stress tensor retains the simple form (3.14).

Consider now an anisotropic fluid with arbitrary stress tensor π^{ab} ($\pi^{ab}u_b = 0$, $\pi^a{}_a = 0$) but with $q^a = 0$. Suppose that ξ^a is a SCKV. If (3.12) with $q^a = 0$ is substituted into

(4.1) with $\psi_{;ab}=0$ and (4.1) is contracted in turn with the tensors $u^a u^b$, h^{ab}, $u^a h^b_{\ c}$, and $h^a_{\ c} h^b_{\ d} - \frac{1}{3} h^{ab} h_{cd}$, then the following four equations are obtained:

$$u^a u^b: \quad \mathscr{L}_\xi \mu + 2\psi(\mu+\Lambda)=0, \tag{6.6}$$

$$h^{ab}: \quad \mathscr{L}_\xi p + 2\psi(p-\Lambda)=0, \tag{6.7}$$

$$u^a h^b_{\ c}: \quad \pi_{cb} v^b = -(\mu+p) v_c, \tag{6.8}$$

$$h^a_{\ c} h^b_{\ d} - \tfrac{1}{3} h^{ab} h_{cd}: \quad \mathscr{L}_\xi \pi_{cd} = 2u_{(c}\pi_{d)b} v^b. \tag{6.9}$$

Equation (4.12) therefore remains valid for general π^{ab} if $q^a=0$, and (4.13) and (4.14) are replaced by (6.7). Using (6.8), (6.9) becomes

$$\mathscr{L}_\xi \pi_{cd} = -2(\mu+p) u_{(c} v_{d)}. \tag{6.10}$$

Since $v_a u^a=0$, it follows from (6.10) that if $\mu+p\neq 0$ and $q^a=0$ then for a SCKV ξ^a,

$$\mathscr{L}_\xi u^a = -\psi u^a \Leftrightarrow \mathscr{L}_\xi \pi_{ab}=0. \tag{6.11}$$

We now give two examples of fluids in which v^a must vanish. First, consider fluids in which the three principal stresses are compressive. This condition will be satisfied in many astrophysical and cosmological situations since we do not in general expect to find a fluid under tension. We have, using (6.8) and $q^a=0$,

$$T_{ab} v^b = -\mu v_a. \tag{6.12}$$

Thus v^a is a spacelike ($v_a u^a=0$) eigenvector of T_{ab} with eigenvalue $-\mu$ that is *negative*. But if the three principal stresses are compressive, the eigenvalues of T_{ab} corresponding to spacelike eigenvectors must be positive,[23,24] and therefore the only permissible solution to (6.12) is $v^a=0$. Although, for a fluid in equilibrium, thermodynamic considerations indicate that the pressure must be positive, negative pressure corresponding to a state of tension in the fluid can exist in nonequilibrium metastable states; spontaneous contraction of the fluid can result in the formation of cavities in the fluid and this can lead to possible negative pressures. In the neighborhood of a critical point, for example, a superheated liquid may have a negative pressure.[25]

Second, consider a relativistic gas. The energy-momentum tensor describing the matter content of a relativistic gas is[26,27]

$$T^{ab} = \sum_A \int_{P_A} p^a p^b f_A \pi_A, \tag{6.13}$$

where summation is over all the species of the gas, p^a is the future-directed four-momentum of a particle of the gas, f_A is the distribution function, and π_A is the coordinate-independent volume element on the mass shell P_A for particles of species A. From (6.13),

$$T_{ab} v^a v^b = \sum_A \int_{P_A} (p_a v^a)^2 f_A \pi_A \geqslant 0, \tag{6.14}$$

while from (6.12),

$$T_{ab} v^a v^b = -\mu v_a v^a \leqslant 0. \tag{6.15}$$

Since v^a is spacelike, we conclude from (6.14) and (6.15) that $v^a=0$ for a relativistic gas described by (6.13).

Finally we examine the possibility of extending the results of Sec. V on equations of state to an anisotropic fluid with arbitrary π_{ab} but $q^a=0$. When $q^a=0$ we have

$$\begin{aligned}(T^{ab} - \tfrac{1}{2}(T-2\Lambda) g^{ab})\xi_b &= \tfrac{1}{2}(\mu+3p-2\Lambda) u^a (u^b \xi_b) \\ &\quad + \tfrac{1}{2}(\mu - p + 2\Lambda) h^{ab}\xi_b + \pi^{ab}\xi_b.\end{aligned} \tag{6.16}$$

We assume that ξ^a is a SCKV so that (5.7) is satisfied. The analysis is the same as in Sec. V with (6.6) and (6.7) taking the place of (4.12)–(4.14).

Suppose first that $\xi_a u^a=0$. In place of (5.12) for a perfect fluid we find that

$$\psi(\mu - p + \Lambda) + \pi^{ab}_{\ \ ;a}\xi_b = 0. \tag{6.17}$$

But from the momentum conservation equation with $q^a=0$,

$$(\mu+p)\dot u^a = -h^{ac}(p_{,c} + \pi_c^{\ b}{}_{;b}), \tag{6.18}$$

and by contracting (6.18) with ξ_a and using again (6.7) we find that

$$\pi^{ab}_{\ \ ;a}\xi_b = 2\psi(p-\Lambda) - (\mu+p)\dot u_a \xi^a. \tag{6.19}$$

Substituting from (6.19) into (6.17) gives

$$(\mu+p)(\psi - \dot u_a \xi^a) = 0, \tag{6.20}$$

which is identically satisfied for any μ and p by virtue of (2.20a).

Second, suppose that ξ^a is parallel to u^a. Then since $\pi_{ab}u^b=0$, π_{ab} makes no contribution to (6.16) and therefore the perfect fluid result (5.14) is again obtained. The statement (5.14) is therefore valid for a SCKV orthogonal to u^a in an anisotropic fluid with arbitrary π_{ab} provided $q^a=0$. In particular it agrees with (5.24) for a fluid with anisotropic pressure since p is related to $p_\parallel$ and $p_\perp$ through (3.13).

VII. CONCLUDING REMARKS

The following argument may explain why a SCKV ξ^a may be expected to satisfy the relation

$$\mathscr{L}_\xi u^a = -\psi u^a \tag{7.1}$$

in an anisotropic fluid with arbitrary π^{ab} and $q^a=0$. When $\psi_{;ab}=0$, ξ^a satisfies[18]

$$\mathscr{L}_\xi R_{ab} = 0, \tag{7.2}$$

and therefore the Ricci tensor is invariant under the mapping ξ^a. (A SCKV is a particular case of a Ricci collineation vector.) Since the eigendirections are an invariant property of a tensor we may expect that the Ricci eigendirections will be invariant under the mapping ξ^a. Now from (3.2) with $q^a=0$ and Einstein's equations (5.5) we have

$$R_{ab}u^b = -\tfrac{1}{2}(\mu+3p-2\Lambda)u_a, \tag{7.3}$$

and therefore u^a is a timelike eigendirection of R_{ab}. Hence we may expect that ξ^a will map the integral curves of u^a into themselves and therefore that

$$\mathscr{L}_\xi u^a = f u^a \tag{7.4}$$

for some scalar function $f(x^c)$. By contracting (7.4) with u_a and using (1.1) it can be verified that $f=-\psi$. For a fluid with anisotropic pressure and vanishing energy flux described by energy-momentum tensor (3.1), we also have

$$R_{ab}n^b = \tfrac{1}{2}(\mu + p_\parallel - 2p_\perp + 2\Lambda)n_a, \tag{7.5}$$

and therefore n^a is a spacelike eigendirection of R_{ab}, which may explain the relation

$$\mathscr{L}_{\xi} n^a = -\psi n^a . \tag{7.6}$$

As well as considering a general anisotropic fluid with arbitrary stress tensor it was also of value to study the special case of a fluid with anisotropic pressure described by stress tensor (3.14). Relations for the anisotropy vector n^a could be developed and also the results derived depended on weaker conditions. For example, although the assumption of positive eigenvalues corresponding to spacelike eigenvectors of T_{ab} is sufficient to derive (7.1) for arbitrary π^{ab} it is not necessary for the special case of a fluid with anisotropic pressure. The spacelike eigenvectors of the energy-momentum tensor (3.1) are n^a and any vector that lies on the two-plane orthogonal to u^a and n^a. The corresponding eigenvalues are $p_{\parallel}$ and $p_{\perp}$. If a pure magnetic field is included, the eigenvalues are $\bar{p}_{\parallel}$ and $\bar{p}_{\perp}$ and from (3.10) $\bar{p}_{\parallel}$ may be negative if the intensity of the magnetic field is sufficiently strong. To derive (7.1) in a fluid with anisotropic pressure we required only (4.18) [or (4.19) if $H \neq 0$] and not $p_{\parallel} > 0$ and $p_{\perp} > 0$ (or $\bar{p}_{\parallel} > 0$ and $\bar{p}_{\perp} > 0$).

Although the kinematic results derived here for a CKV are valid in general, the dynamic results established for a SCKV depend on the energy-momentum tensor. These dynamic results depend crucially on the vanishing of the energy flux vector.

ACKNOWLEDGMENTS

We thank the referee for valuable comments.

DPM thanks the Foundation for Research Development of the Council for Scientific and Industrial Research, Pretoria, for an Operating Grant.

[1]L. Herrera, J. Jiménez, L. Leal, J. Ponce de León, M. Esculpi, and V. Galina, J. Math. Phys. **25**, 3274 (1984).

[2]D. P. Mason and M. Tsamparlis, J. Math. Phys. **26**, 2881 (1985).

[3]R. Maartens and S. D. Maharaj, Class. Quantum Grav. **3**, 1005 (1986).

[4]M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).

[5]L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 778, 2018, 2847 (1985).

[6]G. F. R. Ellis, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 104–182; in *Cargèse Lectures in Physics*, edited by E.Schatzman (Gordon and Breach, New York, 1973), Vol. 6, pp. 1–60. Latin indices run over the four coordinates of space-time. A semicolon denotes covariant differentiation with respect to the metric tensor g_{ab} of space-time [signature $(-+++)$]. An overhead dot denotes covariant differentiation along a fluid particle world line; for example, $\dot{A}^a = A^a{}_{;b} u^b$. The Riemann curvature tensor is defined through the Ricci identity

$$A_{a;bc} - A_{a;cb} = R_{tabc} A^t$$

as

$$R^a{}_{bcd} = \Gamma^a_{bd,c} - \Gamma^a_{bc,d} + \Gamma^a_{cs}\Gamma^s_{bd} - \Gamma^a_{ds}\Gamma^s_{bc} .$$

The Ricci and Einstein tensors are defined as

$$R_{ab} = R^t{}_{atb} ,$$

$$G_{ab} = R_{ab} - \tfrac{1}{2} R g_{ab} ,$$

where $R = R^a{}_a$ is the Ricci scalar. Units are used in which the speed of light in a vacuum and Einstein's gravitational constant are both unity. Einstein's field equations are

$$G_{ab} + \Lambda g_{ab} = T_{ab} ,$$

where Λ is the cosmological constant.

[7]L. Herrera, G. J. Ruggeri, and L. Witten, Astrophys. J. **234**, 1094 (1979).

[8]P. S. Letelier, Phys. Rev. D **22**, 807 (1980).

[9]M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, J. Math. Phys. **22**, 118 (1981).

[10]M. Cosenza, L. Herrera, M. Esculpi, and L. Witten, Phys. Rev. D **25**, 2527 (1982).

[11]S. S. Bayin, Phys. Rev. D **26**, 1262 (1982).

[12]L. Herrera and J. Ponce de León, J. Math. Phys. **26**, 2302 (1985).

[13]G. F. Chew, M. L. Goldberger, and F. E. Low, Proc. R. Soc. London Ser. A **236**, 112 (1956).

[14]R. D. Tenreiro and R. Hakim, Phys. Rev. D **15**, 1435 (1977).

[15]A. Lichnerowicz, *Relativistic Magnetohydrodynamics* (Benjamin, Amsterdam, 1967), Chap. 4.

[16]R. Maartens and S. D. Maharaj, J. Math. Phys. **26**, 2869 (1985).

[17]K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1955), pp. 17 and 52.

[18]If η^a is any vector then (see Ref. 6 for the definition of $R^a{}_{bcd}$ and Ref. 17)

$$\mathscr{L}_\eta R^a{}_{bcd} = (\mathscr{L}_\eta \Gamma^a_{bd})_{;c} - (\mathscr{L}_\eta \Gamma^a_{bc})_{;d} ,$$

$$\mathscr{L}_\eta R_{ab} = (\mathscr{L}_\eta \Gamma^s_{ab})_{;s} - (\mathscr{L}_\eta \Gamma^s_{as})_{;b} ,$$

$$\mathscr{L}_\eta \Gamma^a_{bc} = \tfrac{1}{2} g^{at} [(\mathscr{L}_\eta g_{bt})_{;c} + (\mathscr{L}_\eta g_{ct})_{;b} - (\mathscr{L}_\eta g_{bc})_{;t}] .$$

Hence if ξ^a is a CKV satisfying (1.1) then

$$\mathscr{L}_\xi R_{ab} = -2\psi_{;ab} - g_{ab}\Box\psi ,$$

$$\mathscr{L}_\xi R = -2\psi R - 6\Box\psi ,$$

$$\mathscr{L}_\xi G_{ab} = 2g_{ab}\Box\psi - 2\psi_{;ab} ,$$

where $\Box\psi = g^{ab}\psi_{;ab}$. Since Einstein's field equations are (see Ref. 6)

$$G_{ab} + \Lambda g_{ab} = T_{ab} ,$$

we have

$$\mathscr{L}_\xi T_{ab} = 2(\Box\psi + \Lambda\psi) g_{ab} - 2\psi_{;ab} .$$

This result does not agree exactly with the result of Herrera *et al.*[1] on replacing ψ by $\psi/2$. The factor $\frac{1}{2}$ multiplying $g_{\mu\nu}\Box\psi$ in their Eq. (12) appears to be an arithmetic error.

[19]C. B. G. McIntosh, Phys. Lett. A **50**, 429 (1975).

[20]C. B. G. McIntosh, Gen. Relativ. Gravit. **7**, 199 (1976).

[21]C. B. Collins, J. Math. Phys. **25**, 995 (1984).

[22]A. A. Coley and B. O. J. Tupper, Astrophys. J. **288**, 418 (1985).

[23]J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1971), p. 186.

[24]The stress in the fluid is described by

$$S_{ab} = p h_{ab} + \pi_{ab} .$$

Suppose that $v^a \neq 0$ and define $\bar{v}^a = v^a/v$, where $v = (v_a v^a)^{1/2}$. Then since $\pi_{ab} v^a = -(\mu + p) v_a$ we have

$$S_{ab}\bar{v}^b = -\mu \bar{v}_a ,$$

$$S_{ab}\bar{v}^a\bar{v}^b = -\mu .$$

Thus $\bar{v}^a$ is an eigendirection of S_{ab} and $S_{ab}\bar{v}^a\bar{v}^b$ is *negative*. But $-S_{ab}\bar{v}^a\bar{v}^b$ is the component of the stress vector in the direction $+\bar{v}^a$ acting on the two-plane orthogonal to u^a and $\bar{v}^a$, which is a principal plane of stress. Hence $-S_{ab}\bar{v}^a\bar{v}^b$ is a principal stress, and if the principal stresses in the fluid are compressive, then $S_{ab}\bar{v}^a\bar{v}^b$ must be *positive*, which is a contradiction. We therefore conclude that v^a cannot be nonzero.

[25]L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon, Oxford, 1968), Secs. 12 and 83.

[26]J. Ehlers, in *General Relativity and Gravitation*, edited by R. K. Sachs (Academic, New York, 1971), pp. 1–70.

[27]J. M. Stewart, "Non-equilibrium relativistic kinetic theory," *Lecture Notes in Physics*, Vol. 10 (Springer, Berlin, 1971), Chap. 2.

Power law singularities in the scale covariant theory

A. Beesham
Department of Mathematics and Applied Mathematics, University of Durban-Westville, Private Bag X54001, Durban 4000, South Africa and Department of Applied Mathematics, University of Capetown, Rondebosch 7700, South Africa

(Received 15 April 1986; accepted for publication 28 July 1986)

Power asymptote singularities are discussed in the scale covariant theory of gravitation. Some general results are derived. Special attention is paid to the Friedmann [A. Friedmann, Z. Phys. **10**, 377 (1922)] and Kasner [E. Kasner, Am. J. Math. **43**, 217 (1921)] models. A wider class of behavior is exhibited and it is shown that the results obtained constitute a generalization of the corresponding general relativistic results.

I. INTRODUCTION

All the orthogonal spatially homogeneous cosmologies in general relativity with zero cosmological constant, perfect fluid source, and the usual equation of state, which are expanding at some instant, originate at a singularity at which the energy density of the fluid diverges.[1] These cosmological solutions have been classified on the basis of which power asymptote they admit at the singularity.[2] An interesting question that arises is how this classification changes in alternative theories of gravity. In this article we investigate power asymptote singularities in homogeneous cosmologies in the scale covariant theory[3] and show that more diverse behavior is possible as compared to general relativity. The scale covariant theory is a viable alternative theory to general relativity, which incorporates a variable gravitational "constant."[4,5]

II. POWER ASYMPTOTES IN GENERAL RELATIVITY

In this section we review briefly the concepts needed to describe the power asymptotes in general relativity. Further details can be found in the beautiful paper by Wainwright.[2] In its eigenframe, the expansion tensor θ_{ab} has the form

$$\theta_{ab} = \mathrm{diag}(0,\theta_1,\theta_2,\theta_3),$$

where the θ_α, $\alpha = 1,2,3$, are the eigenvalues associated with the spacelike eigenvectors. Length scales l_α are defined (up to constant scale factors) by

$$\dot{l}_\alpha / l_\alpha = \theta_\alpha,$$

where the overhead dot denotes differentiation with respect to t, t being the clock time along the fluid flow lines. As the singularity at $t=0$ is approached, the diagonal tetrad components θ_α are of the form

$$\theta_\alpha = (P_\alpha/t)[1+O(t^r)],\ \alpha = 1,2,3, \tag{2.1}$$

where the p_α are constants that characterize the power asymptote. The $O(t^r)$ denotes higher-order terms that tend to zero as a power of t.

In addition to the p_α, the following finite limits are defined:

$$\beta_m = \lim_{t\to 0^+} \frac{3\mu}{\theta^2},\ \beta_s = \lim_{t\to 0^+}\frac{3\sigma^2}{\theta^2},\ \beta_c = \lim_{t\to 0^+}\frac{-3R^*}{2\theta^2}, \tag{2.2}$$

where μ, θ, and σ are the matter density, expansion, and shear of the fluid, and R^* is the curvature of the hypersurfaces orthogonal to the fluid flow. Depending upon whether $\beta_m \neq 0$, $\beta_s \neq 0$, or $\beta_c \neq 0$, the matter, shear, or curvature, respectively, are dynamically significant near the singularity. In view of the equation[4]

$$\tfrac{1}{3}\theta^2 = \sigma^2 + \mu - \tfrac{1}{2}R^*, \tag{2.3}$$

the constants in (2.2) satisfy the relation

$$\beta_m + \beta_s + \beta_c = 1. \tag{2.4}$$

Each of the constants β_m, β_s, and β_c is non-negative. If $\beta_m = 1$, then the matter is dynamically important near the singularity, $\beta_s = 1$ means a shear-dominated singularity and $\beta_c = 1$ a curvature-dominated singularity. The following relations hold:

$$\beta_c = 0 \ \wedge\ \beta_m \neq 0 \Rightarrow \beta_s = 0, \tag{2.5}$$

$$\beta_s = 0 \Rightarrow \beta_c = 0. \tag{2.6}$$

The power asymptotes are classified according to the dynamical significance of the matter, shear, and spatial curvature. In addition various subcases are distinguished according to the values of the p_α. In the case of the Friedmann[6] cosmologies, we have

$$\beta_m = 1,\ \beta_s = 0,\ \beta_c = 0, \tag{2.7}$$

$$p_1 = p_2 = p_3 = 2/(3\gamma), \tag{2.8}$$

$$2/\gamma = p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 4/(3\gamma^2), \tag{2.9}$$

whereas in the case of the Kasner[7] model we have

$$\beta_m = 0,\ \beta_s = 1,\ \beta_c = 0, \tag{2.10}$$

$$p_1 = \tfrac{1}{3} + 2(\sin\alpha)/3,\ p_2 = \tfrac{1}{3} + 2(\sin(\alpha + 2\pi/3))/3,$$

$$p_3 = \tfrac{1}{3} + 2(\sin(\alpha+4\pi/3))/3,\ -\pi/6 < \alpha \leqslant \pi/2, \tag{2.11}$$

$$2/\gamma > p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1. \tag{2.12}$$

III. POWER ASYMPTOTES IN THE SCALE COVARIANT THEORY

According to the scale covariant theory, Einstein's field equations are valid in gravitational units, which describe macroscopic phenomena, whereas physical quantities are measured in atomic units, which relate to microscopic phenomena. The metric tensors in the two units are related by a conformal transformation

$$g_{ab} = \phi^2(t_A)(g_A)_{ab},\ 0 < \phi < \infty, \tag{3.1}$$

where the subscript A denotes atomic units. We use the sym-

 0022-2488/86/122995-03$02.50

bol ϕ instead of the more usual β, as we reserve β for the constants appearing in Sec. II. The scale function ϕ is considered here to be a function only of time since we are concerned only with homogeneous cosmologies. Tensors and equations in the scale covariant theory are obtained from the corresponding general relativistic ones by making the conformal transformation (3.1). Covariant differentiation has to be replaced by co-covariant differentiation.[3,8]

Hence we have[3,9,10]

$$\theta = \phi^{-1}(\theta_A + 3\dot{\phi}/\phi), \tag{3.2}$$

$$G\mu = \phi^{-2}G_A\mu_A, \tag{3.3}$$

$$\sigma = \phi^{-1}\sigma_A, \tag{3.4}$$

$$t = \int_0^{t_a} \phi(t_A)dt_A. \tag{3.5}$$

In the scale covariant theory, allowance is made for a possible variation of the gravitational constant G_A.

Since

$$(\dot{l}_A)_\alpha/(l_A)_\alpha = (\theta_A)_\alpha, \tag{3.6}$$

the $(p_A)_\alpha$ are obtained from the equation

$$(\theta_A)_\alpha = [(p_A)_\alpha/t_A][1 + O(t_A^r)], \quad \alpha = 1,2,3. \tag{3.7}$$

Motivated by the corresponding general relativistic definitions in Sec. II, we define the following limits:

$$(\beta_A)_m = \lim_{t_A \to 0^+} \frac{3\mu_A}{\theta_A^2}, \quad (\beta_A)_s = \lim_{t_A \to 0^+} \frac{3\sigma_A^2}{\theta_A^2},$$
$$(\beta_A)_c = \lim_{t_A \to 0^+} \frac{-3R_A^*}{2\theta_A^2}. \tag{3.8}$$

The functional forms that have been considered for ϕ as being consistent with observations are[10]

$$\phi(t_A) = (t_0/t_A)^\epsilon, \quad \epsilon = \pm 1, \pm \tfrac{1}{2}, \tag{3.9}$$

where t_0, ϵ are constants. We shall consider $-1 \leqslant \epsilon < 1$, which includes the above cases except for $\epsilon = 1$. This case can be treated separately since it leads to logarithmic functions [see Eq. (3.5)], but we shall not pursue this case further in this article.

Furthermore, in the scale covariant theory it is assumed that[11,12] the gravitational constant varies as the universe of time, viz.,

$$G_A = B_0/t_A, \quad B_0 = \text{const.} \tag{3.10}$$

From Eq. (2.3) we find that

$$R^* = 2\sigma^2 + 2\mu - \tfrac{2}{3}\theta^2,$$

which together with (3.2)–(3.4) lead to

$$R^* = \phi^{-2}R_A^*. \tag{3.11}$$

From (2.2), (3.8), (3.9), and (3.11) we infer the following.

(i) $\beta_s = 0 \Rightarrow (\beta_A)_s = 0$. So a velocity-dominated singularity[13,14] in gravitational units implies a velocity-dominated singularity in atomic units. Furthermore, $\beta_s > 0 \Rightarrow (\beta_A)_s > 0$.

(ii) $\beta_c = 0 \Rightarrow (\beta_A)_c = 0$ and $\beta_c > 0 \Rightarrow (\beta_A)_c > 0$.

(iii) $(\beta_A)_s = 0 \Rightarrow (\beta_A)_c = 0$.

(iv) $\beta_m = 0 \Rightarrow (\beta_A)_m = 0$. If matter is dynamically negligible in gravitational units, it is also dynamically negligible in atomic units.

(v) Flat three-spaces in gravitational units imply flat three-spaces in atomic units.

(vi) The equation

$$(\beta_A)_m + (\beta_A)_s = (\beta_A)_c = 1$$

does not hold in atomic units except for the case $\phi = 1$, i.e., gravitational units.

IV. FRIEDMANN COSMOLOGIES IN THE SCALE COVARIANT THEORY

The Friedmann-type equation for the Robertson–Walker metric in the scale covariant theory is[3]

$$3(\dot{l}_A/l_A + \dot{\phi}/\phi)^2 + k/l_A^2 - G_A\mu_A - \Lambda_A = 0. \tag{4.1}$$

Since[3,10]

$$l(t) = \phi(t_A)l_A(t_A),$$
$$dt = \phi(t_A)dt_A, \quad \Lambda = \phi^{-2}(t_A)\Lambda_A(t_A),$$

and using (3.3) we find that Eq. (4) becomes the familiar general relativistic Friedmann equation

$$3\dot{l}^2 - G\mu l^2 - \Lambda l^2 = -3k.$$

The asymptotic behavior of this equation for small l or equivalently for small t is well known[15]:

$$l = C_0 t^{2/(3\gamma)}, \quad C_0 = \text{const.} \tag{4.2}$$

Now small atomic times correspond to small gravitational times, and we thus find by transforming back to atomic units that the asymptotic behavior for small times in the scale covariant theory is given by

$$l_A = D_0 t_A^{\epsilon + 2/(3\gamma) - 2\epsilon/(3\gamma)}, \quad D_0 = \text{const.} \tag{4.3}$$

Although l_A does not always tend to zero as $t_A \to 0$, it is easy to verify from the integral[10] of the conservation equation

$$\mu_A \sim 1/(l_A^{3\gamma-2}G), \tag{4.4}$$

and Eqs. (3.10) and (4.3) that, for the range of values of ϵ considered, we always have a singularity as $t_A \to 0$, i.e., the energy density diverges.

We may now state our results for Friedmann cosmologies in the scale covariant theory:

$$(\beta_A)_m = \begin{cases} 1, & \text{if } \phi = \text{const and } G_A = \text{const}, \\ 0, & \text{otherwise;} \end{cases}$$

$$(\beta_A)_s = (\beta_A)_c = 0;$$

$$3\epsilon + 2(1-\epsilon)/\gamma$$
$$= (p_A)_1 + (p_A)_2 + (p_A)_3 > (p_A)_1^2 + (p_A)_2^2 + (p_A)_3^2$$
$$= 3\epsilon^2 + 4\epsilon(1-\epsilon)/\gamma + 4(1-\epsilon)^2/(3\gamma^2).$$

The last relation holds for $\gamma = 1$ and $\gamma = \frac{4}{3}$ (the latter except when $\epsilon = -1$, in which case the inequality becomes an equality).

V. KASNER MODEL IN THE SCALE COVARIANT THEORY

It is well known that in general relativity the shear varies as

$$\sigma^2 \sim t^{-2}. \tag{5.1}$$

In atomic units, the shear varies as[16]

$$\sigma_A^2 = \phi^2(t_A)\sigma^2. \tag{5.2}$$

From relations (5.1), (5.2), (3.5), and (3.9), we find that the shear in atomic units varies as

$$\sigma_A^2 \sim t_A^{-2}. \tag{5.3}$$

Further, the length scale in the Kasner model is given by[16]

$$l_A \sim t^{1/3 + 2\epsilon/3}, \tag{5.4}$$

from which we can calculate θ_A. Thus

$$(\beta_A)_s = \lim_{t\to 0^+} (3\sigma^2/\theta^2) = 1, \tag{5.5}$$

which is the same as the general relativistic result. From the results of Sec. III, it follows that $(\beta_A)_c = 0$ and, since the Kasner model is empty, $(\beta_A)_m = 0$ since $\mu_A = 0$.

From the relation (5.4) we find that

$$(p_A)_1 = \Lambda_1 + (1-\Lambda_1)\epsilon, \quad 3\Lambda_1 = 1 + 2\sin\alpha,$$

$$(p_A)_2 = \Lambda_2 + (1-\Lambda_2)\epsilon, \quad 3\Lambda_2 = 1 + 2\sin(\alpha + 2\pi/3),$$

$$(p_A)_3 = \Lambda_3 + (1-\Lambda_3)\epsilon, \quad 3\Lambda_3 = 1 + 2\sin(\alpha + 4\pi/3).$$

Thus

$$(p_A)_1 + (p_A)_2 + (p_A)_3 = 1 + 2\epsilon, \tag{5.6}$$

and

$$(p_A)_1^2 + (p_A)_2^2 + (p_A)_3^2 = 1 + 2\epsilon^2. \tag{5.7}$$

From Eqs. (5.6) and (5.7), we deduce that for $0<\epsilon<1$ we have

$$(p_A)_1 + (p_A)_2 + (p_A)_3 > (p_A)_1^2 + (p_A)_2^2 + (p_A)_3^2,$$

whereas for $-1 \leqslant \epsilon < 0$, we have

$$(p_A)_1 + (p_A)_2 + (p_A)_3 < (p_A)_1^2 + (p_A)_2^2 + (p_A)_3^2.$$

The equality only occurs for $\epsilon = 0$, i.e., in gravitational units.

Depending upon the value of ϵ, it is possible to have

$$2/\gamma \lesseqgtr (p_A)_1 + (p_A)_2 + (p_A)_3.$$

For $\epsilon \neq 0$, we have

$$(p_A)_1^2 + (p_A)_2^2 + (p_A)_3^2 > 1.$$

VI. CONCLUSION

The two models that we have discussed serve to illustrate the kind of diverse behavior that occurs in the scale covariant theory as compared to general relativity. We notice that for $\epsilon = 0$ (ϕ = const) and G_A = const, our results reduce to the general relativistic results given earlier. Our results thus represent a generalization of the corresponding general relativistic results. We note further that in atomic units it is possible to have $(\beta_A)_m = 0$ even in nonempty Friedmann models. This is not possible in nonempty Friedmann models in gravitational units.

Finally we remark that the value $\epsilon = -\frac{1}{2}$ seems most likely to fit observational results.[17] From our general results, it is easy to work out the power asymptotes for this specific value of ϵ.

[1]C. B. Collins and G. F. R. Ellis, Phys. Rep. **56**, 65 (1979).

[2]J. Wainwright, Gen. Relativ. Gravit. **16**, 657 (1984).

[3]V. Canuto, P. J. Adams, S.-H. Hsieh, and E. Tsiang, Phys. Rev. D **16**, 1643 (1977).

[4]P. S. Wesson, *Gravity, Particles and Astrophysics* (Reidel, Dordrecht, 1980).

[5]C. W. Will, Phys. Rep. **113**, 345 (1984).

[6]A. Friedmann, Z. Phys. **10**, 377 (1922).

[7]E. Kasner, Am. J. Math. **43**, 217 (1921).

[8]P. A. M. Dirac, Proc. R. Soc. London A **333**, 403 (1973).

[9]H. Rahman and S. Banerji, Astrophys. Space Sci. **113**, 405 (1985).

[10]V. M. Canuto, S.-H. Hsieh, and J. R. Owen, Astrophys. J. Suppl. **41**, 263 (1979).

[11]V. Canuto, "Scale covariant cosmology and astrophysical applications," in *Proceedings of the Second Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1982).

[12]V. M. Canuto and I. Goldman, "The strong equivalence principle and its violation," in *Early Evolution of the Universe and Its Present Structure*, edited by G. O. Abell and G. Chincarini (Reidel, Dordrecht, 1983).

[13]D. Eardley, E. Liang, and R. Sachs, J. Math. Phys. **13**, 99 (1972).

[14]E. P. T. Liang, J. Math. Phys. **13**, 386 (1972).

[15]G. F. R. Ellis, "Relativistic cosmology," in *General Relativity and Cosmology*, edited by R. K. Sachs (Academic, New York, 1971).

[16]A. Beesham, Class. Quantum Grav. **3**, 481 (1986).

[17]V. M. Canuto and I. Goldman, Nature **304**, 311 (1983).

The unboundedness of the gravitational partition function

B. Broda
Institute of Physics, Nowotki 149/153, 90-236 Łódź, Poland

(Received 19 January 1986; accepted for publication 16 April 1986)

The unboundedness of the gravitational partition function Z is formally established. First the Euclidean path integral representation for Z, in terms of the renormalized effective action $S^{\text{eff}}_{\text{ren}}$, is derived. Next it is shown that for "strong" fields, $S^{\text{eff}}_{\text{ren}}$ is unbounded from below. The possible influence of the space-time topology is taken into account.

I. INTRODUCTION

One of the most fundamental properties of a physical system is its stability. In the classical case it usually means the boundedness of the energy functional from below. Instead in the quantum case it should be understood as the boundedness of a spectrum of the energy operator H (see Ref. 1). Classical stability was established for gravity (pure and with the matter fields) in the asymptotically flat case[2] and in some other cases.[3] Matter fields were assumed to satisfy the dominant energy condition.[4] It is interesting and important to examine the problem of to what extent the known classical results may be valid in the quantum case. However, we are not able to find spectral properties of the energy operator for gravity directly, instead we have at our disposal some other objects as the partition function Z and the effective action S^{eff} containing the information needed.

This paper is the first one of a series devoted to examining the quantum stability of gravity in different approaches. We start our analysis from the investigation of the partition function Z in the simplest pure gravitational case. We will impose the "volume cutoff" and the boundary conditions by restricting ourselves to closed manifolds, i.e., manifolds without boundaries and compact. The theory will be taken in the Euclidean version.

First we will write down the classical gravitational action S_g with the cosmological term Λ and perform the standard Faddeev–Popov procedure in order to obtain the effective action S^{eff}. The partition function will gain the following path integral representation:

$$Z = \int e^{-S^{\text{eff}}}\, d\mu, \tag{1.1}$$

where $d\mu$ is a properly defined measure. Unfortunately the measure as well as some formal manipulations we will perform do not have a rigorous mathematical meaning; this situation is rather common in gravitation.

Next we will introduce a definition of the analytically renormalized determinant to assure the existence of the effective action S^{eff}. Following the method of Ref. 5 we will rescale the fields to show the unboundedness of the partition function Z (1.1) in the "strong" fields limit. Thus the unboundedness of the partition function is a result of scaling properties of the renormalized effective action $S^{\text{eff}}_{\text{ren}}$.

II. THE PARTITION FUNCTION

The action in general relativity is usually written in the form

$$S_g = -(2\kappa)^{-1} \int (R - 2\Lambda) g^{1/2}\, d^4x, \tag{2.1}$$

where R is the curvature scalar, Λ is the cosmological constant, g is the determinant of the metric $g_{\mu\nu} = \mathrm{diag}(+\,+\,+\,+)$, and κ is the Einstein gravitational constant. We will consider only the so-called closed manifolds, i.e., compact without boundary term.[5] Let us notice also that the cosmological term Λ damps contributions to the path integral coming from large volumes.

We will quantize the theory by the covariant Faddeev–Popov path-integral method.[6] We will treat the gravitational field as a gauge theory with the diffeomorphism group ($\mathrm{Diff}\, M^4$) acting as the gauge group. The diffeomorphism is generated by the displacement of space-time points through a contravariant vector field $\xi^\mu(x)$, i.e., $x^\mu \to x^\mu + \xi^\mu(x)$. Consequently, the effect of this transformation on $g^{\mu\nu}(x)$ takes the form

$$\delta g^{\mu\nu} = -L_\xi\, g^{\mu\nu} = -\xi^\lambda \partial_\lambda g^{\mu\nu} + g^{\mu\lambda} \partial_\lambda \xi^\nu + g^{\lambda\nu} \partial_\lambda \xi^\mu, \tag{2.2}$$

where L is the Lie derivative.

According to the standard procedure we have to fix a gauge and find the proper Faddeev–Popov term. We choose the so-called degenerate harmonic gauge

$$\partial_\nu (g^{1/2} g^{\mu\nu}) = 0. \tag{2.3}$$

If we subject the gauge-fixing term (2.3) to the infinitesimal gauge transformation (2.2) we will get the following ghost contribution to the action:

$$S_{\text{FP}} = -4 \log \det(-\Delta), \tag{2.4}$$

where Δ is the Laplace–Beltrami operator, i.e., $\Delta = g^{-1/2} \partial_\nu (g^{1/2} g^{\nu\lambda} \partial_\lambda)$ (for the proof see Appendix A). It is important to stress that we are not working in any background field formalism but in the normal field one. Thus we should not expect to get a truly covariant Faddeev–Popov term. Nevertheless, we have succeeded in obtaining the covariant Laplace–Beltrami operator in the suitable chosen gauge. The situation has no counterpart in the Yang–Mills case. The choice (2.3) is an extremely convenient one because of the "covariant shape" of the Faddeev–Popov operator, viz. Δ (see Ref. 7).

The gauge-fixing term (2.3) is introduced into the theory with the help of some Lagrange multipliers a_μ. We add to the action the following gauge-fixing expression:

$$S_{\text{GF}} = i \int a_\mu\, \partial_\nu (g^{1/2} g^{\mu\nu})\, d^4x. \tag{2.5}$$

Now we are ready to write down the whole effective action S^{eff}. Collecting the formulas (2.1), (2.4), and (2.5) we have

$$S^{\text{eff}} = S_g + S_{\text{FP}} + S_{\text{GF}}$$
$$= -(2\kappa)^{-1}\int (R-2\Lambda)g^{1/2}\,d^4x - 4\log\det(-\Delta)$$
$$+ i\int a_\mu\,\partial_\nu(g^{1/2}g^{\mu\nu})d^4x. \tag{2.6}$$

We are left only with the problem of finding an appropriate gravitational functional measure, given in the following general form:

$$D[g] = \prod_x f(g_{\alpha\beta})\prod_{\mu<\nu} d(g^{1/2}g^{\mu\nu}), \tag{2.7}$$

where f is an unknown homogeneous function, i.e., $f(\omega^2 g_{\alpha\beta}) = \omega^a f(g_{\alpha\beta})$. Thus we should determine the actual expression for f. Instead of doing so we will give two frequently used variants. Namely, the canonical measure proposed by Fradkin and Vilkovisky[8] is given by $f(g_{\alpha\beta}) = g^{-3/2}g^{00}$ and its homogeneity degree is $a = -14$. On the other hand, the scale-invariant measure is given in the form[9] $f(g_{\alpha\beta}) = g^{-5/2}$, where the homogeneity degree is $a = -20$. Thus the entire homogeneity degrees of the measure $D[g]$ (2.7) are $A = 6$ in the former case and $A = 0$ in the latter. It will appear in the sequel that for our purposes it suffices to guarantee the condition

$$A<8. \tag{2.8}$$

Now we are ready to give the final form of the partition function

$$Z = \int D[g]Da\,e^{-S^{\text{eff}}},$$

where S^{eff} is given by (2.6) and

$$Da = \prod_x\prod_\mu (2\pi)^{-1}\,da_\mu.$$

It is worth noting that the anomaly-free gravitational measure recently given by Fujikawa[10] does not fulfill the restriction (2.8) and therefore this case will not be dealt with in the paper.

III. THE HEAT KERNEL REPRESENTATION OF THE RENORMALIZED DETERMINANTS

As the further analysis indicates (viz. Sec. IV), our final result crucially depends on the behavior of the $\log\det(-\Delta)$ in the effective action S^{eff} (2.6).

First of all, the $\log\det(-\Delta)$ must be properly defined. We shall now derive an extremely convenient (for applications) heat kernel representation for the renormalized determinants. Formally we may write

$$\log\det H = -\int_0^\infty \frac{ds}{s}\,T(s), \tag{3.1}$$

where H is a non-negative operator (equal to $-\Delta$ in our case) and $T(s) = \operatorname{Tr} e^{-sH}$. Unfortunately (3.1) is divergent and must be renormalized. There are many known definitions of the renormalized $\log\det H$, e.g., zeta function, dimensional, point splitting, etc.[11] It seems that the zeta-function renormalization prescription is the most commonly used one in the curved space calculations.[12] However, in this work we will adopt Seiler's renormalization prescription[13] because of its straightforward interpretation in terms of the t'Hooft minimal subtractions in the Feynmann diagram language. It is remarkable that our final result does not depend on the definitions accepted as they differ by scale-invariant terms only.

The heat kernel $G(x,y;s) = \langle x|e^{-sH}|y\rangle$ satisfies the equation

$$\partial_s G(x,y;s) + HG(x,y;s) = 0,$$

where H is an elliptic second-order differential operator with the initial condition $G(x,y;0) = \delta(x-y)$. The short-distance expansion (x is close to y) of the heat kernel was given by DeWitt.[12] We will be interested only in the trace version of it. Thus we obtain the so-called Seeley expansion for small s,

$$\operatorname{Tr} e^{-sH} \equiv T(s) = \sum_{i=0}^{\infty} T_i s^{i-2}, \tag{3.2}$$

where the T_i are the Seeley or the Hadamard–Minakshisundaram–DeWitt ("Hamidew") coefficients.[5,14] For manifolds with boundaries the expansion (3.2) needs additional terms, abandoned here.[15]

Below we will give the first three Seeley ("Hamidew") coefficients for $H = -\Delta$ that are the most important from the renormalization point of view:

$$T_0 = (4\pi)^{-2}\int g^{1/2}\,d^4x, \tag{3.3a}$$

$$T_1 = (4\pi)^{-2}\int \frac{1}{6}Rg^{1/2}\,d^4x, \tag{3.3b}$$

$$T_2 = (4\pi)^{-2}\int\left(\frac{1}{180}R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu} - \frac{1}{180}R^{\alpha\beta}R_{\alpha\beta} + \frac{1}{72}R^2\right)g^{1/2}\,d^4x. \tag{3.3c}$$

Generally speaking, we have two kinds of divergences in (3.1): ultraviolet and infrared ones. The former originates from integration for $s\to 0$ and the latter may appear from integration for $s\to\infty$. The occurrence of the ultraviolet divergences results from the form of the Seeley expansion (3.2). It is easy to see that the first three terms in (3.2) give rise to divergences for $s\to 0$. The infrared divergences are connected with zero modes N of the operator H. Since Δ has a discrete spectrum for a compact manifold[11] we may rewrite $T(s)$ as

$$\sum_{\lambda_n} e^{-s\lambda_n} = \sum_{\lambda_n\neq 0} e^{-s\lambda_n} + N. \tag{3.4}$$

Here $N\neq 0$ gives rise to divergences in (3.1) for $s\to\infty$.

Definition: Let H be a non-negative elliptic differential operator, H_0 its free part, and μ^2 a constant associated with a mass scale. Then the analytically renormalized determinant is defined by the formula

$$\log \det_{AR} H = (\alpha!)^{-1} \int_0^\infty ds \log(\mu^2 s) \left(\frac{d}{ds}\right)^{\alpha+1} \times \{[T(s) - T_0(s)]s^\alpha\}, \qquad (3.5)$$

where $T(s) = \mathrm{Tr}\, e^{-sH}$, $T_0(s) = \mathrm{Tr}\, e^{-sH_0}$, and $[T(s) - T_0(s)]s^\alpha$ is smooth in s for some α.

Remarks: (1) $T_0(s)$ is coming from the normalization of the functional measure.

(2) In our case, i.e., $H = -\Delta$, it suffices to put $\alpha = 1$.

We shall derive now the explicit functional form of the $\log \det_{AR} H$ using the Seeley expansion.

Proposition 1: Let H be a non-negative elliptic second-order differential operator, N and N_0 be the numbers of zero modes for H and H_0, respectively, and C be the Euler constant. Then

$$\log \det_{AR} H = -\int_0^\infty \frac{ds}{s} [T(s) - T_0(s) - T_1 s^{-1} - T_2 e^{-\mu^2 s} - (N - N_0)(1 - e^{-\mu^2 s})] + (1 + C)[T_2 - (N - N_0)].$$

For the proof see Appendix B.

In the next section we will extensively use the following proposition.

Proposition 2:

$$\log \det_{AR} (\rho H) = \log \det_{AR} H + [T_2 - (N - N_0)] \log \rho, \qquad (3.6)$$

where $\rho \in R_+$.

For the proof see Appendix C (see Ref. 16).

IV. SCALING

To find announced unboundedness of the gravitational partition function Z we will have to rescale the fields in the suitable way. The scaling differs from the one proposed in Ref. 5 because it looks rather as a change of the variables in the path integral and therefore is expected to give an equivalent expression. We put

$$g_{\mu\nu}(x) \to \omega^2 g_{\mu\nu}(x), \qquad (4.1)$$

$$a_\mu(x) \to \omega^{-(1/4)A} a_\mu(x), \qquad (4.2)$$

where ω is a real positive constant and A is an entire homogeneity degree of the measure $D[g]$ (2.7). The former transformation, i.e., (4.1), represents an ordinary conformal transformation with a constant factor ω and is the principal one. The latter, i.e., (4.2), is an auxiliary transformation to assure the formal invariance of the measure (2.7).

Equation (4.1) implies

$$R \to \omega^{-2} R, \qquad (4.3)$$

$$\Delta \to \omega^{-2} \Delta. \qquad (4.4)$$

By virtue of (4.1) and (4.3) we have the following scaling property for the classical gravitational action S_g (2.1):

$$-(2\kappa)^{-1} \int (R - 2\Lambda) g^{1/2} d^4x \to \omega^2 S_0 + \omega^4 S_\Lambda \equiv -(2\kappa)^{-1} \omega^2 \int R g^{1/2} d^4x + (2\kappa)^{-1} \omega^4 \int 2\Lambda g^{1/2} d^4x. \qquad (4.5)$$

It is damped for a large value of ω by the cosmological term and tends to zero for small ω.

Using (4.1) and (4.2) we can find the scaling property for the gauge-fixing term S_{GF} (2.5)

$$S_{GF} \to \omega^{2-(1/4)A} S_{GF}. \qquad (4.6)$$

It is bounded for small ω unless the condition (2.8) is not fulfilled.

From (4.4) and (3.6) it follows that

$$\log \det_{AR}(-\Delta) \to \log \det_{AR}(-\Delta) + [T_2 - (N - N_0)] \log \omega^{-2}. \qquad (4.7)$$

It is obvious that in the small ω limit ["strong" fields limit, see (4.3)] the renormalized effective action $S^{\text{eff}}_{\text{ren}}$ is dominated by the term

$$T_2 - (N - N_0). \qquad (4.8)$$

We observe that the measure (2.7) is obviously formally invariant under the transformations (4.1) and (4.2). Consequently, the behavior of the partition function crucially depends on the term (4.8). We see that the whole effect follows in our approach from the renormalized Faddeev–Popov term in small ω limit. Instead, in Hawking's approach a general conformal transformation is performed with a rapidly varying factor ω (inadmissible in our gauge-fixing choice) of the classical action S_g (see Ref. 5).

Now, we will examine the behavior of the expression (4.8). It will be convenient to use the Gauss–Bonnet theorem,[17] which, in our closed four-dimensional case, bears the following simple form:

$$\chi(M) = (32\pi^2)^{-1} \times \int (R^{\alpha\beta\mu\nu} R_{\alpha\beta\mu\nu} - 4R^{\alpha\beta} R_{\alpha\beta} + R^2) g^{1/2} d^4x, \qquad (4.9)$$

where $\chi(M)$ is the Euler number of a manifold M. The Euler number $\chi(M)$ is given by the alternate sum of the Betti numbers B_i,

$$\chi(M) = \sum_{i=0}^{4} (-)^i B_i,$$

where $B_i = \dim H^i(M)$ and $H^i(M)$ is the cohomology class. By virtue of the Poincaré lemma,[17] $B_i = B_{4-i}$ for orientable manifolds. Consequently $\chi(M) = 2B_0 + B_2 - 2B_1$. If the manifold is simply connected, $B_1 (= B_3) = 0$ (see Ref. 14) so $\chi(M) \geqslant 2$. For a connected manifold $B_0 = 1$. Since we have

$$N = N_0 = B_i \qquad (4.10)$$

for the Laplace–deRham (Laplace–Beltrami in our case) operator on i-forms (0-forms, i.e., functions in our case) on closed manifolds, we may confine our attention to the term T_2 only.

By virtue of (4.9) and (4.10), the term (4.8) may be reexpressed in the following manner:

$$T_2-(N-N_0)$$
$$=T_2=\frac{1}{90}\chi(M)+\frac{1}{960\pi^2}$$
$$\times\int\left(R^{\alpha\beta}R_{\alpha\beta}+\frac{1}{2}R^2\right)g^{1/2}\,d^4x. \tag{4.11}$$

It follows from the above considerations that T_2 is always positive for simply connected manifolds independently on the metric $g_{\mu\nu}$; $\chi(M)\geqslant 2$ and the second term in (4.11) is evidently positive ($T_2\geqslant\frac{1}{45}$). Thus the gravitational partition function is unbounded in the small ω limit in this case. However, when nonsimply connected manifolds are taken into account, the situation is less clear. In any case, T_2 should be positive for a sufficiently rapidly varying metric $g_{\mu\nu}$. So it seems that the nonsimple topologies cannot improve the situation considerably.

We shall now present the main result. Assuming that the measure (2.7) makes mathematical sense, we have the following theorem.

Theorem: Let M^4 be an orientable closed simply connected four-dimensional manifold and Z the partition function for gravity, $Z\neq 0$. Then

$$Z=+\infty. \tag{4.12}$$

Proof: Using the scaling properties for the renormalized effective action $S^{\text{eff}}_{\text{ren}}$ described above [(4.5)–(4.7)] we obtain

$$Z=\int D[g]Da\exp(4T_2\log\omega^{-2}-S_{\text{FP}}$$
$$-\omega^{2-(1/4)A}S_{\text{GF}}-\omega^2S_0-\omega^4S_\Lambda).$$

Since $T\geqslant\frac{1}{45}$ for orientable closed simply connected manifolds we have

$$Z\geqslant\exp\left(\frac{4}{45}\log\omega^{-2}\right)\int D[g]Da$$
$$\times\exp(-S_{\text{FP}}-\omega^{2-(1/4)A}S_{\text{GF}}$$
$$-\omega^2S_0-\omega^4S_\Lambda)\underset{\omega\to 0}{\to}+\infty.$$

V. DISCUSSION

We have shown that there are some troubles with the partition function for gravity Z to exist. Namely, it has appeared that by the suitable rescaling of the fields we may convert the functional integrand to an arbitrarily big expression. Actually, the indefiniteness is coming from the renormalized Faddeev–Popov term in the degenerate harmonic gauge. However, it is well known (in the Yang–Mills case) that such gauges usually possess some drawbacks of the Gribov ambiguity type. Therefore one may wonder whether that is the case. As was stressed by Fradkin and Vilkovisky[8] the degenerate harmonic gauge is the most suitable for nonperturbative analysis in quantum gravity because of the "covariant shape" of the Faddeev–Popov operator.

It should be noticed that we do not perform any one-loop approximation. As the classical gravitational action lacks R^2-like terms, log $\det_{\text{AR}}$ may be understood as a redefinition of the action rather than a conventional renormalization of a coupling constant. Thus, our $S^{\text{eff}}_{\text{ren}}$ is in some sense classical and exact.

Referring to the boundary conditions, we only mention that they should not affect the results in the essential way. We must only add some surface terms to the Seeley ("Hamidew") coefficients[15] and to the action.

The only possible chance to improve the situation is the restriction of the space-time topology. Unfortunately, although it helps a little, it does not suffice.

It should be interesting to add to the action some matter fields coupled (minimally or nonminimally, maybe in the supersymmetric way) to gravity. As fermion and boson fields enter the effective action with opposite signs, the result may be radically changed. Roughly speaking, nongauge bosons improve and fermions make the stability worse. Of course, gauge fields introduce additional ghost fields that act in the opposite direction. Thus the situation is more involved and stability depends on topology, in the essential way,[18] and on the field content.

ACKNOWLEDGMENTS

I would like to thank Professor R. Rączka for many helpful discussions and continuous help.

APPENDIX A: DERIVATION OF THE FADDEEV–POPOV TERM

We have

$$\delta g^{\mu\nu}=-\xi^\lambda\partial_\lambda g^{\mu\nu}+g^{\mu\lambda}\partial_\lambda\xi^\nu+g^{\lambda\nu}\partial_\lambda\xi^\mu$$

[see (2.2)] and

$$\delta g=-\xi^\lambda\partial_\lambda g-2g\,\partial_\lambda\xi^\lambda.$$

Denoting $h^{\mu\nu}=g^{1/2}g^{\mu\nu}$, we get

$$\delta h^{\mu\nu}=-\partial_\lambda(\xi^\lambda h^{\mu\nu})+\partial_\lambda(h^{\mu\lambda}\xi^\nu)$$
$$-\partial_\lambda h^{\mu\lambda}\xi^\nu+h^{\nu\lambda}\partial_\lambda\xi^\mu.$$

In the degenerate harmonic gauge, i.e., $\partial_\nu h^{\mu\nu}=0$, we obtain

$$\partial_\nu\delta h^{\mu\nu}=\partial_\nu(h^{\nu\lambda}\partial_\lambda\xi^\mu).$$

The Faddeev–Popov formula now reads

$$\int\bar\theta^\sigma\delta_{\sigma\mu}\,\partial_\nu(h^{\nu\lambda}\partial_\lambda\theta^\mu)d^4x$$
$$=\int\bar\theta^\sigma\delta_{\sigma\mu}\Delta\theta^\mu g^{1/2}\,d^4x,$$

where Δ is Laplace–Beltrami operator. We see that $g^{1/4}$ should be absorbed in the Faddeev–Popov ghost fields in the functional measure. It only causes a redefinition of the gravitational sector in the functional measure.

APPENDIX B: PROOF OF PROPOSITION 1

For $\alpha=1$,

$$\log\det_{\text{AR}}H=\int_0^\infty ds\log(\mu^2s)\frac{d^2}{ds^2}\{[T(s)-T_0(s)]s\} \tag{B1}$$

has the following explicit representation:

$$\log \det_{AR} H = -\int_0^\infty \frac{ds}{s}\,[T(s) - T_0(s) - T_1 s^{-1} - T_2 e^{-\mu^2 s} - (N - N_0)(1 - e^{-\mu^2 s})] + (1 + C)[T_2 - (N - N_0)], \tag{B2}$$

where N and N_0 are the numbers of zero modes for H and H_0, respectively, $T(s) \equiv \operatorname{Tr} e^{-sH}$, $T_0(s) \equiv \operatorname{Tr} e^{-sH_0}$, and C is the Euler constant.

Proof: Integrating twice by parts in (B1) we obtain

$$\log \det_{AR} H = \log(\mu^2 s)\{[T(s) - T_0(s)]s\}'|_0^\infty - [T(s) - T_0(s)]|_0^\infty - \int_0^\infty \frac{ds}{s}\,[T(s) - T_0(s)]. \tag{B3}$$

By virtue of (3.4), we have, for $\Lambda \to +\infty$,

$$\{[T(s) - T_0(s)]s\}'|_{s=\Lambda} = N - N_0, \tag{B4a}$$

$$[T(s) - T_0(s)]|_{s=\Lambda} = N - N_0. \tag{B4b}$$

Instead, by virtue of (3.2) and (3.3a), we have for

$$\{[T(s) - T_0(s)]s\}'|_{s=\epsilon} = T_2, \tag{B5a}$$

$$[T(s) - T_0(s)]|_{s=\epsilon} = T_1 \epsilon^{-1} + T_2. \tag{B5b}$$

Now we observe that

$$\log(\mu^2 \Lambda) = \int_0^\Lambda \frac{ds}{s}\,(1 - e^{-\mu^2 s}) - C,$$

$$\log(\mu^2 \epsilon) = -\int_\epsilon^\infty \frac{ds}{s}\,e^{-\mu^2 s} - C,$$

$$\epsilon^{-1} = \int_\epsilon^\infty s^{-2}\,ds.$$

Inserting these expressions into (B5) and next (B4) and (B5) into (B3) we obtain (B2).

APPENDIX C: PROOF OF PROPOSITION 2

$$\log \det_{AR}(\rho H) = \log \det_{AR} H + [T_2 - (N - N_0)]\log \rho, \tag{C1}$$

where $\rho \in \mathbf{R}$ and $\rho > 0$.

Proof: By (B1) we have

$$\log \det_{AR}(\rho H) = \int_0^\infty ds\,\log(\mu^2 s)\,\frac{d^2}{ds^2}\{[T(\rho s) - T_0(\rho s)]s\}.$$

Now set $s' = \rho s$. Hence

$$\log \det_{AR}(\rho H) = \log \det_{AR} H + \log \rho^{-1}\{[T(s') - T_0(s')]\}|_0^\infty.$$

According to (B4a) and (B5a) we obtain (C1).

[1]R. Rączka, Ann. Phys. (NY) **160**, 355 (1985).
[2]E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[3]L. F. Abbott and S. Deser, Nucl. Phys. B **195**, 76 (1982).
[4]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge, U.P., Cambridge, 1973).
[5]S. W. Hawking, in *General Relativity: An Einstein Centennary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979).
[6]L. D. Faddeev and V. N. Popov, Phys. Lett. B **25**, 30 (1967).
[7]This comes from Refs. 6 and 8.
[8]E. S. Fradkin and G. A. Vilkovisky, Phys. Rev. D **8**, 4241 (1973).
[9]N. P. Konopleva and V. N. Popov, *Gauge Fields* (Harwood Academic, New York, 1981).
[10]K. Fujikawa, Nucl. Phys. B **226**, 437 (1983).
[11]R. M. Wald, Commun. Math. Phys. **70**, 221 (1979).
[12]B. S. DeWitt, in Ref. 5.
[13]E. Seiler, *Proceedings—Gauge Theories: Fundamental Interactions and Rigorous Results* (Brasov, Romania, 1981).
[14]S. W. Hawking, *Recent Developments in Gravitation*, edited by M. Levy and S. Deser (Plenum, New York, 1979).
[15]E. S. Fradkin and A. A. Tseytlin, Nucl. Phys. B **201**, 469 (1983).
[16]Proof is a trivial one in the zeta-function renormalization prescription. See, for example, S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[17]T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).
[18]B. Broda, talk given at International Conference XIII DGMTP, Shumen, Bulgaria, 1984.

Path integral for second-derivative Lagrangian $L = (\kappa/2)\ddot{x}^2 + (m/2)\dot{x}^2 + (k/2)x^2 - j(\tau)x(\tau)$

H. Kleinert
Department of Physics, University of California, San Diego, La Jolla, California 92093

(Received 19 February 1986; accepted for publication 6 August 1986)

For the above second-derivative Euclidean Lagrangian, the quantum statistical probability distribution $(x_b v_b \tau_b | x_a v_a \tau_a)$ that an orbit $x(\tau)$ with initial position x_a and velocity $\dot{x}_a = v_a$ arrives at a final point x_b and velocity $\dot{x}_b = v_b$ is calculated.

I. INTRODUCTION

The behavior of many physical systems cannot be understood without allowing for higher-gradient terms in the field energy. In relativistic quantum field theory such terms have not enjoyed much popularity, due to notorious difficulties with positivity of either the energy or of the metric of the quantum mechanical Hilbert space.[1] In statistical mechanics, however, such terms are ubiquitous and impossible to avoid. Some examples follow.

(1) Polymers on an intermediate distance scale are stiff objects and their energy requires the inclusion of a bending energy which involves the square of the second derivative, $\ddot{x}^2(s)$, where s is the length parameter of the space curve.[2]

(2) The walls of many living cells are free of tension and undergo fluctuations controlled mainly by second-gradient curvature energy.[3] This makes the fluctuations so large that they can be seen in an ordinary light microscope, as first observed on human red blood cells in 1890.[4] These giant fluctuations are crucial to prevent the cells from sticking to each other, in spite of their attractive van der Waals forces.[5]

(3) The formation of microemulsions cannot take place without the ampliphilic soap layer between water and oil losing its surface tension.[6]

(4) The strings of color electric flux lines, which bind quarks and antiquarks, can lose their tension in a phase transition, in which case they are controlled completely by second-gradient elasticity.[7]

(5) Finally, the cosmos at an early stage of evolution may not have been controlled by the Einstein action, but by the Weyl action which involves the square of the curvatures and thus contains the square of two derivatives of the metric. The geophysically observed deviations from Newton's law, when masses come closer to each other than ≈ 200 m, could be a signal for such terms (the sign is correct).[8]

II. THE PATH INTEGRAL

In all these physical situations, the prototype of the fluctuation problem to be solved is the path integral

$$(x_b v_b \tau_b | x_a v_a \tau_a) = \int \mathscr{D}x(\tau) \exp\left(- \int_{\tau_a}^{\tau_b} d\tau\, L(\tau) \right),$$
$$x(\tau_a) = x_a, \quad \dot{x}(\tau_a) = v_a, \qquad (1)$$
$$x(\tau_b) = x_b, \quad \dot{x}(\tau_b) = v_b,$$

with the Euclidean Lagrangian ($\cdot \equiv d/d\tau$)

$$L(\tau) = \frac{\kappa}{2}\ddot{x}^2(\tau) + \frac{m}{2}\dot{x}^2(\tau) + \frac{k}{2}x^2(\tau) - j(\tau)x(\tau). \quad (2)$$

In order to make all integrals convergent we have rotated the time variable t to imaginary values $t = -i\tau$.

After rescaling the variables τ to $\tau = \kappa^{-1/3}\tau_{\rm old}$ and introducing the frequencies ω_1, ω_2 via

$$\omega_1^2 + \omega_2^2 = (m/2)\kappa^{-1/3}, \quad \omega_1^2\omega_2^2 = k\kappa^{1/3}, \qquad (3)$$

we are confronted with the probability distribution

$$(x_b v_b \tau_b | x_a v_a \tau_a) = \int \mathscr{D}x(\tau) \exp\left(- \int_{\tau_a}^{\tau_b} d\tau\, L(\tau) \right),$$
$$x(\tau_a) = x_a, \quad \dot{x}(\tau_a) = v_a, \qquad (4)$$
$$x(\tau_b) = x_b, \quad \dot{x}(\tau_b) = v_b,$$

where L is now the Euclidean Lagrangian

$$L = \tfrac{1}{2}[\ddot{x}^2 + (\omega_1^2 + \omega_2^2)\dot{x}^2 + \omega_1^2\omega_2^2 x^2] - j(\tau)x(\tau) \qquad (5)$$

with an appropriately rescaled current ($j = \kappa^{1/3} j_{\rm old}$). This can be separated into a pure surface term

$$L_{\rm sf} = \frac{d}{dt}\Lambda = \frac{d}{dt}\left[\frac{1}{2}(\dot{x}\ddot{x} - x\dddot{x}) + (\omega_1^2 + \omega_2^2)x\dot{x}\right], \qquad (6a)$$

plus a source term

$$L_{\rm source} = -\int_{\tau_a}^{\tau_b} d\tau\, j(\tau)x(\tau), \qquad (6b)$$

plus a term

$$L_0 = \tfrac{1}{2}x(\ddddot{x} - (\omega_1^2 + \omega_2^2)\ddot{x} + \omega_1^2\omega_2^2 x), \qquad (6c)$$

which vanishes for solutions of the free field equation

$$(\partial_\tau^2 - \omega_1^2)(\partial_\tau^2 - \omega_2^2)x(\tau) = 0. \qquad (7)$$

These correspond to two independent harmonic oscillators and have the general form

$$x_{\rm cl}(\tau) = A\cosh\omega_1(\tau - \tau_a) + B\sinh\omega_1(\tau - \tau_a) + C\cosh\omega_2(\tau - \tau_a) + D\sinh\omega_2(\tau - \tau_a). \quad (8)$$

(The two oscillators can be exhibited by introducing the two auxiliary variables $q_1 = \ddot{x} - \omega_1^2 x$, $q_2 = \ddot{x} - \omega_2^2 x$ and noting that $L_3 = [q_1(\partial_\tau^2 - \omega_1^2)q_1 - q_2(\partial_\tau^2 - \omega_2^2)q_2]/(\omega_1^2 - \omega_2^2)$. The negative sign in front of the second term leads to the difficulties with a quantum mechanical formulation due to an indefinite Hamiltonian.) The proper measure of integration in the path integral (4) is found via the canonical formalism. For a higher-gradient Lagrangian (3) we can follow the method of Ostrogradski,[9] according to which $\dot{x}(\tau) \equiv v(\tau)$ may be considered as an *independent* degree of freedom replacing the Lagrangian (5) by the equivalent one

$$\tilde{L} = \tfrac{1}{2}(\dot{v}^2 + (\omega_1^2 + \omega_2^2)v^2 + \omega_1^2\omega_2^2 x^2) - ip(\dot{x} - v) - jx, \qquad (5')$$

where the Lagrangian multiplier $p(\tau)$ ensures the correct relation between v and $\dot{x}$. The canonical momenta are $p = i(\partial L/\partial \dot{x})$ and $p_v = i(\partial \tilde{L}_E/\partial \dot{v}) = i\dot{v}$ such that the Hamiltonian is

$$\begin{aligned} H(p,x,p_v,v,\tau) &= ip\dot{x} + ip_v\dot{v} + \tilde{L} \\ &= \tfrac{1}{2}(p_v^2 + (\omega_1^2 + \omega_2^2)v^2 \\ &\quad + \omega_1^2\,\omega_2^2 x^2)ip_v - j(\tau)x\,, \end{aligned} \tag{9}$$

with the Hamiltonian equations of motion

$$\begin{aligned} i\dot{p} &= -\frac{\partial H}{\partial x} = -\omega_1^2\,\omega_2^2 x + j(\tau)\,, \\ i\dot{p}_v &= -\frac{\partial H}{\partial v} = -(\omega_1^2 + \omega_2^2)v - ip\,, \\ i\dot{v} &= \frac{\partial H}{\partial p_v} = p_v, \quad i\dot{x} = \frac{\partial H}{\partial p} = iv\,. \end{aligned} \tag{10}$$

It is now straightforward to specify the measure of the path integral. In phase space it has the form

$$\begin{aligned} &(x_b v_b \tau_b | x_a v_a \tau_a) \\ &= \int \mathscr{D}x\,\mathscr{D}v \int \frac{\mathscr{D}p}{2\pi} \int \frac{\mathscr{D}p_v}{2\pi} \\ &\quad\times \exp\left(\int_{\tau_a}^{\tau_v} d\tau (ip\dot{x} + ip_v\dot{v} - H(p,x,p_v,v,\tau))\right), \end{aligned} \tag{11}$$

where $\int \mathscr{D}x$ means, as usual, the product of all

$$\prod_{n=1}^{N} \int_{-\infty}^{\infty} dx_n$$

over the time sliced positions

$$x_n \equiv x(\tau_n), \quad \tau_n \equiv \tau_a + \epsilon n\,,$$

where $\epsilon = (\tau_b - \tau_a)/(N+1)$ and $x_a = x(\tau_0)$, $x_b = x(\tau_{N+1})$ are held fixed, and $\int \mathscr{D}p/2\pi$ is the product of integrals

$$\prod_{n=1}^{N+1} \int_{-\infty}^{\infty} \frac{dp_n}{2\pi}$$

involving all $N+1$ momenta that appear in the canonical term

$$\int_{\tau_a}^{\tau_b} d\tau\, ip\dot{x} = \sum_{n=1}^{N+1} ip_n(x_n - x_{n-1})\,.$$

The same rule applies to the conjugate variable pair v and p_v, which are split into $v_1,\dots,v_{N+1}$ and $p_{v_1},\dots,p_{v_{N+1}}$ with $v_1 = v_a$, $v_{N+1} = v_b$ held fixed and the canonical integral measure is

$$\prod_{n=2}^{N} \int_{-\infty}^{\infty} dv_n \prod_{n=2}^{N+1} \int_{-\infty}^{\infty} \frac{dp_{v_n}}{2\pi}\,. \tag{12}$$

By construction, the amplitude (11) and hence also (4) satisfies the Schrödinger equation

$$\begin{aligned} &(H(-i\partial_x, x, -i\partial_v, v, \tau) + \partial_\tau)(xv\tau|x'v'\tau') \\ &= (-\tfrac{1}{2}\partial_v^2 + (\omega_1^2 + \omega_2^2)v^2 + \omega_1^2\,\omega_2^2 x^2 \\ &\quad + v\partial_x - j(\tau)x + \partial_\tau)(xv\tau|x'v'\tau') \\ &= \delta(x - x')\delta(v - v')\delta(\tau - \tau')\,. \end{aligned} \tag{13}$$

Using (11), it is now straightforward to obtain the measure of integration for the pure x space path integral (3) as follows: Integrating out the variables p_2 and p_N gives a product of δ functions

$$\prod_{n=2}^{N} \delta(x_n - x_{n-1} - \epsilon v_n)\,, \tag{14}$$

which can be used to eliminate the integrals over $v_2,\dots,v_N$, thereby producing a factor $1/\epsilon^{N-1}$. The integrals over $pv_2,\dots,pv_{N+1}$ produce a further factor $(1/\sqrt{2\pi\epsilon})N$. Thus the measure of the path integral (3) can be written as follows:

$$\begin{aligned} \int \mathscr{D}x(\tau) &= \epsilon \prod_{n=1}^{N} \left[\int_{-\infty}^{\infty} \frac{dx_n}{\sqrt{2\pi\epsilon}\,\epsilon}\right] \int_{-\infty}^{\infty} \frac{dp_1}{2\pi} \int_{-\infty}^{\infty} \frac{dp_{N+1}}{2\pi} \\ &\quad\times \exp[ip_{N+1}(x_{N+1} - x_N - \epsilon v_{N+1}) \\ &\quad + ip_1(x_1 - x_0 - \epsilon v_1)]\,. \end{aligned} \tag{15}$$

We may now also integrate out the remaining two momentum integrals, thereby eliminating the integrations over dx_1 and dx_0. For the calculations to come it will, however, be convenient to leave the measure in this form.

Due to the quadratic form of the energy, the integration over the spatial variables can most easily be done following the same procedure as developed for lowest-gradient quadratic energy in Feynman and Hibbs.[10] We expand all orbits around a fixed classical trajectory $x_{\rm cl}(\tau)$, which connects the initial point x_a at velocity v_a with the final point x_b at velocity v_b, and write $x(\tau) = x_{\rm cl}(\tau) + \delta x(\tau)$. The fluctuations $\delta x(\tau)$ then have the property that

$$\begin{aligned} \delta x(\tau_a) &= \delta x(\tau_b) = 0\,, \\ \delta v(\tau_a) &= \delta v(\tau_v) = 0\,. \end{aligned} \tag{16}$$

Inserting this into the action $\mathscr{A} = \int_{\tau_a}^{\tau_b} d\tau\, L(\tau)$, there is a classical contribution coming entirely from the surface term (6a),

$$\begin{aligned} \mathscr{A}_{\rm cl,sf} &= \tfrac{1}{2}[v_b \ddot{x}_b - x_b \dddot{x}_b \\ &\quad + \tfrac{1}{2}(\omega_1^2 + \omega_2^2)x_b v_b - (ab)]\,|_{x = x_{\rm cl}(\tau)}\,, \end{aligned} \tag{17}$$

plus a contribution from the source term (6b)

$$\mathscr{A}_{\rm cl,source} = -\int_{\tau_a}^{\tau_b} d\tau\, j(\tau) x_{\rm cl}(\tau)\,, \tag{18}$$

plus a fluctuation piece,

$$\begin{aligned} \mathscr{A}_{\rm fl} &= \int_{\tau_a}^{\tau_b} d\tau \left\{\frac{1}{2}[(\delta\ddot{x})^2 + (\omega_1^2 + \omega_2^2)(\delta\dot{x})^2 + (\delta x)^2] \right. \\ &\quad \left. - j(\tau)\delta x(\tau)\right\}. \end{aligned} \tag{19}$$

Hence we may write

$$\begin{aligned} &(x_b v_b \tau_b | x_a v_a \tau_a) \\ &= e^{-\mathscr{A}_{\rm cl,sf} - \mathscr{A}_{\rm cl,source}} \int \mathscr{D}\,\delta x(\tau) \\ &\quad\times \exp\left(\int_{\tau_a}^{\tau_b} \left\{\frac{1}{2}[(\delta\ddot{x})^2 + (\omega_1^2 + \omega_2^2)(\delta\dot{x})^2 \right.\right. \\ &\quad \left.\left. + \omega_1^2\,\omega_2^2(\delta x)^2] - j(\tau)\delta x(\tau)\right\}\right), \end{aligned} \tag{20}$$

$$\delta x(\tau_a) = 0, \quad \delta x(\tau_b) = 0\,,$$

$$\delta\dot{x}(\tau_a)=0,\quad \delta\dot{x}(\tau_b)=0,$$

where $\delta x(\tau)$ has now a measure like (15) except that δx_{N+1}, δx_0, δv_{N+1}, and δv_1 vanish at the end points, i.e.,

$$\int \mathscr{D}\,\delta x(\tau)=\epsilon\prod_{n=1}^{N}\left[\int_{-\infty}^{\infty}\frac{dx_n}{\sqrt{2\pi\epsilon}\,\epsilon}\right]\int_{-\infty}^{\infty}\frac{dp_1}{2\pi}$$
$$\times\int_{-\infty}^{\infty}\frac{dp_{N+1}}{2\pi}e^{-i(p_{N+1}\delta x_N-p_1\delta x_1)}.\tag{21}$$

The vanishing of δx_0, δx_{N+1} implies that $\delta x(\tau)$ has only the Fourier components

$$\delta x(\tau)=\sqrt{\frac{2}{\beta}}\sum_{m=1}^{N}\delta x_m\sin\nu_m(\tau-\tau_a),\tag{22}$$

with $\beta\equiv\tau_b-\tau_a$ and frequencies

$$\nu_m=(\pi/\beta)m.\tag{23}$$

In terms of the Fourier components the exponential in the fluctuation factor (20) reads

$$\exp\left\{-\frac{1}{2}\sum_{m=1}^{N}(\Omega_m^2+\omega_1^2)(\Omega_m^2+\omega_2^2)(\delta x_m)^2\right.$$
$$\left.+\int_{\tau_a}^{\tau_b}d\tau\,j(\tau)\sqrt{\frac{2}{\beta}}\sum_m\delta x_m\sin\nu_m(\tau-\tau_a)\right\},$$

where

$$\Omega_m^2\equiv(1/\epsilon^2)(2-2\cos\nu_m\epsilon)=(4/\epsilon^2)\sin^2(\nu_m\epsilon/2)\tag{24}$$

are the squared eigenvalues of the differences $[(x_n-x_{n-1})/\epsilon]$ in Fourier space. Since the Fourier series has a unit Jacobian, the measure (21) becomes

$$\int\mathscr{D}\,\delta x(\tau)=\epsilon\prod_{m=1}^{N}\int_{-\infty}^{\infty}\frac{d\delta x_m}{\sqrt{2\pi\epsilon}\,\epsilon}\int_{-\infty}^{\infty}\frac{dp_1}{2\pi}$$
$$\times\int_{-\infty}^{\infty}\frac{dp_{N+1}}{2\pi}\epsilon^{-i(p_{N+1}\delta x_N-p_1\delta x_1)}.\tag{25}$$

Neglecting for a moment the couplings to p_{N+1} and p_1 and to the current $j(\tau)$, the pure δx_m part of the integrals gives the product

$$\epsilon\left[\prod_{m=1}^{N}(\epsilon^2\Omega_m^2+\epsilon^2\omega_1^2)(\epsilon^2\Omega_m^2+\epsilon^2\omega_2^2)\right]^{-1/2}$$
$$\xrightarrow[N\to\infty]{}\epsilon\sqrt{\frac{\omega_1\epsilon}{\sinh\omega_1\beta}}\sqrt{\frac{\omega_2\epsilon}{\sinh\omega_2\beta}}.\tag{26}$$

In Fourier space, the couplings to p_{N+1}, p_1, and the current amount to

$$-ip_{N+1}\sqrt{\frac{2}{\beta}}\sum_{m=1}^{N}\delta x_m\sin(\nu_m\epsilon)$$
$$+ip_1\sqrt{\frac{2}{\beta}}\sum_{m=1}^{N}\delta x_m\sin(\nu_m(\beta-\epsilon))$$
$$+\int_{\tau_a}^{\tau_b}d\tau\,j(\tau)\sqrt{\frac{2}{\beta}}\sum_{m=1}^{N}\delta x_n\sin\nu_m(\tau-\tau_a).\tag{27}$$

This can be rewritten as

$$\sqrt{\frac{2}{\beta}}\Bigg[-i(p_{N+1}-p_1)\sum_{m=1,3,5,\ldots}\delta x_m\sin(\nu_m\epsilon)$$
$$-i(p_{N+1}+p_1)\sum_{m=2,4,6,\ldots}\delta x_m\sin(\nu_m\epsilon)$$
$$+\int_{\tau_a}^{\tau_b}d\tau\,j(\tau)\sum_{m=1,2,3,\ldots}\delta x_m\sin(\nu_m(\tau-\tau_a))\Bigg].\tag{28}$$

In the absence of a current, these terms lead, after the δx_m integrations, to the additional momentum integrals

$$\int_{-\infty}^{\infty}\frac{dp_1}{2\pi}\int_{-\infty}^{\infty}\frac{dp_2}{2\pi}\exp\left\{-\frac{\epsilon^2}{2}(p_{N+1}-p_1)^2\right.$$
$$\times\frac{2}{\beta}\sum_{m=1,3,5,\ldots}\frac{(1/\epsilon^2)\sin^2(\nu_m\epsilon)}{(\Omega_m^2+\omega_1^2)(\Omega_m^2+\omega_2^2)}$$
$$-\frac{\epsilon^2}{2}(p_{N+1}+p_1)^2$$
$$\left.\times\frac{2}{\beta}\sum_{m=2,4,6,\ldots}\frac{(1/\epsilon^2)\sin^2(\nu_m\epsilon)}{(\Omega_m^2+\omega_1^2)(\Omega_m^2+\omega_2^2)}\right\}.\tag{29}$$

Due to their fast convergence, the sums can be replaced, for $\epsilon\to 0$, by

$$\frac{2}{\beta}\sum_m\frac{\nu_m^2}{(\nu_m^2+\omega_1^2)(\nu_m^2+\omega_2^2)}$$
$$=\frac{2}{\beta}\frac{1}{\omega_1^2-\omega_2^2}\sum_m\left(\frac{\omega_1^2}{\nu_m^2+\omega_1^2}-\frac{\omega_2^2}{\nu_m^2+\omega_2^2}\right).\tag{30}$$

These sums are equal to

$$D_e=\frac{1}{\beta(\omega_1^2-\omega_2^2)}\left(\frac{\omega_1\beta}{2}\coth\frac{\omega_1\beta}{2}-(12)\right),\quad\text{for even } m,$$
$$D_o\equiv\frac{1}{\beta(\omega_1^2-\omega_2^2)}\left(\frac{\omega_1\beta}{2}\tanh\frac{\omega_1\beta}{2}-(12)\right),\quad\text{for odd } m,\tag{31}$$

such that the integrations over p_{N+1}, p_1 yield the further factor

$$\frac{1}{2\pi}\frac{1}{\epsilon^2}\frac{1}{2}\frac{1}{\sqrt{D_eD_o}}=\frac{1}{2\pi\epsilon^2}\frac{\beta}{2}\frac{|\omega_1^2-\omega_2^2|}{[(\omega_1\beta/2)\coth(\omega_1\beta/2)-(12)]^{1/2}[(\omega_1\beta/2)\tanh(\omega_1\beta/2)-(12)]^{1/2}}$$
$$=\frac{\beta}{2\pi\epsilon^2}\frac{(\omega_1^2-\omega_2^2)\sqrt{\sinh\omega_1\beta\sinh\omega_2\beta}}{\sqrt{(\omega_1^2+\omega_2^2)\sinh\omega_1\beta\sinh\omega_2\beta-2\omega_1\omega_2(\cosh\omega_1\beta\cosh\omega_2\beta-1)}}.\tag{32}$$

If the current is nonzero, the expression (29) is replaced by

$$\int_{-\infty}^{\infty}\frac{dp_1}{2\pi}\int_{-\infty}^{\infty}\frac{dp_2}{2\pi}\exp\frac{1}{2}\left\{\frac{2}{\beta}\sum_{m=1,3,5,\ldots}\left(\int_{\tau_a}^{\tau_b}d\tau\,j(\tau)\sin\nu_m(\tau-\tau_a)-i(p_{N+1}-p_1)\sin\nu_m\epsilon\right)\right.$$

$$\times\left(\int_{\tau_a}^{\tau_b} d\tau' j(\tau')\sin\nu_m(\tau'-\tau_a) - i(p_{N+1}-p_1)\sin\nu_m\epsilon\right)[(\Omega_m^2+\omega_1^2)(\Omega_m^2+\omega_2^2)]^{-1}$$
$$+\frac{2}{\beta}\sum_{m=2,4,5,\ldots}\left(\int_{\tau_a}^{\tau_b} d\tau j(\tau)\sin\nu_m(\tau-\tau_a) - i(p_{N+1}+p_1)\sin\nu_m\epsilon\right)$$
$$\times\left(\int_{\tau_a}^{\tau_b} d\tau' j(\tau')\sin\nu_m(\tau'-\tau_a) - i(p_{N+1}+p_1)\sin\nu_m\epsilon\right)[(\Omega_m^2+\omega_1^2)(\Omega_m^2+\omega_2^2)]^{-1}\Bigg\}.$$

This gives an additional term in (29),

$$\exp\left\{\frac{1}{2}\int_{\tau_a}^{\tau_b} d\tau\int_{\tau_a}^{\tau_b} d\tau'\,\hat{G}(\tau,\tau')j(\tau)j(\tau')\right\}$$
$$\times\exp\left\{-i\epsilon(p_{N+1}-p_1)\int_{\tau_a}^{\tau_b} d\tau j(\tau)h_o(\tau)\right.$$
$$\left.-i\epsilon(p_{N+1}+p_1)\int d\tau j(\tau)h_e(\tau)\right\},$$

where

$$\hat{G}(\tau,\tau')=\frac{2}{\beta}\sum_m\frac{\sin\nu_m(\tau-\tau_a)\sin\nu_m(\tau'-\tau_a)}{(\Omega_m^2+\omega_1^2)(\Omega_m^2+\omega_2^2)},$$
$$h_o(\tau)=\frac{2}{\beta}\sum_{m=1,3,5,\ldots}\frac{(1/\epsilon)\sin\nu_m\epsilon\,\sin\nu_m(\tau-\tau_a)}{(\Omega_m^2+\omega_1^2)(\Omega_m^2-\omega_2^2)},\tag{33}$$
$$h_e(\tau)=\frac{2}{\beta}\sum_{m=2,4,6,\ldots}\frac{(1/\epsilon)\sin\nu_m\epsilon\,\sin\nu_m(\tau-\tau_a)}{(\Omega_m^2+\omega_1^2)(\Omega_m^2-\omega_2^2)}.$$

If we now integrate out the momenta p_{N+1}, p_1 the external source yields the factor

$$\exp\left\{\frac{1}{2}\int_{\tau_a}^{\tau_b} d\tau\int_{\tau_a}^{\tau_b} d\tau' j(\tau)G(\tau,\tau')j(\tau')\right\},\tag{34}$$

where

$$G(\tau,\tau')=\hat{G}(\tau,\tau')-\frac{h_e(\tau)h_e(\tau')}{D_e}-\frac{h_o(\tau)h_o(\tau')}{D_o}\tag{35}$$

This is the correlation function of the fluctuations $\delta x(\tau)$,

$$\langle\delta x(\tau)\delta x(\tau')\rangle = G(\tau,\tau').\tag{36}$$

Since $\delta x(\tau)$ vanishes at the end points $\tau=\tau_b$, $\tau'=\tau_a$ and has zero velocities, also $G(\tau,\tau')$ must have this property. Indeed, the vanishing of $G(\tau,\tau_a)$ and $G(\tau_b,\tau)$ is trivial to see. The zero velocity at the end points, on the other hand, is a consequence of the two properties [which follows directly from (33) and (31)]:

$$\frac{d}{d\tau'}G(\tau,\tau')|_{\tau'=\tau_a}=h_e(\tau)+h_o(\tau),\tag{37a}$$
$$\frac{d}{d\tau'}h_{\substack{e\\o}}(\tau')|_{\tau'=\tau_a}=D_{\substack{e\\o}}.\tag{37b}$$

Hence

$$\frac{d}{d\tau'}G(\tau,\tau')|_{\tau'=\tau_a}=0.\tag{38}$$

Collecting all terms, we arrive at the probability distribution

$$(x_b v_b \tau_b|x_a v_a \tau_a)$$
$$=F(\beta)\exp\left\{-\mathscr{A}_{\mathrm{cl,sf}}-\mathscr{A}_{\mathrm{cl,source}}\right.$$
$$\left.+\frac{1}{2}\int_{\tau_a}^{\tau_b} d\tau\int_{\tau_a}^{\tau} d\tau' j(\tau)G(\tau,\tau')j(\tau')\right\},\tag{39}$$

with the fluctuation factor

$$F(\beta)=\frac{1}{2\pi}\sqrt{\omega_1\omega_2}\frac{|\omega_1^2-\omega_2^2|}{\sqrt{(\omega_1^2+\omega_2^2)\sinh\omega_1\beta\sinh\omega_2\beta-2\omega_1\omega_2(\cosh\omega_1\beta\cosh\omega_2\beta-1)}}.\tag{40}$$

The terms $\mathscr{A}_{\mathrm{cl,sf}}$ and $\mathscr{A}_{\mathrm{cl,source}}$ are the only ones that depend on the initial and final variables $x_a v_a$, $x_b v_b$. They will be calculated in the following two sections.

III. THE CLASSICAL ACTION FOR ZERO EXTERNAL SOURCE

The starting point is formula (18). All we have to do is express the quantities $\dot{x}_b, \ddot{x}_b, \dot{x}_a$, and $\ddot{x}_a$ in terms of the initial and final variables $x_b v_b$, $x_b v_b$. For this purpose we invert the matrix relation

$$\begin{pmatrix}x_a\\x_b\\x_a\\x_b\end{pmatrix}=M\begin{pmatrix}A\\B\\C\\D\end{pmatrix},\tag{41}$$

with

$$
M=\begin{pmatrix} 1 & 0 & 1 & 0 \\ c_1 & s_1 & c_2 & s_2 \\ 0 & \omega_1 & 0 & \omega_2 \\ \omega_1 s_1 & \omega_1 c_1 & \omega_2 s_2 & \omega_2 c_2 \end{pmatrix}, \tag{42}
$$

where $c_{\substack{1\\2}} \equiv \cosh \omega_{\substack{1\\2}}\beta$, $s_{\substack{1\\2}} \equiv \sinh \omega_{\substack{1\\2}}\beta$, and find

$$
M^{-1} = (1/|M|)R , \tag{43}
$$

where $|M|$ is the determinant

$$
|M| = (\omega_1^2+\omega_2^2)s_1 s_2 - 2\omega_1\omega_2(c_1c_2-1) , \tag{44}
$$

and thus precisely equal to the expression under the last square root in the fluctuation factor (40). The matrix R is equal to

$$
R=\begin{pmatrix} -\omega_1\omega_2(c_1c_2-1)+\omega_2^2 s_1 s_2 & \omega_1\omega_2(c_1-c_2) & -\omega_1 c_1 s_2+\omega_2 s_1 c_2 & \omega_1 s_2-\omega_2 s_1 \\ \omega_1\omega_2 s_1 c_2-\omega_2^2 c_1 s_2 & -\omega_1\omega_2 s_1+\omega_2^2 s_2 & \omega_1 s_1 s_2-\omega_2(c_1c_2-1) & \omega_2(c_1-c_2) \\ \omega_1^2 s_1 s_2-\omega_1\omega_2(c_1c_2-1) & -\omega_1\omega_2(c_1-c_2) & \omega_1 c_1 s_2-\omega_2 s_1 c_2 & -\omega_1 s_2+\omega_2 s_1 \\ -\omega_1^2 s_1 c_2+\omega_1\omega_2 c_1 s_2 & \omega_1^2 s_1-\omega_1\omega_2 s_2 & \omega_1(c_1c_2-1)+\omega_2 s_1 s_2 & -\omega_1(c_1-c_2) \end{pmatrix}. \tag{45}
$$

This gives

$$
\ddot{x}_b = (\omega_1^2 c_1,\omega_1^2 s_1,\omega_2^2 c_2,\omega_2^2 s_2)M^{-1}\begin{pmatrix} x_a \\ x_b \\ v_a \\ v_b \end{pmatrix} = \frac{1}{|M|}\begin{pmatrix} \omega_1^3\omega_2(c_1-c_2)+(12) \\ -\omega_1^3\omega_2(c_1c_2-1)+(12)+2\omega_1^2\omega_2^2 s_1 s_2 \\ -\omega_1^3 s_2+\omega_1^2\omega_2 s_1+(12) \\ \omega_1^3 c_1 s_2-\omega_1^2\omega_2 s_1 c_2+(12) \end{pmatrix}^T \begin{pmatrix} x_a \\ x_b \\ v_a \\ v_b \end{pmatrix}, \tag{46}
$$

$$
\ddot{x}_a = (\omega_1^2,0,\omega_2^2,0)M^{-1}\begin{pmatrix} x_a \\ x_b \\ v_a \\ v_b \end{pmatrix} = \frac{1}{|M|}\begin{pmatrix} -\omega_1^3\omega_2(c_1c_2-1)+(12)+2\omega_1^2\omega_2^2 s_1 s_2 \\ \omega_1^3\omega_2(c_1-c_2)+(12) \\ -\omega_1^3 c_1 s_2+\omega_1^2\omega_2 s_1 c_2+(12) \\ \omega_1^3 s_2-\omega_1^2\omega_2 s_1+(12) \end{pmatrix}^T \begin{pmatrix} x_a \\ x_b \\ v_a \\ v_b \end{pmatrix}, \tag{47}
$$

$$
\dddot{x}_b = (\omega_1^3 s_1,\omega_1^3 c_1,\omega_2^3 s_2,\omega_2^3 c_2)M^{-1}\begin{pmatrix} x_a \\ x_b \\ v_a \\ v_b \end{pmatrix} = \frac{1}{|M|}\begin{pmatrix} \omega_1^4\omega_2 s_1+(12) \\ -\omega_1^4\omega_2 s_1 c_2+\omega_1^3\omega_2^2 c_1 s_2+(12) \\ \omega_1^3\omega_2(c_1-c_2)+(12) \\ \omega_1^4 s_1 s_2-\omega_1^3\omega_2(c_1c_2-1)+(12) \end{pmatrix}^T \begin{pmatrix} x_a \\ x_b \\ v_a \\ v_b \end{pmatrix}, \tag{48}
$$

and, upon inserting this into (18), the classical action

$$
\begin{aligned}
\mathscr{A}_{\mathrm{cl,sf}} = (1/2|M|)\{&(\omega_1^2-\omega_2^2)[(\omega_1 c_1 s_2-\omega_2 s_1 c_2)(v_b^2+v_a^2)-2(\omega_1 s_2-\omega_2 s_1)v_b v_a] \\
&-2\omega_1\omega_2[(\omega_1^2+\omega_2^2)(c_1c_2-1)-2\omega_1\omega_2 s_1 s_2](v_b x_b-v_a x_a)+2\omega_1\omega_2(\omega_1^2-\omega_2^2)(c_1-c_2)(v_b x_a-v_a x_b) \\
&+\omega_1\omega_2(\omega_1^2-\omega_2^2)(\omega_1 s_1 c_2-\omega_2 c_1 s_2)(x_b^2+x_a^2)-2\omega_1\omega_2(\omega_1^2-\omega_2^2)(\omega_1 s_1-\omega_2 s_2)x_b x_a\} .
\end{aligned} \tag{49}
$$

In the absence of external currents, this can be inserted into Eq. (39) giving the desired probability distribution. Before we go on to calculating the full $j\neq 0$ contributions, it is useful to first study a few properties of the $j=0$ result.

IV. THE PARTITION FUNCTION AT $j=0$

The quantum statistical partition function of the $j=0$ system is obtained by setting $x_b = x_a \equiv x$, $x_b = v_a \equiv v$, in which case the classical action becomes

$$
\mathscr{A}_{\mathrm{cl,sf}} = ax^2 + bv^2 , \tag{50}
$$

with

$$
\begin{aligned}
a &= (1/|M|)(\omega_1^2-\omega_2^2)(\omega_1(c_1-1)s_2-\omega_2(c_2-1)s_1) , \\
b &= (1/|M|)(\omega_1^2-\omega_2^2)\omega_1\omega_2 \\
&\quad\times(\omega_1(c_2-1)s_1-\omega_2(c_1-1)s_2) ,
\end{aligned} \tag{51}
$$

and forming the trace

$$
Z=\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dv (xv\tau_b|xv\tau_a) . \tag{52}
$$

This yields

$$
\begin{aligned}
Z &= F(\beta)\frac{\pi}{\sqrt{ab}} \\
&= F(\beta)\frac{\pi\sqrt{|M|}}{\sqrt{(\omega_1^2-\omega_2^2)^2\omega_1\omega_2(c_1-1)(c_2-1)}}
\end{aligned}
$$

$$= \frac{1}{2} \frac{1}{\sqrt{(c_1 - 1)(c_2 - 1)}}$$

$$= \frac{1}{2\sinh(\omega_1\beta/2)} \frac{1}{2\sinh(\omega_2\beta/2)}. \tag{53}$$

The result factorizes into the partition functions of the two harmonic oscillators contained in the system. This could also have been obtained directly from (4) by summing over all *periodic* paths, which would have given

$$Z = \prod_{m=0,\pm 2,\pm 4} \frac{1}{[(\epsilon^2\Omega_m^2 + \epsilon^2\omega_1^2)(\epsilon^2\Omega_m^2 + \epsilon^2\omega_2^2)]^{1/2}}$$

$$= \frac{1}{2\sinh(\omega_1\beta/2)} \frac{1}{2\sinh(\omega_2\beta/2)}. \tag{54}$$

In this product, the integer m runs through all even numbers, positive as well as negative, since periodic paths have the Fourier expansion

$$x(\tau) = \frac{1}{\sqrt{\beta}} \sum_{m=0,\pm 2,\pm 4} (e^{i\nu_m\tau} x_m + \text{c.c.}), \tag{55}$$

with $x_m = x^*_{-m}$.

V. LIMITING CASES

Let us check our result (39) at $j=0$ by looking at a couple of limiting cases that have been solved before. Taking $\omega_2 = 0$, $\omega_1 = \omega$, the Hamiltonian (9) reduces to that of a harmonic oscillator in the variable v with an external linear potential ipv. The integral over $\mathscr{D}x$ in (11) forces $p(\tau)$ to be a constant (via the canonical term $\exp \int_{\tau_a}^{\tau_b} d\tau\, ipx$ in the integrand) and the path integral (11) can be written as the Fourier transform

$$(x_b v_b \tau_b | x_a v_a \tau_a) = \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{-ip(x_b - x_a)} (v_b \tau_b | v_a \tau_a)_p \tag{56}$$

of the following probability distribution:

$$(v_b \tau_b | v_a \tau_a)_p \equiv \int \mathscr{D}v(\tau) \times \exp\left[-\int_{\tau_a}^{\tau_b} d\tau \left(\frac{\dot{v}^2}{2} + \frac{\omega^2}{2} v^2 + ipv \right) \right]. \tag{57}$$

This path integral is well known. It is obtained by a simple shift of the standard oscillator expression[10]

$$(v_b \tau_b | v_a \tau_a) = \int \mathscr{D}v(\tau) \exp\left[-\int_{\tau_a}^{\tau_b} d\tau \left(\frac{\dot{v}^2}{2} + \frac{\omega^2}{2} v^2 \right) \right]$$

$$= \sqrt{\frac{\omega}{2\pi \sinh \omega\beta}} \exp\left\{ -\frac{\omega}{2\sinh \omega\beta} \times [\cosh \omega\beta (v_b^2 + v_a^2) - 2 v_b v_a] \right\}, \tag{58}$$

with $s \equiv \sinh \omega\beta$, $c \equiv \cosh \omega\beta$, namely,

$$(v_b \tau_b | v_a \tau_a)_p = \sqrt{\frac{\omega}{2\pi s}} \exp\left\{ -\frac{\omega}{2s} [c(v_b^2 + v_a^2) - 2 v_b v_a] - ip \frac{c-1}{\omega s}(v_b + v_a) - \frac{p^2}{2\omega^2}\left(\beta - 2\frac{c-1}{\omega s}\right) \right\}, \tag{59}$$

whereupon (56) becomes

$$(x_b v_b \tau_b | x_a v_a \tau_a) = \sqrt{\frac{\omega}{2\pi s}} \frac{1}{\sqrt{2\pi\beta}} \frac{\omega}{\sqrt{1-\rho}} \times \exp\left\{ -\frac{\omega}{2s}[c(v_b^2 + v_a^2) - 2v_b v_a] - \frac{\omega^2}{2\beta} \frac{1}{1-\rho} \left[x_b - x_a - \frac{\beta}{2}\rho(v_b + v_a) \right]^2 \right\}, \tag{60}$$

with

$$\rho \equiv 2\frac{c-1}{\omega\beta s} = \frac{\tanh(\omega\beta/2)}{(\omega\beta/2)}. \tag{61}$$

Taking the trace of (60) with respect to the velocity variable, the distribution acquires the simple form

$$(x_b \tau_b | x_a \tau_a) = \frac{1}{2\sinh(\omega\beta/2)} \times \frac{1}{\sqrt{2\pi\beta/\omega}} e^{-(\omega/2\beta)(x_b - x_a)^2}. \tag{62}$$

The prefactor $1/[2\sinh(\omega\beta/2)]$ accounts for the partition function of the harmonic oscillator associated with the variable v. Apart from that, expression (62) shows the standard mean-square end-to-end distance of a random chain, namely $\langle (x_b - x_a)^2 \rangle = \beta/\omega$. It has the same linear behavior in β as in the absence of the stiffness term $\ddot{x}^2$.

It is easy to verify that our general expression (39) with (40) and (49) reduces to (60) for $\omega_2 = 0$. Indeed, then

$$|M| \to \omega\omega_2[\omega\beta s - 2(c-1)] = \omega\omega_2\omega\beta s(1-\rho)$$

and

$$\mathscr{A}_{\text{cl,sf}} \to \frac{1}{2\beta s(1-\rho)} \times \{(\omega\beta c - s)(v_b^2 + v_a^2) - 2(\omega\beta - s) v_b v_a - 2\omega(c-1)(v_b x_b - v_a x_a - v_b x_a + v_a x_b) + \omega^2 s (x_b - x_a)^2\}, \tag{63}$$

giving the correct exponent as well as the fluctuation factor in (60).

If we also let $\omega \to 0$, then $1 - \rho \to \frac{1}{12}\omega^2\beta^2$ and the action reduces to the simple expression

$$\mathscr{A}_{\text{cl,sf}} \to (1/2\beta)(v_b - v_a)^2 + (6/\beta^3) \times [x_b - x_a - (\beta/2)(v_b + v_a)]^2, \tag{64}$$

which could have been found directly from the classical orbit

$$x - x_a + v_a\tau + x_2\tau^2 + x_3\tau^3 \tag{65}$$

after adjusting the parameters x_2, x_3 to the initial and final values

$$x_3 = -(2/\beta^3)[x_b - x_a - (\beta/2)(v_b + v_a)],$$

$$x_2 = (3/\beta^2)(x_b - x_a) + (1/2\beta)(v_b - v_a) - (3/2\beta)(v_b + v_a). \tag{66}$$

The transition probability becomes

$$(x_b v_b \tau_b | x_a v_a \tau_a) = (\sqrt{3}/\pi\beta^2) e^{-\mathscr{A}_{\rm cl,sf}} . \quad (67)$$

Another useful limit is that of $\omega_2 \to \omega_1 \equiv \omega$. Setting $\omega_1 = \omega + \epsilon, \omega_2 = \omega - \epsilon$, the determinant becomes

$$|M| \to 4\epsilon^2 (s^2 - \omega^2\beta^2) . \quad (68)$$

Inserting the limit

$$\begin{aligned} c_{\substack{1\\2}} &= c(1 \pm \epsilon\beta \tanh(\omega\beta)) + (\epsilon^2/2)\beta^2 + \cdots , \\ s_{\substack{1\\2}} &= s(1 \pm \epsilon\beta \coth(\omega\beta)) + (\epsilon^2/2)\beta^2 + \cdots , \end{aligned} \quad (69)$$

and using $\coth(\omega\beta) - \tanh(\omega\beta) = (1/sc)$ we find the classical action

$$\begin{aligned} \mathscr{A}_{\rm cl,1} = \frac{\omega}{s^2 - \omega^2\beta^2} \{ & (sc - \omega\beta)(v_b^2 + v_a^2) \\ & - 2(s - c\omega\beta) v_b v_a \\ & - \omega(s^2 + \omega^2\beta^2)(v_b x_b - v_a x_a) \\ & + 2\omega s (v_b x_a - v_a x_b) + \omega^2 (sc + \omega\beta)(x_b^2 + v_a^2) \\ & - 2\omega^2 (s + c\omega\beta) v_b x_a \} , \end{aligned} \quad (70)$$

and hence

$$(x_b v_b \tau_b | x_a v_a \tau_a) = \frac{1}{\pi} \frac{\omega^2}{\sqrt{s^2 - \omega^2\beta^2}} e^{-\mathscr{A}_{\rm cl,sf}} . \quad (71)$$

In the limit $\omega \to 0$, this reduces again to (67) with (64), as it should.

VI. THE SOURCE TERMS

The source appears in (18) and the last term in (20). First we calculate (18):

$$\mathscr{A}_{\rm cl,source} = - \int_{\tau_a}^{\tau_b} d\tau\, x_{\rm cl}(\tau) j(\tau) , \quad (72)$$

where $x_{\rm cl}(\tau)$ is given by (8) with $A, B, C,$ and D expressed in terms of $x_a v_a x_b v_b$ via the matrix M^{-1} of Eq. (42). Hence

$$x_{\rm cl}(\tau) = \frac{1}{|M|} R \begin{pmatrix} \cosh \omega_1(\tau - \tau_a) \\ \sinh \omega_1(\tau - \tau_a) \\ \cosh \omega_2(\tau - \tau_a) \\ \sinh \omega_2(\tau - \tau_a) \end{pmatrix} . \quad (73)$$

In the ordinary harmonic oscillator, the usual way of giving the classical solution is more symmetrical in τ_a and τ_b

$$\begin{aligned} x_{\rm cl} = (1/\sinh \omega\beta) & (x_b \sinh \omega(\tau - \tau_a) \\ & + x_a \sinh \omega(\tau_b - \tau)) . \end{aligned} \quad (74)$$

It displays directly the interpolation between x_a and x_b. We can also bring (73) to such a form, which, however, is now much more involved. By expanding $x_{\rm cl}$ into the four solutions

$$\begin{aligned} f_a(\tau) &= \omega_2 \sinh \omega_1(\tau - \tau_a) - \omega_1 \sinh \omega_2(\tau - \tau_a) , \\ f_b(\tau) &= \omega_2 \sinh \omega_1(\tau_b - \tau) - \omega_1 \sinh \omega_2(\tau_b - \tau) , \\ g_a(\tau) &= \cosh \omega_1(\tau - \tau_a) - \cosh \omega_2(\tau - \tau_a) , \\ g_b(\tau) &= \cosh \omega_1(\tau_b - \tau) - \cosh \omega_2(\tau_b - \tau) , \end{aligned} \quad (75)$$

which have the boundary properties

$$\begin{aligned} & f_a(\tau_a) = 0, \quad f'_a(\tau_a) = 0 , \\ & f_b(\tau_b) = 0, \quad f'_b(\tau_b) = 0 , \\ & g_a(\tau_a) = 0, \quad g'_b(\tau_b) = 0 , \\ & g_b(\tau_b) = 0, \quad g'_b(\tau_b) = 0 , \end{aligned} \quad (76)$$

it is straightforward to form the linear combination, with the correct initial and final values

$$\begin{aligned} x_{\rm cl}(\tau) = & - (1/|M|) \\ & \times \{ [x_b(\omega_1 s_1 - \omega_2 s_2) - v_b(c_1 - c_2)] f_a(\tau) \\ & + [x_a \omega_1 s_1 - \omega_2 s_2) - v_a(c_1 - c_2)] f_b(\tau) \\ & - [x_b \omega_1\omega_2(c_1 - c_2) - v_b(\omega_2 s_1 - \omega_1 s_2)] g_a(\tau) \\ & - [x_a \omega_1\omega_2(c_1 - c_2) \\ & - v_a(\omega_2 s_1 - \omega_1 s_2)] g_b(\tau) \} . \end{aligned} \quad (77)$$

This may be more useful than (73), for some purposes.

Let us now turn to the fluctuation part of the external source term in (39). Notice that it is sufficient to calculate the odd and even sums

$$\begin{aligned} \hat{G}_o(\tau,\tau') &= \frac{2}{\beta} \sum_{m=1,3,5,\ldots} \frac{\sin \nu_m(\tau - \tau_a) \sin \nu_m(\tau' - \tau_a)}{(\Omega_m^2 + \omega_1^2)(\Omega_m^2 + \omega_2^2)} , \\ \hat{G}_e(\tau,\tau') &= \frac{2}{\beta} \sum_{m=2,4,6,\ldots} \frac{\sin \nu_m(\tau - \tau_a) \sin \nu_m(\tau' - \tau_a)}{(\Omega_m^2 + \omega_1^2)(\Omega_m^2 + \omega_2^2)} . \end{aligned} \quad (78)$$

Then

$$\hat{G}(\tau,\tau') = \hat{G}_o(\tau,\tau') + \hat{G}_e(\tau,\tau') \quad (79)$$

and the functions $h_o(\tau)$, $h_e(\tau')$ are simply found from the derivatives [compare (37a)]

$$h_{\substack{o\\e}}(\tau) = \lim_{\tau' \to \tau_a} \frac{\partial}{\partial \tau'} \hat{G}_{\substack{o\\e}}(\tau,\tau') . \quad (80)$$

In the sums (78) we can replace Ω_m^2 by ν_m^2, due to their fast convergence, and write

$$\hat{G}_o(\tau,\tau') = [1/(\omega_2^2 - \omega_1^2)](G_o^{\omega_1}(\tau,\tau') - G_o^{\omega_2}(\tau,\tau')) , \quad (81)$$

$$\hat{G}_e(\tau,\tau') = [1/(\omega_2^2 - \omega_1^2)](G_e^{\omega_1}(\tau,\tau') - G_e^{\omega_2}(\tau,\tau')) , \quad (82)$$

where

$$\begin{aligned} G_{\substack{o\\e}}^{\omega}(\tau,\tau') &= \frac{2}{\beta} \sum_{\substack{m=1,3,5,\ldots\\2,4,6}} \frac{\sin \nu_m(\tau - \tau_a) \sin \nu_m(\tau' - \tau_a)}{(\nu_m^2 + \omega^2)} \\ &= \pm \frac{2}{\beta} \sum_{\substack{m=1,3,5\\2,4,6}} \frac{\sin \nu_m(\tau_b - \tau) \sin \nu_m(\tau' - \tau_a)}{(\nu_m^2 + \omega^2)} \end{aligned} \quad (83)$$

are the odd and even frequency parts of the correlation function of the ordinary harmonic oscillator. They, in turn, are simply obtained from the standard boson and fermion correlation functions

$$\begin{aligned} G_{\rm B}(\tau) &= \frac{1}{\beta} \sum_{m=0,\pm 2,\pm 4,\ldots} e^{-i\nu_m\tau} \frac{1}{\nu_m^2 + \omega^2} \\ &= \frac{1}{2\omega} \frac{\cosh \omega[\tau - (\beta/2)]}{\sinh(\omega\beta/2)} , \quad \tau \in (0,\beta) , \end{aligned} \quad (84)$$

$$G_{\rm F}(\tau) = \frac{1}{\beta} \sum_{m=\pm 1,\pm 3,\pm 4} e^{i\nu_m\tau} \frac{1}{\nu_m^2 + \omega^2}$$

$$= -\frac{1}{2\omega}\frac{\sinh\omega[\tau-(\beta/2)]}{\cosh(\omega\beta/2)}, \quad \tau\in(0,\beta). \tag{85}$$

For $\tau=0$ these coincide with the sums appearing in Eqs. (30) and (31), as they should.

Notice that the right-hand side is valid only for $\tau\in(0,\beta)$. Outside this interval, the functions have to be continued periodically or antiperiodically for G_e or G_o. An explicit representation which shows this property is obtained by rewriting

$$G_B(\tau) = \sum_{l=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{d\nu}{2\pi}e^{-i\nu(\tau+l\beta)}\frac{1}{\nu^2+\omega^2}, \tag{86}$$

$$G_F(\tau) = \sum_{l=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{d\nu}{2\pi}e^{-i\nu(\tau+l\beta)}e^{\pi i l}\frac{1}{\nu^2+\omega^2}, \tag{87}$$

where the sums over all integer numbers l squeeze the ν integrals into the appropriate sums (84) and (85). Performing the integrals over ν gives

$$G_B(\tau) = \frac{1}{2\omega}\sum_l(\theta(\tau+l\beta)e^{-\omega(\tau+l\beta)} + \theta(\tau-l\beta)e^{\omega(\tau+l\beta)}), \tag{88}$$

$$G_F(\tau) = \frac{1}{2\omega}\sum_l(-)^l(\theta(\tau+l\beta)e^{-\omega(\tau+l\beta)} + \theta(\tau-l\beta)e^{\omega(\tau+l\beta)}). \tag{89}$$

For $\tau\in(0,\beta)$, the sums split into $l=0,1,2,\ldots$ and $l=-1,-2,-3,\ldots$ and can be performed to yield the results (84) and (85). For $\tau\in(\beta,2\beta)$, however, these have to be replaced by

$$G_B(\tau) = \frac{1}{2\omega}\frac{\cosh\omega[\tau-(3\beta/2)]}{\sinh(\omega\beta/2)} \tag{90}$$

$$[\tau\in(\beta,2\beta)].$$

$$G_F(\tau) = \frac{1}{2\omega}\frac{\sinh\omega[\tau-(3\beta/2)]}{\cosh(\omega\beta/2)} \tag{91}$$

When forming the appropriate combinations of these correlation functions in (83) and (84), we have to distinguish the cases $\tau+\tau'<\tau_a+\tau_b$, $\tau+\tau'>\tau_a+\tau_b$. In the first case we find

$$G_e^\omega(\tau,\tau') = -\frac{1}{2\omega\sinh(\omega\beta/2)} \times\sinh\omega\left(\tau-\frac{\tau_b+\tau_a}{2}\right)\sinh\omega(\tau'-\tau_a),$$

$$\text{for } \tau>\tau'\in(\tau_a,\tau_b), \quad \tau+\tau'<\tau_a+\tau_b, \tag{92}$$

$$G_o^\omega(\tau,\tau') = \frac{1}{2\omega\cosh(\omega\beta/2)} \times\cosh\omega\left(\tau-\frac{\tau_b+\tau_a}{2}\right)\sinh\omega(\tau'-\tau_a). \tag{93}$$

In the second case

$$G_e^\omega(\tau,\tau') = \frac{1}{2\omega\sinh(\omega\beta/2)} \times\sinh\omega(\tau_b-\tau)\sinh\omega\left(\tau'-\frac{\tau_a+\tau_b}{2}\right),$$

$$\text{for } \tau>\tau'\in(\tau_a,\tau_b), \quad \tau+\tau'>\tau_a+\tau_b, \tag{94}$$

$$G_o^\omega(\tau,\tau') = -\frac{1}{2\omega\cosh(\omega\beta/2)} \times\sinh\omega(\tau_b-\tau)\cosh\omega\left(\tau'-\frac{\tau_a+\tau_b}{2}\right). \tag{95}$$

As a check we add the even and odd results and find

$$G^\omega(\tau,\tau') = (\omega\sinh\omega\beta)^{-1} \times\sinh\omega(\tau_b-\tau)\sinh\omega(\tau'-\tau_a), \quad \tau>\tau', \tag{96}$$

in either case, which is the correct correlation function

$$G^\omega(\tau,\tau') = \langle\delta x(\tau)\delta x(\tau')\rangle|_{\text{oscill}} = \frac{2}{\beta}\sum_{m=1,2,\ldots}\frac{\sin\nu_m(\tau-\tau_a)\sin\nu_m(\tau'-\tau_a)}{\nu_m^2+\omega^2} \tag{97}$$

appearing in the path integral of the ordinary harmonic oscillator.[11] Inserting (94)–(97) into (82) we find the odd and even parts of the correlation function $\hat{G}(\tau,\tau')$:

$$\hat{G}_e(\tau,\tau') = -\frac{1}{(\omega_2^2-\omega_1^2)}\left(\frac{1}{2\omega_1 s_1}\sinh\omega_1\left(\tau-\frac{\tau_b+\tau_a}{2}\right) \times\sinh\omega_1(\tau'-\tau_a)-(12)\right),$$

$$\text{for } \tau>\tau'\in(\tau_a,\tau_b), \quad \tau+\tau'<\tau_a+\tau_b, \tag{98}$$

$$\hat{G}_o(\tau,\tau') = \frac{1}{(\omega_2^2-\omega_1^2)}\left(\frac{1}{2\omega_1 c_1}\cosh\omega_1\left(\tau-\frac{\tau_b+\tau_a}{2}\right) \times\sinh\omega_1(\tau'-\tau_a)-(12)\right),$$

and

$$\hat{G}_e(\tau,\tau') = \frac{1}{\omega_2^2-\omega_1^2}\left(\frac{1}{2\omega_1 s_1}\sinh\omega_1(\tau_b-\tau) \times\sinh\omega_1\left(\tau'-\frac{\tau_a+\tau_b}{2}\right)-(12)\right),$$

$$\text{for } \tau>\tau'\in(\tau_b,\tau_a), \quad \tau+\tau'>\tau_a+\tau_b, \tag{99}$$

$$\hat{G}_o(\tau,\tau') = -\frac{1}{\omega_2^2-\omega_1^2}\left(\frac{1}{2\omega_1 c_1}\sinh\omega_1(\tau_b-\tau) \times\cosh\omega_1\left(\tau'-\frac{\tau_a+\tau_b}{2}\right)-(12)\right).$$

Adding up the even and odd parts we find, according to formula (79),

$$\hat{G}(\tau,\tau') = -\frac{1}{\omega_1^2-\omega_2^2}\left[\frac{1}{\omega_1 s_1}\sinh\omega_1(\tau_b-\tau) \times\sinh\omega_1(\tau'-\tau_a)-(12)\right] \tag{100}$$

in either case. This is the first part of the correlation function $\langle\delta x(\tau)\delta x(\tau')\rangle$ in Eq. (35).

Since we have treated the even and odd parts separately, it is now easy to find other pieces $h_e(\tau)$, $h_e(\tau')$ from the limits (80)

$$h_e(\tau)=\frac{1}{\omega_1^2-\omega_2^2}\left[\frac{1}{2\sinh(\omega_1\beta/2)}\times\sinh\left(\omega_1\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\right)-(12)\right],$$

$$h_o(\tau)=-\frac{1}{\omega_1^2-\omega_2^2}\left[\frac{1}{2\cosh(\omega_1\beta/2)}\times\cosh\left(\omega_1\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\right)-(12)\right]. \qquad (101)$$

As a cross check, we may form

$$\lim_{\tau\to\tau_a}\frac{\partial}{\partial\tau}h_g(\tau),$$

which gives D_e, D_o, as it should [compare with (37b) and (31)].

Combining (103) and (104) and using D_e, D_o we obtain the complete correlation function of the fluctuations [recall (35)]:

$$\begin{aligned}\langle\delta x(\tau)\delta x(\tau')\rangle=G(\tau,\tau')=&-\frac{1}{\omega_1^2-\omega_2^2}\left\{\left(\frac{1}{\omega_1 s_1}\sinh\omega_1(\tau_b-\tau)\sinh\omega_1(\tau'-\tau_a)-(12)\right)\right.\\&+\frac{1}{2}\frac{1}{(\omega_1\coth(\omega_1\beta/2)-(12))}\left(\frac{1}{\sinh(\omega_1\beta/2)}\sinh\omega_1\left(\tau-\frac{\tau_a+\tau_b}{2}\right)-(12)\right)\\&\times\left(\frac{1}{\sinh(\omega_1\beta/2)}\sinh\omega_1\left(\tau'-\frac{\tau_a+\tau_b}{2}\right)-(12)\right)\\&+\frac{1}{2}\frac{1}{(\omega_1\tanh(\omega_1\beta/2)-(12))}\left(\frac{1}{\cosh(\omega_1\beta/2)}\cosh\omega_1\left(\tau-\frac{\tau_a+\tau_b}{2}\right)-(12)\right)\\&\left.\times\left(\frac{1}{\cosh(\omega_1\beta/2)}\cosh\omega_1\left(\tau'-\frac{\tau_a+\tau_b}{2}\right)-(12)\right)\right\}.\end{aligned} \qquad (102)$$

As a final check we verify once more that this Green's function vanishes at the end points together with its time derivatives. This completes the calculation of the probability distribution $(x_bv_b\tau_b|x_av_a\tau_a)$. The result is Eq. (39) with the prefactor (40), the classical surface term (49), the classical source term (72) with $x_{\rm cl}(\tau)$ given in (73) or (77), and the fluctuation part of the source term given by the correlation function (102).

VII. LIMITING FORMS OF SOURCE TERMS

For completeness, let us perform the limits $\omega_2\to0$, $\omega_2\to0$, $\omega_1\to0$, and $\omega_1\to\omega_2$ on the source terms. For $\omega_2=0$, $\omega_1=\omega$ the classical solution (77) reduces to

$$\begin{aligned}x_{\rm cl}=&-\frac{1}{\beta\omega(1-\rho)}\\&\times\left\{\left[\left(x_b-\frac{\rho}{2}\beta v_b\right)(\sinh\omega(\tau-\tau_a)-\omega(\tau-\tau_a))\right.\right.\\&\left.+\left(x_a-\frac{\rho}{2}\beta v_a\right)(\sinh\omega(\tau_b-\tau)-\omega(\tau_b-\tau))\right]\\&-\left[\left(x_b\frac{\rho}{2}\beta\omega-\frac{v_b}{\omega}\right)\left(1-\frac{\omega\beta}{s}\right)(\cosh\omega(\tau-\tau_a)-1)\right.\\&+\left(x_a\frac{\rho}{2}\beta\omega-\frac{v_a}{\omega}\right)\left(1-\frac{\omega\beta}{s}\right)\\&\left.\left.\times(\cosh\omega(\tau_b-\tau)-1)\right]\right\}.\end{aligned} \qquad (103)$$

If also $\omega\to0$,

$$\begin{aligned}x_{\rm cl}=&-12\left\{\left(x_b-\frac{\beta}{2}v_b\right)\frac{(\tau-\tau_a)^3}{6\beta^3}-(ba)\right.\\&\left.-\left(x_b-\frac{\beta}{6}v_b\right)\frac{(\tau-\tau_a)^2}{4\beta^2}+(ba)\right\}.\end{aligned} \qquad (104)$$

In the limit $\omega_1-\omega_2=2\epsilon\to0$, the functions (75) tend towards

$$f_a(\tau)\to2\epsilon(\omega(\tau-\tau_a)\cosh\omega(\tau-\tau_a)-\sinh\omega(\tau-\tau_a)),$$

$$g_a(\tau)\to2\epsilon(\tau-\tau_a)\sinh\omega(\tau-\tau_a),$$

with analogous limits for $f_b(\tau)$, $g_b(\tau)$, and the classical solutions become

$$\begin{aligned}x_{\rm cl}(\tau)=&-\frac{1}{s^2-\omega^2\beta^2}\left\{\left(x_b\left(s+\omega\beta\frac{c}{s}\right)-v_bs\right)(\omega(\tau-\tau_a)\right.\\&\times\cosh(\tau-\tau_a)-\sinh\omega(\tau-\tau_a))-(ab)\\&-x_b\left(\omega\beta s+\frac{v_b}{\omega}\left(s-\omega\beta\frac{c}{s}\right)\right)\omega(\tau-\tau_a)\\&\left.\times\sinh\omega(\tau-\tau_a)+(ab)\right\}.\end{aligned} \qquad (105)$$

The fluctuation part of the source contribution has the following limits: for $\omega_2\to0$, $\omega_1=\omega$,

$$\hat G(\tau,\tau')=-\frac{1}{\omega^2}\left[\frac{1}{\omega s}\sinh\omega(\tau_b-\tau)\sinh\omega(\tau'-\tau_b)-\frac{(\tau_b-\tau)(\tau'-\tau_a)}{\beta}\right]; \qquad (106)$$

for $\omega_2\to0$, $\omega_1\to0$,

$$\hat G(\tau,\tau')\to-\frac{1}{6\beta}(\tau_b-\tau)(\tau'-\tau_a)\times[(\tau_b-\tau)^2+(\tau'-\tau_a)^2-\beta^2]; \qquad (107)$$

for $\omega_2\to\omega_1=\omega$,

$$\begin{aligned}\hat G(\tau,\tau')\to\frac{1}{2\omega^3s}&\left[\left(1+\omega\beta\frac{c}{s}\right)\sinh\omega(\tau_b-\tau)\sinh\omega(\tau'-\tau_a)\right.\\&-\sinh\omega(\tau_b-\tau)\omega(\tau'-\tau_a)\cosh\omega(\tau'-\tau_a)\end{aligned}$$

$$-\omega(\tau_b-\tau)\cosh\omega(\tau_b-\tau)\sinh\omega(\tau'-\tau_a)]\,; \tag{108}$$

with the latter reducing properly to (107) in the limit $\omega\to 0$.

The functions $h_e(\tau)$, $h_o(\tau)$ become, for $\omega_2\to 0$, $\omega_1=\omega$,

$$h_e(\tau)=\frac{1}{\omega^2}\left(\frac{\sinh\omega(\tau-(\tau_a+\tau_b)/2)}{2\sinh(\omega\beta/2)}-\frac{\tau-(\tau_a+\tau_b)/2}{\omega\beta}\right), \tag{109}$$

$$h_o(\tau)=-\frac{1}{\omega^2}\left(\frac{\cosh\omega(\tau-(\tau_a+\tau_b)/2)}{2\cosh(\omega\beta/2)}-\frac{1}{2}\right); \tag{110}$$

for $\omega_2\to 0$, $\omega_1\to 0$,

$$h_e(\tau)=\frac{1}{6\beta}\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\left[\left(\tau-\frac{\tau_a+\tau_b}{2}\right)^2-\frac{1}{4}\beta^2\right], \tag{111}$$

$$h_o(\tau)=-\frac{1}{4}\left[\left(\tau-\frac{\tau_a+\tau_b}{2}\right)^2-\frac{1}{4}\beta^2\right]; \tag{112}$$

and for $\omega_2\to\omega_1=\omega$,

$$h_e(\tau)=\frac{1}{4\omega^2\sinh(\omega\beta/2)}\times\left(\omega\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\cosh\omega\left(\tau-\frac{\tau_a+\tau_b}{2}\right)-\frac{\omega\beta}{2}\coth\frac{\omega\beta}{2}\sinh\omega\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\right), \tag{113}$$

$$h_o(\tau)=-\frac{1}{4\omega^2\cosh(\omega\beta/2)}\times\left(\omega\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\sinh\omega\left(\tau-\frac{\tau_a+\tau_b}{2}\right)-\frac{\omega\beta}{2}\tanh\frac{\omega\beta}{2}\cosh\omega\left(\tau-\frac{\tau_a+\tau_b}{2}\right)\right); \tag{114}$$

and the quantities D_e, D_o, for $\omega_2\to 0$, $\omega_1=\omega$,

$$D_e=\frac{1}{\beta\omega^2}\left(\frac{\omega\beta}{2}\coth\frac{\omega\beta}{2}-1\right), \tag{115}$$

$$D_o=\frac{1}{\beta\omega^2}\frac{\omega\beta}{2}\tanh\frac{\omega\beta}{2}; \tag{116}$$

for $\omega_2\to 0$, $\omega_1\to 0$,

$$D_e=\tfrac{1}{12}\beta, \tag{117}$$

$$D_o=\tfrac{1}{4}\beta; \tag{118}$$

and for $\omega_2\to\omega_1=\omega$,

$$D_e=\frac{\beta}{8}\left(\frac{\coth(\omega\beta/2)}{\omega\beta/2}-\frac{1}{\sinh^2(\omega\beta/2)}\right), \tag{119}$$

$$D_o=\frac{\beta}{8}\left(\frac{\tanh(\omega\beta/2)}{\omega\beta/2}+\frac{1}{\cosh^2(\omega\beta/2)}\right). \tag{120}$$

Combining these $\hat{G}$, h, and D as required by (35) we obtain the limiting terms of the correlation function $G(\tau,\tau')$.

VIII. SECOND QUANTIZATION

Frequently one is not interested in studying the behavior of a single fluctuating-line-like object but wants to consider grand-canonical ensembles of these. It is then convenient to introduce a single fluctuating field whose Feyman diagrams are capable of representing all the different individual line contributions. For the usual random chain with a Lagrangian $(D/2a)\dot{x}^2$ in D dimensions, it is well known how to achieve this goal. For open chains of a given length L the appropriate field is $\psi(\mathbf{x},\tau)$ and has the action[12]

$$\mathscr{A}=\int_0^D d\tau\int d^Dx\{\psi^+(\partial_\tau-\mu)\psi+\psi^+H(-i\,\partial_x)\psi\}, \tag{121}$$

where $H(\mathbf{p})=\mathbf{p}^2/(2D/a)$ is the Hamiltonian and μ is the chemical potential of a chain element. For loops of any length L, with a distribution e^{-m^2L}, the fields $\varphi(x)$ depend only on the spatial variable $\mathbf{x}$ and the action is of the Klein-Gordon type[13]

$$\mathscr{A}=\int d^Dx\,\varphi(\mathbf{x})(H(-i\partial_x)+m^2)\varphi(\mathbf{x})=\frac{1}{2}\int d^Dx[(\partial\varphi(\mathbf{x}))^2+m^2\varphi^2(\mathbf{x})]\,. \tag{122}$$

In the present case where the Lagrangian contains a second time derivative, a second quantization can be achieved by introducing, for open chains of a given length L, a field $\psi(x,v,\tau)$ which depends on position, velocity, and time with an action

$$\mathscr{A}=\int_0^L d\tau\int d^Dx\int d^Dx\times\{\psi^+(\partial_\tau-\mu)\psi+\psi^+H(-i\,\partial_x,\mathbf{x},-\mathrm{i}\,\partial_v,\mathbf{v},\tau)\psi\}, \tag{123}$$

where $H(p,x,p_v,v,\tau)$ is a Hamiltonian of the type (9) in D dimensions. For closed chains of any length one has, similarly, a field $\varphi(\mathbf{x},\mathbf{v})$ and an action

$$\mathscr{A}=\frac{1}{2}\int d^Dx\,d^Dv\,\varphi(\mathbf{x},\mathbf{v})(H(\mathbf{p},\mathbf{x},\mathbf{p}_v,\mathbf{v},\tau)+m^2)\varphi(\mathbf{x},\mathbf{v})\,.$$

IX. CONCLUSION

We have calculated the exact amplitude for fluctuating orbits $x(\tau)$ governed by the general second-gradient Lagrangian (2). The results is given by Eq. (39) with the fluctuation prefactor (40), the classical action (49), the classical source action (72) and (77), and the fluctuation part of the source given by (102).

ACKNOWLEDGMENT

The author is grateful to Professor B. Zimm for a useful discussion and to Professor N. Kroll and Professor J. Kuti for their hospitality at the University of California, San Diego.

This work was supported in part by the Deutsche Forschungsgemeinschaft under Grant No. Kl 256 and by UCSD/DOE Contract No. DEAT-03-81ER40029.

[1]See, for example, A. Pais and G. E. Uhlenbeck, Phys. Rev. 79, 145 (1950); A. O. Barut and G. Mullen, Ann. Phys. (NY) 20, 184, 203 (1962).

[2]R. H. Harris and J. E. Hearst, J. Chem. Phys. 44, 2595 (1966). See also, H.

Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971), p. 52ff or his review article in Pure Appl. Chem. **46**, 135 (1976) and in Annu. Rev. Phys. Chem. **25**, 179 (1974); **35**, 123 (1984).
[3]W. Helfrich, Z. Naturforsch. **30c**, 81 (1975); **28c**, 693 (1974); P. B. Canham, J. Theor. Biol. **26**, 61 (1970).
[4]F. Browicz, Zbl. Med. Wiss. **28**, 625 (1890).
[5]W. Helfrich and R. M. Servuss, Nuovo Cimento **D3**, 137 (1984); W. Janke and H. Kleinert, Phys. Lett. A **117**, 353 (1986) and (in press).
[6]P. G. deGennes and C. Taupin, J. Chem. Phys. **86**, 7794 (1982).
[7]H. Kleinert, Phys. Lett. B **174** (1986); A. M. Polyakov, Nucl. Phys. B **268**, 406 (1986).
[8]See, for example, K. Akama, Y. Chikashige, T. Matsuki, and H. Terazawa, Prog. Theor. Phys. **60**, 868 (1978); K. Akama and H. Terazawa, Gen. Relativ. Gravit. **15**, 201 (1983); S. Adler, Phys. Lett. **95B**, 241 (1980); A. Zee, Phys. Rev. D **23**, 858 (1981). For geophysical deviations from Newton's law, see F. D. Stacey and G. J. Tuck, Nature (London) **292**, 230 (1981); S. C. Holding and G. J. Tuck, Nature (London) **307**, 714 (1984); G. W. Gibbons and B. F. Whiting, Nature (London) **291**, 636 (1981). For higher gradient gravity, see K. Stelle, Gen. Relativ. Gravit. **9**, 353 (1978). The alternative explanation of the effect via 1955 proposal of Lee and Yang by perphoton, recently advocated by E. Fishbach *et al.*, Phys. Rev. Lett. **56**, 3 (1986) seems to have problems since the systematics observed by them from Eötvös' 1922 data is 16 times larger than the geophysical observations.
[9]M. Ostrogradski, Mem. Acad. St. Petersburg, **VI4**, 385 (1850). See also, Whittaker, *Analytical Dynamics* (Cambridge U. P., London, 1932), 4th ed., p. 265.
[10]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[11]K. F. Freed, J. Chem. Phys. **34**, 453 (1971); M. G. Bawendi and K. F. Freed, *ibid.* **83**, 2491 (1985).
[12]K. F. Freed, Adv. Chem. Phys. **22**, 1 (1972).
[13]H. Kleinert, *Gauge Theory of Stresses and Defects* (World Scientific, Singapore, to be published).

Probability density for Bose–Einstein and Fermi–Dirac particles: Slater–Kahn functions

M. D. Kostin
Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544

(Received 21 February 1986; accepted for publication 6 August 1986)

A nonlinear form of the Schrödinger equation is used to derive a partial differential equation for a Slater–Kahn function, which in turn yields the probability density of Bose–Einstein and Fermi–Dirac particles at a given temperature.

I. INTRODUCTION

The Slater sum for a quantum system at equilibrium at temperature T has been defined by the expression[1]

$$S(x,\beta) = \sum_n [\psi_n(x)]^2 \exp(-\beta E_n), \tag{1}$$

where $\psi_n(x)$ is the normalized wave function of state n, E_n is its energy eigenvalue, $\beta = 1/kT$, k is Boltzmann's constant, and the sum extends over all quantum states. The probability density is then given by

$$P(x,\beta) = S(x,\beta)/Z(\beta), \tag{2}$$

where the partition function $Z(\beta)$ is obtained by integrating $S(x,\beta)$ over all positions:

$$Z(\beta) = \int S(x,\beta)dx. \tag{3}$$

Important properties of the Slater sum for a single particle were derived by Uhlenbeck and Gropper.[2] Special attention was given by Kahn[3] to the Slater sum for two Fermi–Dirac or two Bose–Einstein particles,

$$K(x_1,x_2,\beta) = \frac{1}{2}\sum_m \sum_n [\psi_m(x_1)\psi_n(x_2) \pm \psi_n(x_1)\psi_m(x_2)]^2 \exp(-\beta E_m - \beta E_n), \tag{4}$$

where the $+$ sign refers to Bose–Einstein particles and the $-$ sign refers to Fermi–Dirac particles. To evaluate the two-particle Slater–Kahn function directly from (4) is very difficult since one would have to obtain all of the eigenfunctions and eigenvalues and carry out the double summation. The problem is especially difficult when the eigenfunctions ψ_m must be calculated numerically, since the higher eigenfunctions have numerous oscillations, and cannot usually be calculated with accuracy. It is the purpose of this paper to derive a partial differential equation, the solution of which can be used to obtain a Slater–Kahn distribution function for Fermi–Dirac or Bose–Einstein particles.

II. SLATER–KAHN FUNCTIONS

The probability density $P(x,\beta)$ for finding a Fermi–Dirac or Bose–Einstein particle at position x when the system is at reciprocal temperature β is given by

$$P(x,\beta) = K(x,\beta)/Z(\beta), \tag{5}$$

where $K(x,\beta)$ is the Slater–Kahn function obtained by integrating (4) over one of the particles,

$$K(x,\beta) = \int K(x,x_2,\beta)dx_2, \tag{6}$$

and $Z(\beta)$ is the partition function of the system,

$$Z(\beta) = \int K(x,\beta)dx. \tag{7}$$

Substituting (4) into (6) and carrying out the integration gives us

$$K(x,\beta) = \frac{1}{2}\sum_m \sum_n [R_m(x) + R_n(x) \pm 2\delta_{mn}R_m(x)] \exp(-\beta E_m - \beta E_n), \tag{8}$$

where $R_n(x) = [\psi_n(x)]^2$. The probability density $R_n(x)$ satisfies the nonlinear differential equation studied by Bohm,[4]

$$(R_n''/R_n) - [(R_n')^2/2R_n^2] = (4m/\hbar^2)[V(x) - E_n]. \tag{9}$$

It has been shown that for a nonlinear equation of the form (9), it is possible to derive a linear equation by multiplying (9) by R_n^2 and differentiating[5]:

$$(\hbar^2/m)R_n''' + 8[E_n - V(x)]R_N' - 4V'R_n(x) = 0. \tag{10}$$

Next, we introduce the Slater–Kahn function

$$K_1(x,\beta) = \tfrac{1}{2}[K_B(x,\beta) - K_F(x,\beta)], \tag{11}$$

where $K_B(x,\beta)$ is the Slater–Kahn function for Bose–Einstein particles and $K_F(x,\beta)$ is that for Fermi–Dirac particles. Substituting (8) into (11) yields

$$K_1(x,\beta) = \sum_n R_n(x) \exp(-2\beta E_n). \tag{12}$$

We multiply (10) by $\exp(-2\beta E_n)$ and sum over all states. Noting that

$$\frac{\partial K_1}{\partial x} = \sum_n R_n'(x) \exp(-2\beta E_n), \tag{13}$$

$$\frac{\partial^2 K_1}{\partial \beta\, \partial x} = -2\sum_n R_n'(x) \exp(-2\beta E_n)E_n, \tag{14}$$

$$\frac{\partial^3 K_1}{\partial x^3} = \sum_n R_n'''(x) \exp(-2\beta E_n), \tag{15}$$

we obtain

 0022-2488/86/123014-02$02.50

$$\frac{\hbar^2}{4m}\frac{\partial^3 K_1}{\partial x^3} = \frac{\partial^2 K_1}{\partial \beta\,\partial x} + 2V\frac{\partial K_1}{\partial x} + V' K_1(x,\beta). \tag{16}$$

Both $K_B(x,\beta)$ and $K_F(x,\beta)$ can be derived from $K_1(x,\beta)$. To do this, we introduce the Slater–Kahn function $K_2(x,\beta)$:

$$K_2(x,\beta) = \tfrac{1}{2}[K_B(x,\beta) + K_F(x,\beta)]\,. \tag{17}$$

Substituting (8) into (17) gives us

$$K_2(x,\beta) = \left[\sum_m \exp(-\beta E_m)\right]\sum_n R_n(x)\exp(-\beta E_n). \tag{18}$$

The integral over $K_1(x,\beta)$ with respect to position will be denoted by $Y_1(\beta)$:

$$Y_1(\beta) = \int K_1(x,\beta)dx. \tag{19}$$

Substituting (12) into (19) and noting that wave functions $\psi_n(x)$ are normalized so that

$$\int R_n(x)dx = 1, \tag{20}$$

we have

$$Y_1(\beta) = \sum_n \exp(-2\beta E_n). \tag{21}$$

Combining (12), (18), and (21) gives us

$$K_2(x,\beta) = Y_1(\beta/2)K_1(x,\beta/2). \tag{22}$$

Thus, both $K_B(x,\beta)$ and $K_F(x,\beta)$ can be obtained from $K_1(x,\beta)$:

$$K_B(x,\beta) = K_2(x,\beta) + K_1(x,\beta), \tag{23}$$

$$K_F(x,\beta) = K_2(x,\beta) - K_1(x,\beta). \tag{24}$$

The corresponding probability densities are given by

$$P_B(x,\beta) = K_B(x,\beta)/Z_B(\beta), \tag{25}$$

$$P_F(x,\beta) = K_F(x,\beta)/Z_F(\beta), \tag{26}$$

where the partition functions are

$$Z_B(\beta) = \int K_B(x,\beta)dx = Y_2(\beta) + Y_1(\beta), \tag{27}$$

$$Z_F(\beta) = \int K_F(x,\beta)dx = Y_2(\beta) - Y_1(\beta), \tag{28}$$

and $Y_2(\beta)$ is the integral of $K_2(x,\beta)$ with respect to position

$$Y_2(\beta) = \int K_2(x,\beta)dx. \tag{29}$$

Finally, substituting (18) into (29) yields an expression for Y_2 in terms of Y_1:

$$Y_2(\beta) = [Y_1(\beta/2)]^2. \tag{30}$$

Thus, both probability densities $P_B(x,\beta)$ and $P_F(x,\beta)$ can be obtained from the solution of the differential equation (16) for the Slater–Kahn function $K_1(x,\beta)$.

In summary, a linearized equation derived from a nonlinear form of the Schrödinger equation has been used to derive a partial differential equation for a Slater–Kahn distribution function. It has been shown that the solution of this equation can be used to derive the probability density for Bose–Einstein and Fermi–Dirac particles.

ACKNOWLEDGMENT

This research was supported in part by a grant from the National Science Foundation.

[1]J. C. Slater, Phys. Rev. **38**, 237 (1931).

[2]G. E. Uhlenbeck and L. Gropper, Phys. Rev. **41**, 79 (1932).

[3]B. Kahn, *On the Theory of the Equation of State* (North-Holland, Amsterdam, 1938) [reprinted in *Studies in Statistical Mechanics*, edited by J. De Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1965), Vol. III, pp. 277ff].

[4]D. Bohm, Phys. Rev. **85**, 166, 180 (1952).

[5]M. D. Kostin, J. Chem. Phys. **47**, 5450 (1967).

The occupation statistics for variously shaped particles on a rectangular $2 \times N$ lattice space

R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

J. L. Hock
Department of Electrical Engineering and Computer Science, Marquette University, Milwaukee, Wisconsin 53233

(Received 10 September 1985; accepted for publication 14 August 1986)

Utilizing set theoretic arguments, matrices of shift operators have been constructed, the determinants of which yield recursions that describe exactly the occupational degeneracies for three differently shaped particles (trimers, L-particles, T-particles) distributed on a $2 \times N$ rectangular lattice. On the basis of these recursions, the expectation, normalization, dispersion, and continuous representation of the statistics have been calculated and compared for each kind of particle.

I. INTRODUCTION

The statistics that govern the occupation of lattice spaces by complex particles (which occupy more than one lattice site) are different in three ways from comparable statistics for simple particles (which occupy a single lattice site): (1) there is no statistical equivalence between particles and vacant sites; (2) the occupation of a site insures that at least one of the nearest neighbor sites is also occupied; and (3) the possible existence of isolated vacant sites cannot serve as the sole criterion for determining whether or not a site can be occupied.

For these reasons, it has been difficult to formulate, in an analytic way, the statistics (and kinetics) of occupation for complex particles distributed on a lattice. It has long been recognized[1–3] that the kinetics of occupation of even one-dimensional lattices by dimers cannot be handled in a statistical manner. However, considerable progress has been made in handling the statistics of dimers on $2 \times N$ lattices[4] and other two-dimensional spaces.[5–12] This work has culminated in a beautiful paper by Phares *et al.*[13] The occupational degeneracy for dimers on a quasi-three-dimensional space has also been considered.[14]

The occupation statistics for trimer particles (occupying three linearly contiguous sites) distributed on a $3 \times N$ lattice has also been treated.[15]

The purpose of the present paper is to develop and compare the occupational statistics for several kinds of more complicated particles distributed on a $2 \times N$ lattice. The method we employ is not limited to rectangular $2 \times N$ lattices, or even to rectangular lattices, but the magnitudes of the calculations involved can rapidly become formidable for lattice spaces of greater size and dimensionality, as well as for more complex lattices (e.g., hexagonal lattices).

To illustrate the method used to treat more complex particles and to contrast the results obtained, we will treat trimers, L-particles, and T-particles (see Fig. 1). These three kinds of particles are chosen to examine the influence of the degree of rotational freedom and particle size (i.e., number of lattice sites occupied by each kind of particle) on the occupational degeneracy. Thus, the trimers have no rotational freedom but may be moved from one row to another and occupy three lattice sites. The L-particles also occupy three lattice sites, but they have complete rotational freedom. The T-particles occupy four lattice sites and exhibit top-to-bottom rotation only.

Although we limit our discussion to these three kinds of particles, it should be clear that the method used can be extended to particles of any configuration.

On the basis of set theoretic arguments, we now develop recursion relations that describe exactly the occupational degeneracy for three kinds of (indistinguishable) particles distributed on a $2 \times N$ rectangular lattice.

For the purpose of establishing the necessary recursions, we define an $\alpha_j(N)$-space to be a $2 \times N$ rectangular lattice space with j sites deleted from either the lower or upper left-hand corner (see Fig. 2). We first use these

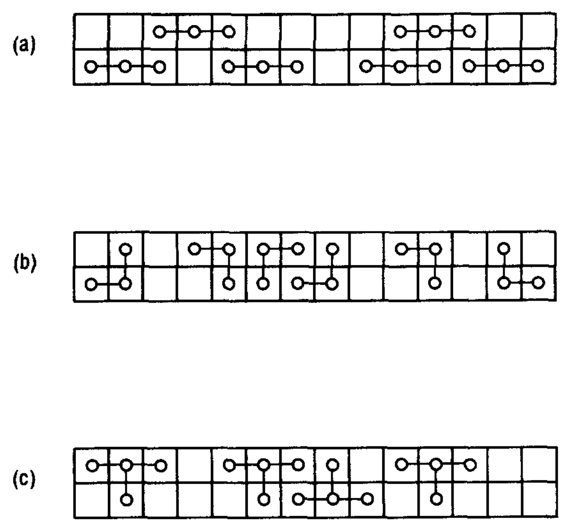


FIG. 1. (a) shows six trimers arranged on a $[2 \times (14)]$ rectangular lattice space; (b) shows six L-particles arranged on a $[2 \times (14)]$ rectangular lattice space; (c) shows four T-particles arranged on a $[2 \times (14)]$ rectangular lattice space.

 0022-2488/86/123016-06$02.50

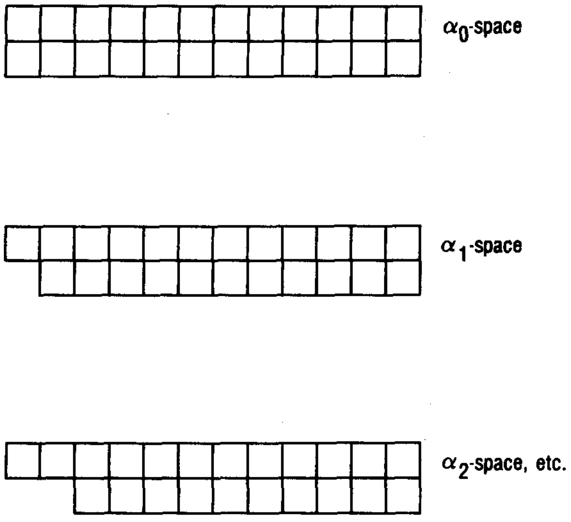


FIG. 2. This figure serves to define the three $[2\times N]$ rectangular lattice spaces, α_0, α_1, and α_2.

$\alpha_j(N)$-spaces to treat indistinguishable trimer particles.

II. TRIMER PARTICLES

We inquire as to the state of occupation of the upper, left-hand site(s) of an α_j-space. Referring to Fig. 3, we see that $A_j[N,q]$, the number of unique arrangements of q indistinguishable trimers on an $\alpha_j(N)$ lattice can be decomposed as follows:

$$A_0[N,q] = A_1[N,q] + A_3[N,q-1]\,, \tag{1a}$$

$$A_1[N,q] = A_0[N-1,q] + A_2[N-1,q-1]\,, \tag{1b}$$

$$A_2[N,q] = A_1[N-1,q] + A_1[N-2,q-1]\,, \tag{1c}$$

$$A_3[N,q] = A_2[N-1,q] + A_0[N-3,q-1]\,. \tag{1d}$$

Equations (1a)–(1d) describe the decomposition of the sets of arrangements of trimers on the various α_j-spaces. Thus, Eq. (1a) says that the number of ways of arranging q trimers on an α_0 lattice is just the number of ways of arranging the trimers when the upper left-hand site is vacant *plus* the number of arrangements when the upper left hand compartment is occupied. (See the three upper left-hand spaces in Fig. 3.) These are the only two logical alternatives. Since these sets are disjoint, their union is merely their sum.

If we associate the shift operator R with the reduction of the value of N by 1, and the shift operator S with the reduction of the value of q by 1, then

$$A_i[N-j,q-k] = R^jS^kA_i[N,q]\,. \tag{2}$$

Because these shift operators commute, they may be manipulated in a manner analogous to the way in which scalar variables are treated. Eqs. (1a)–(1d) may then be written in matrix form:

$$\begin{pmatrix} -1 & 1 & 0 & S \\ R & -1 & RS & 0 \\ 0 & R+R^2S & -1 & 0 \\ R^3S & 0 & R & -1 \end{pmatrix}\begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix} = 0\,. \tag{3}$$

For Eq. (3) to have a nontrivial solution, the determinant of the shift operator matrix should annihilate the solution space $(A_0A_1A_2A_3)$.[16–18] Thus, the determinant of the

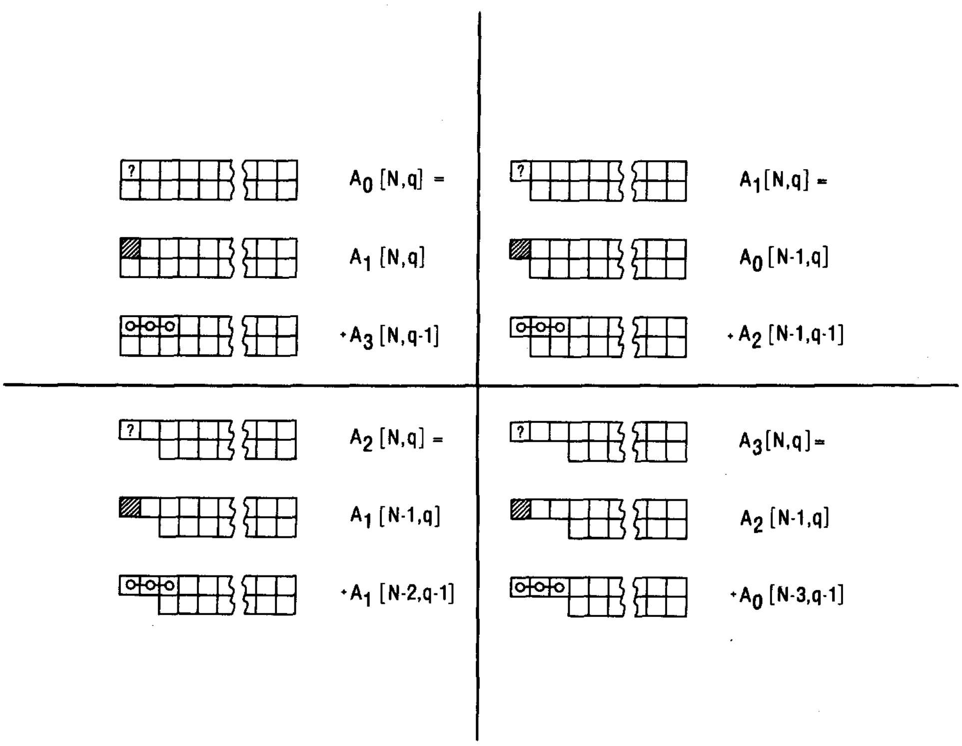


FIG. 3. The decomposition of the degeneracies A_0, A_1, A_2, and A_3.

shift operator matrix, operating on *any* of the A_j's, yields the same recursion:

$$
\begin{aligned}
A[N,q] = & A[N-1,q] + A[N-2,q-1] \\
& + A[N-3,q-1] + 2A[N-3,q-2] \\
& + A[N-4,q-2] - A[N-5,q-3] \\
& - A[N-6,q-4]. \qquad (4)
\end{aligned}
$$

Incidentally, the shift-operator matrix contained in Eq. (3) may be readily extended to λ-bell particles (where λ is the number of linearly contiguous sites occupied by each particle). The corresponding shift operator matrix is a $[\lambda+1]\times[\lambda+1]$ field in which the elements of the main diagonal are -1; the elements of the diagonal just below the main diagonal always contain R and the main reverse diagonal elements always contain $R^{r-1}S$, where r is the number of the row ($1\leqslant r<\lambda+1$); in addition, the element of the first row, second column is 1. All other matrix elements vanish.

The normalization for the statistics described in Eq. (4), Δ_N, is

$$\Delta_N \equiv \sum_{q=0} A[N,q]. \qquad (5)$$

Using Eq. (4), one obtains from Eq. (5) a recursion for the normalization:

$$
\begin{aligned}
\Delta_N = & \Delta_{N-1} + \Delta_{N-2} + 3\Delta_{N-3} \\
& + \Delta_{N-4} - \Delta_{N-5} - \Delta_{N-6}. \qquad (6)
\end{aligned}
$$

If we assume that $\Delta_N = kR^N$, then Eq. (6) becomes

$$R^6 - R^5 - R^4 - 3R^3 - R^2 + R + 1 = 0, \qquad (7)$$

which has the following roots:

$$
\begin{aligned}
& R_1 = 2.147\,899\,04, \\
& R_2 = -0.341\,163\,901 + i[1.161\,5414], \\
& R_3 = -0.341\,163\,901 - i[1.161\,5414], \\
& R_4 = -0.539\,495\,17 + i[0.368\,989\,40], \\
& R_5 = -0.539\,495\,17 - i[0.368\,989\,40], \\
& R_6 = 0.682\,327\,803.
\end{aligned} \qquad (8)
$$

Thus, a general solution for Δ_N is a linear combination of these solutions, i.e.,

$$\Delta_N = \sum_{j=1}^{6} k_j R_j^N, \qquad (9)$$

where, using the initial values from Table I, we obtain

$$
\begin{aligned}
& k_1 = 0.372\,922\,451, \\
& k_2 = 0.237\,569\,540 - i[0.149\,876\,314], \\
& k_3 = 0.237\,569\,540 + i[0.149\,876\,314], \\
& k_4 = 0.227\,162\,317 + i[0.047\,611\,525], \\
& k_5 = 0.227\,162\,317 - i[0.047\,611\,525], \\
& k_6 = 0.105\,506\,147.
\end{aligned} \qquad (10)
$$

As $N\to\infty$, $\Delta_N \cong k_1 R_1^N$, i.e.,

$$\Delta_N = [(0.372\,922\,451][2.147\,898\,904]^N. \qquad (11)$$

If we define the coverage θ to be the fraction of all lattice sites that are occupied, then the expectation of θ may be written

$$\langle\theta\rangle_N \equiv 3\langle q\rangle_N/2N, \qquad (12)$$

where

$$
\begin{aligned}
\langle q\rangle_N & \equiv \left\{\sum_q qA[N,q]\right\}\left\{\sum_q A[N,q]\right\}^{-1} \\
& = \frac{1}{\Delta_N}\sum_q qA[N,q]. \qquad (13)
\end{aligned}
$$

It follows that

$$\langle\theta\rangle_N = \frac{3}{2N\Delta_N}\sum_q qA[N,q]. \qquad (14)$$

Utilizing the recursion, Eq. (4), and assuming that, for sufficiently large N,

$$\langle\theta\rangle_N \cong \langle\theta\rangle_{N-1} \cong \langle\theta\rangle_{N-2} \cong \cdots \equiv \langle\theta\rangle, \qquad (15)$$

we obtain

$$\langle\theta\rangle = \frac{3}{2}\left\{\frac{\Delta_{N-2} + 5\Delta_{N-3} + 2\Delta_{N-4} - 3\Delta_{N-5} - 4\Delta_{N-6}}{\Delta_{N-1} + 2\Delta_{N-2} + 9\Delta_{N-3} + 4\Delta_{N-4} - 5\Delta_{N-5} - 6\Delta_{N-6}}\right\}, \qquad (16)$$

or, using Eq. (11), we may write Eq. (16) as

$$
\begin{aligned}
\langle\theta\rangle & = \frac{3}{2}\left\{\frac{R_1^4 + 5R_1^3 + 2R_1^2 - 3R_1 - 4}{R_1^5 + 2R_1^4 + 9R_1^3 + 4R_1^2 - 5R_1 - 6}\right\} \\
& \cong 0.582\,762\,012. \qquad (17)
\end{aligned}
$$

Thus, assuming the validity of the central limit theorem, the maximum number of arrangements of trimers on a $2\times N$ lattice occurs when the lattice is approximately 58.3% filled. The open circles in Fig. 4 show $A[39,\theta]$ as a function of θ, according to Eq. (4), i.e., the open circles are the exact occupational degeneracy as a function of coverage.

The dispersion in θ, $\langle\theta^2\rangle_N$, is defined by

$$
\begin{aligned}
\langle\theta^2\rangle_N & \equiv \frac{9}{4N^2}\langle q^2\rangle_N \\
& = \frac{9}{4N^2\Delta_N}\sum_q q^2 A[N,q]. \qquad (18)
\end{aligned}
$$

Utilizing the recursion, Eq. (4), and assuming that $\langle\theta^2\rangle_N$ is of the form

$$\frac{\alpha}{N} + \langle\theta\rangle_N^2,$$

TABLE I. Showing the occupational degeneracy for trimers distributed on a $2\times N$ lattice for various values of N and q.

$N \backslash q$	0	1	2	3	4	5	6	Δ_N
1	1							1
2	1							1
3	1	2	1					4
4	1	4	4					9
5	1	6	9	0				16
6	1	8	18	8	1			36
7	1	10	31	30	9			81
8	1	12	48	72	36			169
9	1	14	69	142	114	20	1	361
10	1	16	94	248	289	120	16	784

where α is a constant, we obtain

$$\sigma_N \equiv [\langle\theta^2\rangle_N - \langle\theta\rangle_N^2]^{1/2} = 0.472\,257\,165\,N^{-1/2}. \tag{19}$$

It follows that for large values of N, $A[N,q]$ may be represented as a Gaussian distribution

$$A\,[N,\theta] = A_{\max}\exp\{-[\theta-\langle\theta\rangle]^2/2\sigma_N^2\}, \tag{20}$$

where

$$A_{\max} = [0.315\,028\,641][2.147\,898\,904]^N N^{-1/2}. \tag{21}$$

The solid curve in Fig. 4 has been calculated according to Eq. (20) for $N=39$, and normalized with respect to $A_{\max}=6.69\times10^{11}$.

III. *L*-PARTICLES

By the method outlined above, a recursion for the occupational degeneracy of L-particles distributed on a $2\times N$ lattice, can be obtained as follows [see Fig. 1(b)]: By a set theoretic argument utilized to obtain Eq. (4), we obtain for L-particles

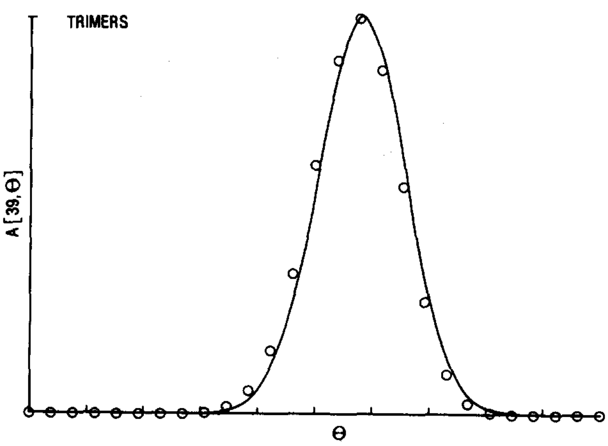


FIG. 4. This figure shows the occupational degeneracy, $A[39,\theta]$, as a function of $\theta\equiv(3/2)(q/N)$ for trimers. The open circles show $A[39,\theta]$ according to Eq. (4). The solid curve has been calculated from Eqs. (19)–(21). Both have been normalized to $A_{\max}=6.69\times10^{11}$.

$$A\,[N,q] = A\,[N-1,q] + 4A\,[N-2,q-1] + 2A\,[N-3,q-2]. \tag{22}$$

The normalization is seen to be

$$\Delta_N \equiv \sum_q A\,[N,q] = \Delta_{N-1} + 4\Delta_{N-2} + 2\Delta_{N-3}. \tag{23}$$

Again, assuming $\Delta_N = kR^N$, it is found that

$$R^3 - R^2 - 4R - 2 = 0,$$

which has the roots

$$R_1 = 2.732\,050\,81, \quad R_2 = -0.732\,050\,81, \quad R_3 = -1, \tag{24}$$

where

$$k_1 = 0.577\,350\,266, \quad k_2 = -0.577\,350\,266, \quad k_3 = 1. \tag{25}$$

Using $\Delta_N \cong k_1 R_1^N$ for large N,

$$\Delta_N \cong [0.577\,350\,266][2.732\,050\,81]^N. \tag{26}$$

Then the expectation of θ, defined by

$$\langle\theta\rangle_N \equiv \frac{3}{2N}\langle q\rangle_N = \frac{3}{2N\Delta_N}\sum_q qA\,[N,q],$$

is

$$\langle\theta\rangle \cong 0.633\,974\,596. \tag{27}$$

The standard deviation, σ_N, defined by

$$\sigma_N \equiv [\langle\theta^2\rangle_N - \langle\theta\rangle_N^2]^{1/2} \cong 0.412\,156\,501\,N^{-1/2}, \tag{28}$$

so that for large values of N, $A\,[N,q]$ can be represented by a Gaussian distribution:

$$A\,[N,\theta] = A_{\max}\exp\{-(\theta-\langle\theta\rangle)^2/2\sigma_N^2\}, \tag{29}$$

where

$$A_{\max} \cong (0.838\,259\,61)[2.732\,050\,81]^N N^{-1/2}. \tag{30}$$

The open circles in Fig. 5 are the exact values of $A[39,\theta]$ calculated from Eq. (22) and the solid curve has been calculated according to Eq. (29) for $N=39$ and normalized with respect to $A_{\max}=1.42\times10^{16}$.

IV. *T*-PARTICLES

For T-particles the recursion for the occupational degeneracy is quite simple:

$$A\,[N,q] = A\,[N-1,q] + A\,[N-2,q-1] + A\,[N-3,q-1], \tag{31}$$

with the normalization

$$\Delta_N = \Delta_{N-1} + \Delta_{N-2} + \Delta_{N-3}. \tag{32}$$

The roots of the cubic associated with Eq. (32) are

$$R_1 = 1.839\,286\,76, \quad R_2 = -0.419\,643\,378 + i[0.606\,290\,729], \quad R_3 = -0.419\,643\,378 - i[0.606\,290\,729], \tag{33}$$

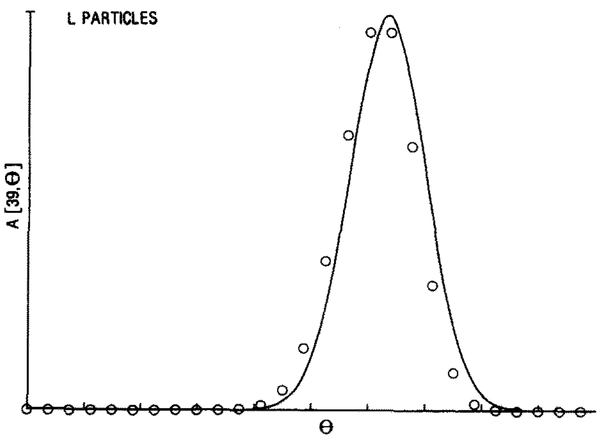


FIG. 5. This figure shows the occupational degeneracy, $A[39,\theta]$, as a function of $\theta \equiv (3/2)(q/N)$ for L-particles. The open circles show $A[39,\theta]$ according to Eq. (22). The solid curve has been calculated from Eqs. (28)–(30). Both have been normalized to $A_{\max} = 1.42\times 10^{16}$.

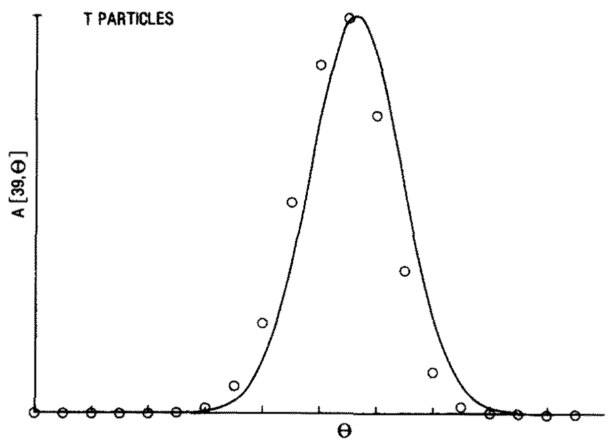


FIG. 6. The degeneracy $A[39,\theta]$ is shown as a function of $\theta \equiv 2q/N$ for T-particles. The open circles have been determined using Eq. (31) and the solid curve has been calculated from Eqs. (37)–(39). Both have been normalized to $A_{\max} \equiv 2.31\times 10^{9}$.

with the associated coefficients

$$\begin{aligned} k_1 &= 0.435\,616\,379\,, \\ k_2 &= 0.282\,191\,146 - i[0.359\,246\,981]\,, \\ k_3 &= 0.282\,191\,146 + i[0.359\,246\,98]\,. \end{aligned} \tag{34}$$

Thus for large N

$$\Delta_N \cong k_1 R_1^N = (0.435\,616\,379)[1.839\,286\,76]^N, \tag{35}$$

$\langle\theta\rangle_N$, the expectation of θ, is then

$$\begin{aligned} \langle\theta\rangle_N &\equiv \frac{2}{N}\langle q\rangle_N = \frac{2}{N\Delta_N}\sum_q qA[N,q] \\ &\cong 0.564\,383\,61\,. \end{aligned} \tag{36}$$

In a similar way, the standard deviation becomes

$$\sigma_N = [0.505\,185\,901]N^{-1/2}\,. \tag{37}$$

Then, for large N, $A[N,\theta]$ can be represented as

$$A[N,\theta] \cong A_{\max}\exp\{-(\theta-\langle\theta\rangle)^2/2\sigma_N^2\}\,, \tag{38}$$

where

$$A_{\max} \cong (0.688\,007\,293)[1.839\,286\,76]^N N^{-1/2}\,. \tag{39}$$

The open circles in Fig. 6 are the exact values of $A[39,\theta]$, as calculated from Eq. (31), while the solid curve shows $A[39,\theta]$ as a function of θ as calculated from Eq. (38) and normalized with respect to $A_{\max} = 2.31\times 10^9$.

V. COMPARISON OF STATISTICS

It is apparent from Eqs. (4), (22), and (31), as well as from the occupational degeneracy recursion for simple particles distributed on a $2\times N$ lattice,

$$\begin{aligned} A[N,q] = A[N-1,q] &+ 2A[N-1,q-1] \\ &+ A[N-1,q-2]\,, \end{aligned}$$

that the size, shape, and number of degrees of freedom of a particular kind of particle are not obviously reflected in the complexity of the respective recursion.

It is interesting to note that the expectations of the coverage for the three kinds of particles under consideration are all greater than that for simple particles, i.e., are greater than 0.5. From these expectations,

$$\begin{aligned} \langle\theta\rangle &\cong 0.582\,762\,012 \quad \text{(trimers)}\,, \\ \langle\theta\rangle &\cong 0.633\,974\,596 \quad (L\text{-particles})\,, \\ \langle\theta\rangle &\cong 0.564\,383\,610 \quad (T\text{-particles})\,, \end{aligned}$$

and from the general expression for the expectation,

$$\langle\theta\rangle = \frac{s}{2N\Delta_N}\sum_q qA[N,q] = \frac{s}{2N}\langle q\rangle\,,$$

where s is the number of sites occupied by each kind of particle, we see that although $\langle q\rangle$ is greater for simple particles, it cannot compensate for the fact that each simple particle occupies only one site.

While s is the same for trimers and L-particles, $\langle\theta\rangle$ is greater for the latter because there are more unique ways to arrange L-particles than trimers on a $2\times N$ lattice. Even though s is greater for T-particles, they cannot be arranged in as many unique ways as can either trimers or L-particles.

The maximum number of arrangements for each kind of particle (for $N = 39$),

$$\begin{aligned} A_{\max} &= 6.69\times 10^{11} \quad \text{(trimers)}\,, \\ A_{\max} &= 1.42\times 10^{16} \quad (L\text{-particles})\,, \\ A_{\max} &= 2.31\times 10^{9} \quad (T\text{-particles})\,, \end{aligned}$$

also reflects the fact that at any coverage, the T-particles cannot be arranged in as many ways as either the trimers or the L-particles. Here $A_{\max}$ for simple particles on the same lattice would be 2.73×10^{22}.

The standard deviation of the coverage for the three kinds of particles considered,

$$\sigma = 0.472\,257\,165\, N^{-1/2} \quad (\text{trimers}),$$
$$\sigma = 0.412\,156\,501\, N^{-1/2} \quad (L\text{-particles}),$$
$$\sigma = 0.505\,185\,901\, N^{-1/2} \quad (T\text{-particles}),$$

are to be compared to

$$\sigma = 0.707\,184\,951\, N^{-1/2},$$

for simple particles. These values indicate that the compactness and rotational freedom of the L-particles result in a narrower distribution. It is also interesting to note that all three more complicated particles have considerably sharper distributions than the distribution for simple particles.

An examination of Figs. 4–6 reveals that the discrete values of $A[39,\theta]$ for each kind of particle appear to be shifted uniformly along the θ axis, relative to the corresponding curves representing the continuous distribution. This indicates that at $N=39$, the values of the dispersions are more accurate than the values of the expectations.

VI. CONCLUSION

We have determined recursions that enumerate exactly the multiplicity of arrangement of indistinguishable trimer, L-, and T-particles distributed on a $2\times N$ rectangular lattice. Utilizing these recursions, we have calculated the expectation, normalization, dispersion, and continuous representation of the occupation statistics for each kind of particle. Comparisons among the statistics for each kind of particle and with the statistics for simple particles have been made.

[1]D. Lichtman and R. B. McQuistan, "Exact occupation statistics for one-dimensional arrays of dumbbells," J. Math. Phys. **8**, 2441 (1967).

[2]R. B. McQuistan, "Exact occupational kinetics for one-dimensional arrays of dumbbells," J. Math. Phys. **9**, 1680 (1968).

[3]R. B. McQuistan, "Vacancy annihilation for one-dimensional dumbbell kinetics," J. Math. Phys. **10**, 2205 (1969).

[4]R. B. McQuistan, "Exact recursion relation for $2\times N$ arrays of dumbbells," J. Math. Phys. **11**, 3095 (1970).

[5]R. B. McQuistan, S. J. Lichtman, and L. P. Levine, "Occupation statistics for parallel dumbbells on a $2\times N$ lattice space," J. Math. Phys. **13**, 242 (1972).

[6]R. C. Grimson, "Exact formulas for $2\times N$ arrays of dumbbells," J. Math. Phys. **15**, 214 (1974).

[7]M. A. Rashid, "Occupation statistics for exact recursion relations for occupation by dumbbells on a $2\times N$ array," J. Math. Phys. **15**, 474 (1974).

[8]A. F. Antippa and A. J. Phâres, "General formalism solving linear recursion relations," J. Math. Phys. **18**, 173 (1977).

[9]A. J. Phares, "A general theory of linear finite-difference equations with a few applications," J. Math. Phys. **18**, 1838 (1977).

[10]A. J. Phares, "A multidimensional extension of the combinatorics function technique. II. Linear and inhomogeneous partial difference equations," J. Math. Phys. **22**, 1030 (1981).

[11]A. J. Phares and K. G. H. Simmons, "Fibonacci numbers revisited: A combinatorics function point of view," Am. J. Phys. **48**, 159 (1980).

[12]J. L. Hock and R. B. McQuistan, "A note on the occupational degeneracy for dimers on a saturated two-dimensional lattice space," Discrete Appl. Math. **8**, 101 (1984).

[13]A. J. Phares, "The occupation statistics for indistinguishable dumbbells on a rectangular lattice space," J. Math. Phys. **25**, 1756 (1984).

[14]J. L. Hock and R. B. McQuistan, "The occupation statistics for indistinguishable dumbbells on a $2\times 2\times N$ lattice space," J. Math. Phys. **24**, 1859 (1983).

[15]J. L. Hock, P. E. Licato, and R. B. McQuistan, "The occupational statistics for indistinguishable trimers on a $3\times N$ lattice space," J. Math. Phys. **23**, 2185 (1982).

[16]R. B. McQuistan and J. L. Hock, "An exact recursion for the composite nearest-neighbor degeneracy for a $2\times N$ lattice space," J. Math. Phys. **25**, 261 (1984).

[17]J. L. Hock and R. B. McQuistan, "An exact recursion for the occupational degeneracy of dumbbells on a hexagonal lattice space," J. Chem. Phys. **83**, 3626 (1985).

[18]D. Walikainen and R. B. McQuistan, "The statistics of interacting dumbbells on a $2\times N$ lattice space," J. Math. Phys. **26**, 815 (1985).

Virtual power and thermodynamics for electromagnetic continua with interfaces

N. Daher
Laboratoire de Physique et Métrologie des Oscillateurs du Centre National de la Recherche Scientifique associé à l'Université de Franche-Comté-Besançon, 32 avenue de l'Observatoire, 25000 Besançon, France
G. A. Maugin
Laboratoire de Mécanique, Théorique associé au Centre National de la Recherche Scientifique, Université Pierre et Marie Curie, tour 66, 4 place Jussieu, 75230 Paris Cedex 05, France

(Received 5 February 1986; accepted for publication 6 August 1986)

This paper presents a systematic and rational formulation of the electromagnetic theory of deformable and fluent bodies swept out by singular surfaces that may carry their own thermodynamics (interfaces). The treatment is based on the principle of virtual power for finite velocity fields, which is so formulated that, when combined, for real velocity fields, with the first principle of thermodynamics in global form, it yields directly the so-called energy theorem both in the bulk and at the singular surface. Then the corresponding rates of entropy production are deduced after introduction of the second principle of thermodynamics. The various alternate expressions of the ponderomotive force, couple, and electromagnetic energy, obtained in the bulk from the Lorentz theory of electrons are developed across the singular surface by means of the generalized transport and Green–Gauss theorems. Finally, an extension of the constitutive theory (well established in the bulk) is given to account for surface phenomena in the case of an electromagnetic fluid. Thermodynamical restrictions are discussed, and comparisons are made with previous works.

I. INTRODUCTION

In recent years consistent descriptions of mechanical and electromagnetic continuous media have been obtained on the basis of the energy method known as the "principle of virtual power" (for finite velocity fields) rather than the classical vectorial approach. Indeed, in 1973 Germain pointed out the interest of using this method for nonsimple materials such as those exhibiting a microstructure[1,2] and in 1980 Maugin showed that this method is particularly suited to the description of the interactions of electromagnetic fields with deformable bodies.[3] Recently Daher and Maugin used this type of formulation to describe purely mechanical continuous media swept out by various singular surfaces and/or lines, thus giving to that principle a range of application as wide as that of the classical vectorial approach while, of course, retaining all the advantages that it already owns and adding some concerning the transversality conditions.[4]

In the present work, a description of continuous media presenting singular surfaces and including electromagnetic effects is given on the basis of the electron theory of Lorentz (for the evaluation of forces, couples, and energies of electromagnetic nature), the principle of virtual power, and the first and second principles of thermodynamics (for the obtainment of field and constitutive equations). The subject matter has been dealt with, at length, in the bulk so that emphasis is placed here especially on surface phenomena by the use of the generalized transport and Green–Gauss theorems.

It is salient to recall that (i) the electron theory of Lorentz consists of a spatial averaging procedure applied to an assembly of nonrelativistic bound point charges contained within a "microelement" and considered to be under the influence of a Lorentz force[5,6]; (ii) a singular surface may be a strong discontinuity in the well known sense granted in continuum mechanics (e.g., shock) or an interface between two phases, which, in many cases, can also be conveniently simulated by a strong discontinuity, but the latter then has material properties in the same way as the bulk phases[4,7]; (iii) a virtual power is a linear continuous form on a set of virtual velocities. The dual quantity to a "velocity" is a "force." The selection of a space of admissible velocities fixes, via this duality, the degree of refinement of the description of forces acting on the system. For so-called internal forces for which one ultimately needs to construct constitutive equations we suppose that the principle of objectivity applies, which, in turn, implies that the dual "velocity field" is objective.[3] In particular, we note that when the medium is spanned by a discontinuity, the principle of objectivity induces an additional internal virtual power, which accounts for the relative motion of the medium and the singularity.[4]

For simplicity, we restrict ourselves to the case where only surface charges and currents are allowed to exist at the interface, while all electromagnetic contributions may be discontinuous across the singular surface. Surface electromagnetic fields have been considered by other authors[7] using a direct postulational approach of global balance laws, or by the present authors[8] for specific applications.

The notation used is recalled in Sec. II. In a general manner we use indifferently the direct (intrinsic) dyadic notation or the notation of Cartesian tensors in rectangular

 0022-2488/86/123022-14$02.50

coordinate systems. The Galilean form of Maxwell's equations is given in Sec. III. This Galilean formulation is not a severe limitation since material velocities encountered in practical applications are considerably less than the speed of light. The nonlinear electromechanical equations are developed in Sec. IV on the basis of the principle of virtual power and the two fundamental principles of thermodynamics written globally for the whole specimen. The first of these, written in a rotationally invariant form, allows one to exhibit the nonlinear contribution of magnetization and electric polarization to the Cauchy stress tensor of the deformable body and it yields directly the transversality condition relative to the surface stress tensor (while this is postulated in the classical vectorial approach). The local volume and surface equations of balance of momentum, angular momentum, and energy and the local statement of the entropy principle follow from this global formulation when real velocity fields are considered. Finally, the above-mentioned local equations are written in a particular case in order to be compared to a previous work performed by Maugin and Eringen on the basis of the vectorial approach.[5] In Sec. V we consider the special case of electromagnetic fluids, for which constitutive equations are constructed using the complete apparatus of nonlinear continuum thermodynamics and its most recent developments. It is shown that the bulk constitutive equations are exactly the same as the ones derived by Eringen[6] while the surface constitutive equations generalize a previous work performed by Bedeaux, Albano, and Mazur.[9]

It is in the nature of the subject matter that the algebra required be long and tedious. Because of the lack of space, we have often indicated only the guideline of the derivation. To render the paper self-contained, however, useful identities, definitions, mathematical properties, and integral transformations have been recalled or derived in the Appendices.

All through the paper analogies or symmetries have been made between bulk and surface equations insofar as possible. In particular, it is shown, in Appendix A, that the effective Lorentz and ponderomotive forces, written in terms of effective charges and currents and expressed with respect to a frame moving with the infinitesimal element of matter, are more convenient to deal with than any other alternate form. The above-mentioned remark takes its full importance when dealing with more complex media such as piezoelectric semiconductors where the electromagnetic continuum is to be split in separate continua.[8]

Finally, we notice that the principle of virtual power may be stated in two different manners, either following a systematic procedure, where the electromagnetic tensor and momentum are introduced through a so-called first gradient theory,[1,3] or by the introduction of the electromagnetic (ponderomotive) forces in the same way as gravitational forces are usually introduced. The first, more fundamental, statement is given in Appendix C while the second one, which is more practical, is given in Sec. IV. Naturally, the two alternate forms are mathematically equivalent.

II. NOTATION

We use the classical notation of continuum mechanics. The general nonlinear deformation of a body B between its reference configuration K_R at time t_0 and its present configuration K_t at time t is represented, at fixed t, by the diffeomorphism

$$x_i = \mathfrak{X}_i(X_K,t) \tag{2.1}$$

where X_K, $K = 1, 2, 3$, and x_i, $i = 1,2,3$, denote the position in rectangular coordinate systems—which need not coincide—in K_R and K_t, respectively, of the same material "particle," the latter concept being understood in the usual continuous framework. The material body B occupies the volume D-Σ of Euclidean physical space E^3 at time t and it is spanned by a singular surface $\Sigma(t)$. The boundary of the volume is noted ∂D-Σ with unit outward normal $\mathbf{n}$. The absolute velocity of $\Sigma(t)$, with respect to a fixed Galilean frame R_G (the so-called laboratory frame of electrodynamics), is noted $\mathbf{v}$ and the unit oriented normal to $\Sigma(t)$ is $\hat{\mathbf{n}}$. The boundary on ∂D of the singular surface, noted $\partial\Sigma$, is equipped with a unit tangent $\mathbf{t}$ and unit normal $\boldsymbol{\tau}$ in the local tangent plane to Σ such that $\boldsymbol{\tau} = \mathbf{n}\times\mathbf{t}$. The vector-valued function $\mathfrak{X}$ is assumed to be sufficiently differentiable in its arguments in D-Σ so as to allow for the forthcoming manipulation. The velocity field $\mathbf{v}$, the direct motion gradient $\mathbf{F}$, the inverse motion gradient $\mathbf{F}^{-1}$, the rate-of-strain tensor $\mathbf{D}$, and the rate-of-rotation tensor $\boldsymbol{\Omega}$ of particles at regular points in D-Σ or ∂D-Σ are classically defined by

$$\mathbf{v} = \frac{\partial \mathfrak{X}}{\partial t}\bigg|_{\text{fixed } X_K} = \{v_i\}, \tag{2.2}$$

$$\mathbf{F} = \left\{x_{i,K} = \frac{\partial \mathfrak{X}_i}{\partial X_K}\right\}, \quad \mathbf{F}^{-1} = \left\{X_{K,i} = \frac{\partial X_K}{\partial x_i}\right\} \tag{2.3}$$

$(J = \det \mathbf{F} > 0 \text{ always})$,

$$\mathbf{D} = \{D_{ij} = v_{(i,j)} \equiv \tfrac{1}{2}(v_{i,j} + v_{j,i}) = D_{ji}\}, \tag{2.4}$$

and

$$\boldsymbol{\Omega} = \{\Omega_{ij} = v_{[i,j]} \equiv \tfrac{1}{2}(v_{i,j} - v_{j,i}) = -\Omega_{ji}\}, \tag{2.5}$$

respectively, so that

$$x_{i,K}\, X_{K,j} = \delta_{ij}, \quad X_{K,i}x_{i,L} = \delta_{KL}, \tag{2.6}$$

where δ_{ij} and δ_{KL} are Kronecker symbols. The Einstein summation convention is understood. Cartesian tensor notation and direct dyadic notation are used indifferently. In the latter case the gradient operators are

$$\boldsymbol{\nabla} = \left\{\frac{\partial}{\partial x_i};\ i = 1,2,3\right\}, \quad \boldsymbol{\nabla}_R = \left\{\frac{\partial}{\partial X_K};\ K = 1,2,3\right\}. \tag{2.7}$$

The divergence of *non-necessarily* symmetric second-order tensors is taken with respect to the *last* index, e.g.,

$$(\operatorname{div} \mathbf{t})_i = t_{ij,j}. \tag{2.8}$$

When material quantities are attached to the singular surface Σ, the corresponding fields are denoted by a superimposed caret "$\hat{\ }$" and the surface is said to be *thermodynamic*; otherwise it is said to be *free*.[4] For instance $\hat{\mathbf{v}}$ is the absolute velocity, with respect to R_G, of particles which belong to Σ. As these particles cannot leave Σ, we obviously have

$$\hat{\mathbf{v}}\cdot\hat{\mathbf{n}} = \mathbf{v}\cdot\hat{\mathbf{n}}, \tag{2.9}$$

while for a *free* singular surface we must necessarily set $\hat{\mathbf{v}} = \boldsymbol{\nu}$.[4]

The cut of the material body D by a singular surface $\Sigma(t)$ in two regular regions D^- and D^+,[4] $\hat{\mathbf{n}}$ being oriented from D^- to D^+, requires the introduction of the following notation:

$$\mathbf{v}^{\pm} = \begin{cases} \mathbf{v}H(D^{\pm}), \\ \mathbf{v} \quad \text{if} \quad \mathbf{x} \in \partial D^{\pm}, \\ \text{uniform limit of } v(D^{\pm}) \text{ for } D^{\pm} \ni \mathbf{x} \to \Sigma^{\pm} \text{ along } \hat{\mathbf{n}}^{\pm}, \end{cases} \tag{2.10}$$

where H is the characteristic (Heaviside) function of a set. The symbols $[\cdots]$ and $\langle \cdots \rangle$ indicate, respectively, the jump and mean value of their enclosures at $\Sigma(t)$, e.g.,

$$[A] = A^+ - A^-, \quad \langle A \rangle = \tfrac{1}{2}(A^+ + A^-), \tag{2.11}$$

where $A^{\pm}$ are the uniform limits of the field A (regular in D-Σ but presenting a *finite* discontinuity at Σ) in approaching Σ on its two faces along its normals $\hat{\mathbf{n}}^{\pm}$.

In the forthcoming development we need the balance of mass, which cannot be deduced from the principle of virtual power. Let ρ and $\hat{\rho}$ be the matter volume density in D-Σ and the matter surface density on Σ, respectively. The total mass conservation reads

$$\frac{d}{dt}\int_{D\text{-}\Sigma} \rho \, dv + \frac{\hat{d}}{dt}\int_{\Sigma} \hat{\rho}\, da = 0, \tag{2.12}$$

where the "material" time derivatives are defined by

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla, \quad \frac{\hat{d}}{dt} = \frac{\partial}{\partial t} + \hat{\mathbf{v}}\cdot\nabla. \tag{2.13}$$

By using a "generalized" version of transport theorems[4] (see also Appendix D), the global statement (2.12) is shown to imply the following local ones:

$$\frac{d\rho}{dt} + \rho\nabla\cdot\mathbf{v} = 0 \quad \text{in } D\text{-}\Sigma, \tag{2.14}$$

$$\frac{\hat{d}\hat{\rho}}{dt} + \hat{\rho}\hat{\nabla}\cdot\hat{\mathbf{v}} + [m] = 0 \quad \text{on } \Sigma, \tag{2.15}$$

where

$$m = \rho(\mathbf{v} - \boldsymbol{\nu})\cdot\hat{\mathbf{n}} = \rho(\mathbf{v} - \hat{\mathbf{v}})\cdot\hat{\mathbf{n}} \tag{2.16}$$

is the so-called *mass transfer* across Σ. Equivalently, Eqs. (2.14) and (2.15) may be written as

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \quad \text{in } D\text{-}\Sigma, \tag{2.17}$$

$$\frac{\bar{\delta}}{\delta t}\hat{\rho} + \hat{\nabla}\cdot(\hat{\rho}\hat{\mathbf{v}}) + [m] = 0 \quad \text{on } \Sigma, \tag{2.18}$$

where

$$\frac{\bar{\delta}}{\delta t} = \frac{\partial}{\partial t} + \hat{\mathbf{v}}\cdot\hat{\mathbf{n}}D = \frac{\partial}{\partial t} + \boldsymbol{\nu}\cdot\hat{\mathbf{n}}D, \quad D = \hat{\mathbf{n}}\cdot\nabla, \quad \hat{\nabla}_i = P_{ij}\nabla_j, \tag{2.19}$$

$$P_{ij} = \delta_{ij} - \hat{n}_i\hat{n}_j, \quad \Omega = -\tfrac{1}{2}\nabla\cdot\hat{\mathbf{n}}.$$

III. MAXWELL'S EQUATIONS

Let $\mathbf{E}$, $\mathbf{B}$, $\mathbf{D}$, $\mathbf{H}$, $\mathbf{J}$, q_f, $\mathbf{P}$, and $\mathbf{M}$, denote the electric field, the magnetic induction, the electric displacement, the magnetic field, the total current, the volume density of free charges, the electric polarization per unit volume, and the magnetization per unit volume, all evaluated in the fixed Galilean frame R_G at time t. At internal points of the regular region D-Σ in K_t, Maxwell's equations are classically expressed by (c = velocity of light in vacuum; Lorentz–Heaviside units are used so that neither factor 4π nor vacuum quantities ϵ_0 and μ_0 appear),

$$\nabla\times\mathbf{E} + \frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} = 0, \quad \nabla\cdot\mathbf{B} = 0, \tag{3.1}$$

$$\nabla\times\mathbf{H} - \frac{1}{c}\frac{\partial\mathbf{D}}{\partial t} = \frac{1}{c}\mathbf{J}, \quad \nabla\cdot\mathbf{D} = q_f, \tag{3.2}$$

with

$$\mathbf{P} = \mathbf{D} - \mathbf{E}, \quad \mathbf{M} = \mathbf{B} - \mathbf{H}. \tag{3.3}$$

On taking the divergence of the first equation of Eq. (3.2) and accounting for the second equation of (3.2) one obtains the equation of conservation of electric charges as

$$\frac{\partial q_f}{\partial t} + \nabla\cdot\mathbf{J} = 0. \tag{3.4}$$

It seems convenient to define an effective charge q^{eff} and an effective total current, in R_G, as

$$q^{\text{eff}} = q_f - \nabla\cdot\mathbf{P}, \quad \mathbf{J}^{\text{eff}} = \mathbf{J} + \frac{\partial\mathbf{P}}{\partial t} + c\nabla\times\mathbf{M}, \tag{3.5}$$

so that Eq. (3.2) transforms to

$$\nabla\times\mathbf{B} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} = \frac{1}{c}\mathbf{J}^{\text{eff}}, \quad \nabla\cdot\mathbf{E} = q^{\text{eff}}. \tag{3.6}$$

Let $\mathscr{E}$, $\mathscr{B}$, $\mathscr{H}$, $\mathscr{P}$, $\mathscr{J}$, and $\mathscr{M}$ be the same fields as $\mathbf{E}$, $\mathbf{B}$, $\mathbf{H}$, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{M}$, but referred to a frame moving with the infinitesimal element of matter at time t, [the so-called comoving frame R_c $(\mathbf{x},t)$]. In the *Galilean approximation*, which is sufficient for the present purpose, we have the following transformation laws between R_G and R_c $(\mathbf{x},t)$[5,6]:

$$\mathscr{E} = \mathbf{E} + (1/c)\mathbf{v}\times\mathbf{B}, \tag{3.7}$$

$$\mathscr{B} = \mathbf{B} - (1/c)\mathbf{v}\times\mathbf{E}, \tag{3.8}$$

$$\mathscr{H} = \mathbf{H} - (1/c)\mathbf{v}\times\mathbf{D} = \mathscr{B} - \mathscr{M}, \tag{3.9}$$

$$\mathscr{J} = \mathbf{J} - q_f\mathbf{v}, \tag{3.10}$$

$$\mathscr{P} = \mathbf{P}, \quad \mathscr{M} = \mathbf{M} + (1/c)\mathbf{v}\times\mathbf{P}. \tag{3.11}$$

The lack of symmetry between the last two formulas reflects the Galilean approximation. The vector field $\mathscr{E}$ is usually called the *electromotive intensity* while $\mathscr{J}$ is none other than the *conduction current*.

On account of (3.7)–(3.11), the bulk equations (3.1) and (3.2) and (3.4) take on the following "Galilean" form[5,6,10]:

$$\nabla\times\mathscr{E} + (1/c)\overset{*}{\mathbf{B}} = 0, \quad \nabla\cdot\mathbf{B} = 0, \tag{3.12}$$

$$\nabla\times\mathscr{H} - (1/c)\overset{*}{\mathbf{D}} = (1/c)\mathscr{J}, \quad \nabla\cdot\mathbf{D} = q_f, \tag{3.13}$$

and

$$\mathscr{D}q_f + \nabla\cdot\mathscr{J} = 0, \tag{3.14}$$

where the convected-time derivatives for a vector $\mathbf{A}$ and a scalar a are defined by

$$\overset{*}{\mathbf{A}} \equiv \frac{d\mathbf{A}}{dt} - (\mathbf{A}\cdot\nabla)\mathbf{v} + \mathbf{A}(\nabla\cdot\mathbf{v})$$

$$= \frac{\partial \mathbf{A}}{\partial t} + \mathbf{v}(\nabla\cdot\mathbf{A}) + \nabla\times(\mathbf{A}\times\mathbf{v}), \quad (3.15)$$

$$\mathscr{D}a \equiv \frac{da}{dt} + a\nabla\cdot\mathbf{v}. \quad (3.16)$$

In the same way as in Eqs. (3.6), we may write Eq. (3.13) in the following alternate form:

$$\nabla\times\mathscr{B} - (1/c)\overset{*}{\mathbf{E}} = (1/c)\mathscr{J}^{\text{eff}}, \quad \nabla\cdot\mathbf{E} = q^{\text{eff}}, \quad (3.17)$$

where the effective conduction current $\mathscr{J}^{\text{eff}}$ has been defined as

$$\mathscr{J}^{\text{eff}} = \mathscr{J} + \overset{*}{\mathbf{P}} + c\nabla\times\mathscr{M}. \quad (3.18)$$

On account of (2.9), and after some calculations and manipulations, the jump conditions at the interface associated with the bulk equations (3.1), (3.2), (3.4), and (3.6) in R_G and (3.12)–(3.14) and (3.17) in $R_c(\mathbf{x},t)$, transform, respectively, to[5,6] in R_G,

$$\hat{\mathbf{n}}\times[\mathbf{E}] - (1/c)(\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})[\mathbf{B}] = 0, \quad [\mathbf{B}]\cdot\hat{\mathbf{n}} = 0, \quad (3.19)$$

$$\hat{\mathbf{n}}\times[\mathbf{H}] + (1/c)(\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})[\mathbf{D}] = (1/c)\,\hat{\mathbf{J}}, \quad [\mathbf{D}]\cdot\hat{\mathbf{n}} = \hat{\mathbf{q}}_f, \quad (3.20)$$

$$\frac{\bar{\delta}}{\delta t}\hat{q}_f + \hat{\nabla}\cdot\hat{\mathbf{J}} + [Q] = 0, \quad (3.21)$$

and

$$\hat{\mathbf{n}}\times[\mathbf{B}] + (1/c)(\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})[\mathbf{E}] = (1/c)\hat{\mathbf{J}}^{\text{eff}},$$

$$[\mathbf{E}]\cdot\hat{\mathbf{n}} = \hat{\mathbf{q}}^{\text{eff}}; \quad (3.22)$$

in $R_c(\mathbf{x},t)$,

$$\hat{\mathbf{n}}\times[\mathscr{E} + (1/c)\mathbf{B}\times(\mathbf{v}-\hat{\mathbf{v}})] = 0, \quad [\mathbf{B}]\cdot\hat{\mathbf{n}} = 0, \quad (3.23)$$

$$\hat{\mathbf{n}}\times[\mathscr{H} - (1/c)\mathbf{D}\times(\mathbf{v}-\hat{\mathbf{v}})] = (1/c)\,\hat{\mathscr{J}},$$

$$[\mathbf{D}]\cdot\hat{\mathbf{n}} = \hat{q}_f, \quad (3.24)$$

$$\hat{\mathscr{D}}\hat{q}_f + \hat{\nabla}\cdot\hat{\mathscr{J}} + [Q] = 0, \quad (3.25)$$

and

$$\hat{\mathbf{n}}\times[\mathscr{B} - (1/c)\mathbf{E}\times(\mathbf{v}-\hat{\mathbf{v}})] = (1/c)\,\hat{\mathscr{J}}^{\text{eff}},$$

$$[\mathbf{E}]\cdot\hat{\mathbf{n}} = \hat{q}^{\text{eff}}, \quad (3.26)$$

with

$$\hat{\mathscr{D}}\hat{a} = \frac{\hat{d}}{dt}\hat{a} + \hat{a}\hat{\nabla}\cdot\hat{\mathbf{v}}, \quad (3.27)$$

$$Q = (\mathbf{J} - q_f\hat{\mathbf{v}})\cdot\hat{\mathbf{n}}, \quad (3.28)$$

$$\hat{\mathscr{J}} = \hat{\mathbf{J}} - \hat{q}_f\hat{\mathbf{v}},$$

$$\hat{\mathscr{J}}^{\text{eff}} = \hat{\mathscr{J}} + \hat{\mathbf{n}}\times[\mathbf{P}\times(\mathbf{v}-\hat{\mathbf{v}})] + c\hat{\mathbf{n}}\times[\mathscr{M}], \quad (3.29)$$

$$\hat{q}^{\text{eff}} = \hat{q}_f - [\mathbf{P}]\cdot\hat{\mathbf{n}},$$

$$\hat{\mathbf{J}}^{\text{eff}} = \hat{\mathbf{J}} - (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})[\mathbf{P}] + c\hat{\mathbf{n}}\times[\mathbf{M}], \quad (3.30)$$

where $\hat{\mathscr{J}}$, $\hat{\mathbf{J}}$, $\hat{q}_f$, $\hat{\mathscr{J}}^{\text{eff}}$, $\hat{\mathbf{J}}^{\text{eff}}$, and $\hat{q}^{\text{eff}}$ are the surface conduction current, the total surface current, the surface density of electric charge, the effective surface conduction current, the effective total surface current, and the effective surface density of electric charge. In writing the above jump equations we have assumed that there is neither surface polarization density nor surface magnetization density on $\Sigma(t)$.

The boundary conditions on ∂D-Σ are obtained by considering a singular surface that is material ($\mathbf{v} = \hat{\mathbf{v}} = \mathbf{v}$). Thus Eqs. (3.23)–(3.25) yield

$$\mathbf{n}\times[\mathscr{E}] = 0, \quad [\mathbf{B}]\cdot\mathbf{n} = 0, \quad (3.31)$$

$$\mathbf{n}\times[\mathscr{H}] = (1/c)\hat{\mathscr{J}}, \quad [\mathbf{D}]\cdot\mathbf{n} = \hat{q}_f, \quad (3.32)$$

$$\hat{\mathscr{D}}\hat{q}_f + \hat{\nabla}\cdot\hat{\mathscr{J}} + [\mathscr{J}]\cdot\mathbf{n} = 0. \quad (3.33)$$

We easily notice that Eqs. (3.27) and (3.28) are immediately deduced from Maxwell's equations written in $R_c(\mathbf{x},t)$. This justifies the introduction of electromagnetic fields related to the comoving frame $R_c(\mathbf{x},t)$.

IV. THERMOELECTROMECHANICAL EQUATIONS

A. General principles in global form

The thermoelectromechanical balance laws (with the exception of Maxwell's equations recalled in Sec. III which are *not* of a mechanical nature) may be deduced in an elegant manner from three general principles written in global form for the material volume D swept out by the singular surface Σ. These are the principle of virtual power and the first and second principles of thermodynamics. We refer the reader to a review paper[3] for general features and the manner to account for electric polarization and magnetization effects in the presence, or absence, of electromagnetic ordering and to Ref. 4 for the purely mechanical case of singular surfaces and interfaces.

We call ${}^tP_a$, ${}^tP_i$, ${}^tP_v$, ${}^tP_c$, ${}^tP_e$, K, $\hat{K}$, E, $\hat{E}$, N, $\hat{N}$, U^{em}, $\dot{Q}_h$, and $\mathscr{N}$, respectively, the total power of inertia forces, the total power of internal forces, the total power of "volume" forces, the total power of contact forces, the total power of prescribed forces (of any type, in the bulk, on surfaces or on lines), the total kinetic energy of the volume (regular region), the total kinetic energy of material points belonging to the singular *thermodynamical* surface, the total internal energy of the regular material region, the total internal energy of the singular thermodynamical surface, the total entropy of the regular material region, the total surface entropy of the singular thermodynamical surface, the total electromagnetic energy of the electromagnetic fields on D-Σ on account of magnetic dipoles, the total rate of supply of heat, and the total rate of supply of entropy. In the sequel a superscript asterisk will indicate a virtual field or the value of an expression in such a field (which is not, in general, a solution of the actual problem).

1. Principle of virtual power

In a Galilean frame and for an absolute Newtonian chronology the total virtual power of inertial forces of the system balances the sum total of the powers of internal and external forces impressed on the system for any virtual velocity fields. With the above notation, this reads

$${}^tP_a^* = {}^tP_i^* + {}^tP_v^* + {}^tP_c^*. \quad (4.1)$$

2. First principle of thermodynamics

The time rate of change of the total energy contained in D-Σ and on Σ is equal to the sum of the power developed by prescribed forces, the energy supply by radiation in D-Σ and on Σ, and the flux of energy through the boundaries ∂D-Σ and $\partial\Sigma$. This may be stated as

$$\frac{d}{dt}(K+E)+\frac{\hat{d}}{dt}(\hat{K}+\hat{E})+\frac{d}{dt}U^{\text{em}}={}^{t}P_e+\dot{Q}_h\,. \tag{4.2}$$

3. Second principle of thermodynamics

For any thermodynamical process the time rate of change of the total entropy of the system is never less than the sum of the total entropy supply and the total flux of entropy through the boundaries. Mathematically, this reads

$$\frac{d}{dt}N+\frac{\hat{d}}{dt}\hat{N}\geqslant\dot{\mathscr{N}}\,. \tag{4.3}$$

For a general magnetizable, electrically polarized, heat conducting deformable conductor, the expression to be carried out in Eqs. (4.1)–(4.3) are given, or constructed, as follows:

$$\begin{aligned}{}^{t}P_a^* &= \int_{D\text{-}\Sigma}\rho\frac{dv_i}{dt}v_i^*\,dv\\ &\quad+\int_\Sigma\left\{\hat{\rho}\frac{\hat{d}\hat{v}_i}{dt}+[m(v_i-\hat{v}_i)]\right\}\hat{v}_i^*\,da\,,\end{aligned} \tag{4.4}$$

$${}^{t}P_i^*=-\int_{D-\Sigma}p_i^*\,dv-\int_\Sigma({}^{\Sigma}p_i^*+\hat{p}_i^*)da, \tag{4.5}$$

$$\begin{aligned}{}^{t}P_v^* &= \int_{D\text{-}\Sigma}\left[(f_i+f_i^{\text{em}})v_i^*+\rho\mathscr{E}_i\left(\frac{d\pi_i}{dt}\right)^*\right.\\ &\quad\left.+\rho B_i\left(\frac{d\mu_i}{dt}\right)^*\right]dv+\int_\Sigma(\hat{f}_i+\hat{f}_i^{\text{em}})\,\hat{v}_i^*\,da\,,\end{aligned} \tag{4.6}$$

$${}^{t}P_c^*=\int_{\partial D\text{-}\Sigma}(T_i+T_i^{\text{em}})v_i^*\,da+\int_{\partial\Sigma}\hat{T}_i\,\hat{v}_i^*\,dl\,, \tag{4.7}$$

$$\begin{aligned}{}^{t}P_e &= \int_{D\text{-}\Sigma}f_iv_i\,dv+\int_{\partial D\text{-}\Sigma}T_iv_i\,da\\ &\quad+\int_\Sigma\hat{f}_i\hat{v}_i\,da+\int_{\partial\Sigma}\hat{T}_i\hat{v}_i\,dl\,,\end{aligned} \tag{4.8}$$

$$\begin{aligned}\dot{Q}_h &= \int_{D\text{-}\Sigma}\rho h\,dv-\int_{\partial D\text{-}\Sigma}\mathbf{q}\cdot\mathbf{n}\,da\\ &\quad+\int_\Sigma\hat{\rho}\hat{h}\,da-\int_{\partial\Sigma}\hat{\mathbf{q}}\cdot\boldsymbol{\tau}\,dl\,,\end{aligned} \tag{4.9}$$

$$\begin{aligned}\dot{\mathscr{N}} &= \int_{D\text{-}\Sigma}\rho\sigma\,dv-\int_{\partial D\text{-}\Sigma}\boldsymbol{\phi}\cdot\mathbf{n}\,da\\ &\quad+\int_\Sigma\hat{\rho}\hat{\sigma}\,da-\int_{\partial\Sigma}\hat{\boldsymbol{\phi}}\cdot\boldsymbol{\tau}\,dl\,,\end{aligned} \tag{4.10}$$

and

$$K=\int_{D\text{-}\Sigma}\frac{1}{2}\rho\mathbf{v}^2\,dv\,,\quad \hat{K}=\int_\Sigma\frac{1}{2}\hat{\rho}\hat{\mathbf{v}}^2\,da\,, \tag{4.11}$$

$$E=\int_{D\text{-}\Sigma}\rho e\,dv,\quad \hat{E}=\int_\Sigma\hat{\rho}\hat{e}\,da\,, \tag{4.12}$$

$$N=\int_{D\text{-}\Sigma}\rho\eta\,dv,\quad \hat{N}=\int_\Sigma\hat{\rho}\hat{\eta}\,da\,, \tag{4.13}$$

$$U^{\text{em}}=\int_{D\text{-}\Sigma}\frac{1}{2}(\mathbf{E}^2+\mathbf{B}^2-2\mathscr{M}\cdot\mathbf{B})\,dv\,. \tag{4.14}$$

In Eq. (4.5) we have introduced the following quantities:

$$p_i^*=\sigma_{ij}D_{ij}^*-\rho^LE_i(D_J\pi_i)^*-\rho^LB_i(D_J\mu_i)^*\,, \tag{4.15}$$

$${}^{\Sigma}p_i^*=[\mathscr{T}_i(v_i-\hat{v}_i)^*]\,, \tag{4.16}$$

and

$$\hat{p}_i^*=\hat{\sigma}_{ij}\hat{D}_{ij}^*\,, \tag{4.17}$$

with

$$\boldsymbol{\pi}=\mathbf{P}/\rho,\quad \boldsymbol{\mu}=\mathscr{M}/\rho\,, \tag{4.18}$$

and

$$(D_J\mathbf{a})_i=\frac{da_i}{dt}-\Omega_{ij}a_j\,, \tag{4.19}$$

$$\hat{D}_{ij}=\hat{v}_{(i,j)}\equiv\tfrac{1}{2}(\hat{v}_{i,j}+\hat{v}_{j,i})\,. \tag{4.20}$$

In the above set expressions, $\mathbf{f}$ and $\hat{\mathbf{f}}$ are volume and surface forces of purely mechanical origin. Here $\mathscr{T}$ and $\mathbf{T}$ are surface tractions across the singular surface and on the boundary ∂D-Σ, while $\hat{\mathbf{T}}$ is a lineal traction. Here e and $\hat{e}$ are internal energies per unit mass, η and $\hat{\eta}$ are entropies per unit mass, and $\mathbf{q}$ and $\hat{\mathbf{q}}$ are the heat flux vectors. The symmetric tensor σ_{ij} is called the intrinsic stress tensor (not to be mistaken for the Cauchy stress tensor to which it is only a symmetric contribution[3]); $\hat{\sigma}_{ij}$ is the corresponding intrinsic surface stress tensor (an essentially two-dimensional geometric object in the absence of so-called membrane forces[4]; see, below, the transversality condition). Constitutive equations will have to be constructed for these two tensors. Here ${}^LE_i$ and ${}^LB_i$ may be referred to as the local electric field and the local magnetic field. They reflect the interactions that take place between the polarization field and the matter and the magnetization field and the matter. The quantities σ and $\hat{\sigma}$, $\boldsymbol{\phi}$ and $\hat{\boldsymbol{\phi}}$ are usually related to h, $\hat{h}$, $\mathbf{q}$, and $\hat{\mathbf{q}}$ and the thermodynamical temperatures θ and $\hat{\theta}$ ($\theta>0$, $\inf\theta=0$; $\hat{\theta}>0$, $\inf\hat{\theta}=0$), where $\hat{\theta}$ is the temperature field attributed to the surface by, e.g., $\hat{\theta}=\partial\hat{e}/\partial\hat{\eta}$. These relations will be specified later on. Finally, $\mathbf{f}^{\text{em}}$ is the volume ponderomotive force in the Galilean approximation and $\mathbf{T}^{\text{em}}$ and $\hat{\mathbf{f}}^{\text{em}}$ are the corresponding electromagnetic forces at the boundary of the body and at the interface Σ. Before specifying the latter, we refer the reader to Ref. 3 or Appendix C for the construction of expressions such as (4.6) and (4.15) and to Ref. 4 for the construction of expressions such as (4.4) and (4.16).

According to a semimicroscopic approach we have (Appendix A)[11]

$$\begin{aligned}\mathbf{f}^{\text{em}} &= q^{\text{eff}}\mathscr{E}+(1/c)\{\mathscr{J}^{\text{eff}}-c\nabla\times\mathscr{M}\}\times\mathbf{B}\\ &\quad+\operatorname{div}(\mathscr{E}\otimes\mathbf{P})+(\nabla\mathbf{B})\cdot\mathscr{M}\\ &= \operatorname{div}\mathbf{t}^{\text{em}}-\frac{\partial\mathbf{G}}{\partial t}\quad\text{in } D\text{-}\Sigma\,,\end{aligned} \tag{4.21}$$

$$\begin{aligned}\hat{f}^{\text{em}} &= \hat{q}^{\text{eff}}\hat{\mathscr{E}}+(1/c)\{\hat{\mathscr{J}}^{\text{eff}}-c\hat{\mathbf{n}}\times[\mathscr{M}]\}\times\langle\mathbf{B}\rangle\\ &\quad+[\mathscr{E}\otimes\mathbf{P}]\cdot\hat{\mathbf{n}}+(\langle\mathscr{M}\rangle\cdot[\mathbf{B}])\cdot\hat{\mathbf{n}}\\ &= [t^{\text{em}}+\mathbf{G}\otimes\hat{\mathbf{v}}]\cdot\hat{\mathbf{n}}\quad\text{across }\Sigma\,,\end{aligned} \tag{4.22}$$

$$\mathbf{T}^{\text{em}}=-(\mathbf{t}^{\text{em}}+\mathbf{G}\otimes\mathbf{v})\cdot\mathbf{n}\quad\text{on }\partial D\text{-}\Sigma\,, \tag{4.23}$$

where

$$t_{ij}^{\text{em}} = E_iE_j + B_iB_j + \mathscr{E}_i P_j - \mathscr{M}_iB_j - \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2 - 2\mathscr{M}\cdot\mathbf{B})\delta_{ij}, \tag{4.24}$$

$$\mathbf{G} = (1/c)\mathbf{E}\times\mathbf{B}, \quad \hat{\mathscr{E}} = \langle\mathbf{E}\rangle + (1/c)\hat{\mathbf{v}}\times\langle\mathbf{B}\rangle. \tag{4.25}$$

The quantity

$$C_{ij}^{\text{em}} = t_{[ij]}^{\text{em}} = \mathscr{E}_{[i} P_{j]} + B_{[i} \mathscr{M}_{j]} \tag{4.26}$$

is the tensor dual of the *ponderomotive couple*. Notice the symmetry between $\mathbf{f}^{\text{em}}$ and $\hat{\mathbf{f}}^{\text{em}}$ in Eqs. (4.21) and (4.22).

B. Local electromechanical equations

On assuming that Eq. (4.1) remains good for any of the virtual fields, $\mathbf{v}^*$, $\hat{\mathbf{v}}^*$, $\hat{\mathbf{v}}^{\pm *}$ (at Σ), $D\hat{\mathbf{v}}^*$, $(d\boldsymbol{\pi}/dt)^*$, and $(d\boldsymbol{\mu}/dt)^*$ and any element of volume, surface, and line, we obtain at once, after use of the generalized Stokes' theorem accounting for the presence of a discontinuity surface (Ref. 4)—see Appendix D—the following local equations:

$$\rho\frac{dv_i}{dt} = t_{ij,j} + f_i + f_i^{\text{em}} \quad \text{in } D\text{-}\Sigma, \tag{4.27}$$

$$T_i + T_i^{\text{em}} = t_{ij}\, n_j \quad \text{on } \partial D\text{-}\Sigma, \tag{4.28}$$

$$\hat{\rho}\frac{\hat{d}\hat{v}_i}{dt} + [m(v_i - \hat{v}_i)] = [\mathscr{T}_i] + (\hat{\nabla}_j + 2\Omega\hat{n}_j)\hat{\sigma}_{ij} + \hat{f}_i + \hat{f}_i^{\text{em}} \quad \text{on } \Sigma, \tag{4.29}$$

$$\hat{T}_i = \hat{\sigma}_{ij}\tau_j \quad \text{along } \partial\Sigma, \tag{4.30}$$

$$\mathscr{T}_i^{\pm} = t_{ij}^{\pm}\, \hat{n}_j \ \text{ on } \Sigma^{\pm}, \quad \hat{\sigma}_{ij}\hat{n}_j = 0 \quad \text{on } \Sigma, \tag{4.31}$$

$${}^LE_i + \mathscr{E}_i = 0, \quad {}^LB_i + B_i = 0 \quad \text{in } D\text{-}\Sigma. \tag{4.32}$$

The *a priori* nonsymmetric Cauchy stress tensor t_{ij} has been defined by

$$t_{ij} = \sigma_{ij} + {}^LE_{[i}\, P_{j]} + {}^LB_{[i}\, \mathscr{M}_{j]}. \tag{4.33}$$

The local statement of the balance of angular momentum is contained in the above mentioned equation. By taking the skew-symmetric part of Eq. (4.33) we obtain

$$t_{[ij]} = {}^LE_{[i}P_{j]} + {}^LB_{[i}\mathscr{M}_{j]}. \tag{4.34}$$

On using Eqs. (4.31) we can rewrite Eq. (4.29) in the following more conventional form:

$$\hat{\rho}\frac{\hat{d}\hat{v}_i}{dt} + [m(v_i - \hat{v}_i)] = [t_{ij}]\hat{n}_j + \hat{\nabla}_j\hat{\sigma}_{ij} + \hat{f}_i + \hat{f}_i^{\text{em}} \quad \text{on } \Sigma. \tag{4.35}$$

C. Local thermodynamical equations

If we combine Eq. (4.1) written for *real* velocity fields (no asterisks), with Eq. (4.2) and account for the demonstrable electromagnetic energy identity (see Appendix B),

$$\frac{d}{dt}U^{\text{em}} = -\int_{D\text{-}\Sigma}\left\{\mathbf{f}^{\text{em}}\cdot\mathbf{v} + \rho\mathscr{E}\frac{d\boldsymbol{\pi}}{dt} + \rho\mathbf{B}\frac{d\boldsymbol{\mu}}{dt} + \mathscr{J}\cdot\mathscr{E}\right\}dv - \int_{\partial D\text{-}\Sigma}(\mathbf{T}^{\text{em}}\cdot\mathbf{v} + \mathscr{S}\cdot\mathbf{n})da - \int_{\Sigma}(\hat{\mathbf{f}}^{\text{em}}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + [\tau_{ij}^{\text{em}}(v_i - \hat{v}_i)]\hat{n}_j)da, \tag{4.36}$$

with

$$\mathscr{S} = c\mathscr{E}\times\mathscr{H}, \tag{4.37}$$

$$\tau_{ij}^{\text{em}} = t_{ij}^{\text{em}} + G_i\hat{v}_j - {}^{\Sigma}\mathscr{S}_{ij}^d - \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2 - 2\mathscr{M}\cdot\mathbf{B})\delta_{ij},$$
$${}^{\Sigma}\mathscr{S}_{ij}^d = \langle\mathscr{E}_i\rangle D_j + \langle\mathscr{H}_i\rangle B_j - (\langle\mathscr{E}\rangle\cdot\mathbf{D} + \langle\mathscr{H}\rangle\cdot\mathbf{B})\delta_{ij}, \tag{4.38}$$

where the first of these is Poynting's vector in $R_c(\mathbf{x},t)$, we obtain the following global expression for the so-called energy theorem:

$$\frac{d}{dt}E + \frac{\hat{d}}{dt}\hat{E} + \dot{K}_{\text{ex}}(\Sigma) + {}^tP_i = \dot{Q}_h + \dot{Q}_{\text{em}}, \tag{4.39}$$

where we have set

$$\dot{K}_{\text{ex}}(\Sigma) = \left(\frac{dK}{dt} + \frac{\hat{d}\hat{K}}{dt}\right) - {}^tP_a = \int_{\Sigma}[(m/2)(\mathbf{v} - \hat{\mathbf{v}})^2]da, \tag{4.40}$$

$$\dot{Q}_{\text{em}} = \int_{D\text{-}\Sigma}\mathscr{J}\cdot\mathscr{E}\, dv + \int_{\partial D\text{-}\Sigma}\mathscr{S}\cdot\mathbf{n}\, da + \int_{\Sigma}\hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle da + \int_{\Sigma}[\tau_{ij}^{\text{em}}(v_i - \hat{v}_i)]\hat{n}_j\, da. \tag{4.41}$$

The quantity defined by Eq. (4.40) is the so-called excess rate of kinetic energy.[4]

Accounting for generalized transport theorems and balances of mass, from Eq. (4.39) we deduce the local forms of the energy theorems as

$$\rho\frac{de}{dt} = p_i + \mathscr{J}\cdot\mathscr{E} + \rho h - \nabla\cdot\tilde{\mathbf{q}} \quad \text{in } D\text{-}\Sigma, \tag{4.42}$$

$$\hat{\rho}\frac{\hat{d}\hat{e}}{dt} = \hat{p}_i + {}^{\Sigma}p_i + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + \hat{\rho}\hat{h} - \hat{\nabla}\cdot\hat{q} + [\tau_{ij}^{\text{em}}(v_i - \hat{v}_i) - \hat{q}_j]\hat{n}_j - [m\{(e - \hat{e}) + \tfrac{1}{2}(\mathbf{v} - \hat{\mathbf{v}})^2\}] \quad \text{on } \Sigma, \tag{4.43}$$

where we assumed the transversality condition

$$\hat{\mathbf{q}}\cdot\hat{\mathbf{n}} = 0 \tag{4.44}$$

and we have set

$$\tilde{\mathbf{q}} = \mathbf{q} - \mathscr{S}. \tag{4.45}$$

It remains to exploit the global inequality (4.3). To that purpose we assume that σ, $\hat{\sigma}$, $\boldsymbol{\phi}$, and $\hat{\boldsymbol{\phi}}$ are given by

$$\sigma = h/\theta, \quad \hat{\sigma} = \hat{h}/\hat{\theta}, \quad \boldsymbol{\phi} = \tilde{\mathbf{q}}/\theta, \quad \hat{\boldsymbol{\phi}} = \hat{\mathbf{q}}/\hat{\theta}. \tag{4.46}$$

Only the volume entropy flux differs from the usual ratio of the heat vector to the temperature, which means that nonsimple thermodynamical processes are involved (compare Ref. 6, p. 129). Accounting for (4.46), the local forms of (4.3) read

$$\rho\theta\frac{d\eta}{dt} > \rho h - \nabla\cdot\tilde{\mathbf{q}} + \boldsymbol{\phi}\cdot\nabla\theta \quad \text{in } D\text{-}\Sigma, \tag{4.47}$$

$$\hat{\rho}\hat{\theta}\frac{\hat{d}\hat{\eta}}{dt} > \hat{\rho}\hat{h} - \hat{\nabla}\cdot\hat{\mathbf{q}} + \hat{\boldsymbol{\phi}}\cdot\hat{\nabla}\hat{\theta} - [m(\eta - \hat{\eta})\hat{\theta}] - [\hat{\theta}\boldsymbol{\phi}]\cdot\hat{\mathbf{n}} \quad \text{on } \Sigma. \tag{4.48}$$

D. The Clausius–Duhem inequality

Introducing the Helmholtz free energy densities Ψ and $\hat{\Psi}$ by

$$\Psi = e - \eta\theta, \quad \hat{\Psi} = \hat{e} - \hat{\eta}\hat{\theta}, \tag{4.49}$$

and accounting for Eqs. (4.42), (4.43), (4.15)–(4.17), (4.31), and (4.33), from (4.47) and (4.48) we are led to the Clausius–Duhem inequalities in the form

$$-\rho\left(\frac{d\Psi}{dt}+\eta\,\frac{d\theta}{dt}\right)+t_{ij}v_{i,j}+\rho\mathscr{E}_i\,\frac{d\pi_i}{dt}+\rho B_i\,\frac{d\mu_i}{dt}$$
$$+\mathscr{J}\cdot\mathscr{E}-\boldsymbol{\phi}\cdot\nabla\theta\geqslant 0\quad\text{in } D\text{-}\Sigma \tag{4.50}$$

and

$$-\hat{\rho}\left(\frac{\hat{d}\hat{\Psi}}{dt}+\hat{\eta}\,\frac{\hat{d}\hat{\theta}}{dt}\right)+\hat{\sigma}_{ij}\hat{D}_{ij}+\langle\mathscr{E}\rangle\cdot\hat{\mathscr{J}}-\hat{\boldsymbol{\phi}}\cdot\hat{\nabla}\hat{\theta}$$
$$+[(t_{ij}+\tau^{\rm em}_{ij})(v_i-\hat{v}_i)-\phi_j(\theta-\hat{\theta})]\hat{n}_j$$
$$-[m\{\eta(\theta-\hat{\theta})+(\Psi-\hat{\Psi})+\tfrac{1}{2}(\mathbf{v}-\hat{\mathbf{v}})^2\}]\geqslant 0\,. \tag{4.51}$$

Finally, if we perform the following Legendre transformation on Ψ:

$$\tilde{\Psi}=\Psi-\mathscr{E}\cdot\boldsymbol{\pi}-\mathbf{B}\cdot\boldsymbol{\mu}\,, \tag{4.52}$$

define the new symmetric stress tensor $\tilde{t}_{ij}$ by

$$\tilde{t}_{ij}=\sigma_{ij}-P_{(i}\mathscr{E}_{j)}-\mathscr{M}_{(i}B_{j)}+(\mathbf{P}\cdot\mathscr{E}+\mathscr{M}\cdot\mathbf{B})\delta_{ij}$$
$$=t_{ij}-P_i\mathscr{E}_j-\mathscr{M}_iB_j+(\mathbf{P}\cdot\mathscr{E}+\mathscr{M}\cdot\mathbf{B})\delta_{ij}\,, \tag{4.53}$$

and use the definition (3.15) of a convected time derivative, we can rewrite the volume inequality (4.50) in the useful equivalent forms

$$-\rho\left(\frac{d\tilde{\Psi}}{dt}+\eta\,\frac{d\theta}{dt}\right)+t_{ij}v_{i,j}-P_i\,\frac{d\mathscr{E}_i}{dt}$$
$$-\mathscr{M}_i\,\frac{dB_i}{dt}+\mathscr{E}\cdot\mathscr{J}-\frac{\tilde{\mathbf{q}}}{\theta}\cdot\nabla\theta\geqslant 0\,, \tag{4.54}$$

or

$$-\rho\left(\frac{d\tilde{\Psi}}{dt}+\eta\,\frac{d\theta}{dt}\right)+\tilde{t}_{ij}D_{ij}-P_i\,\overset{*}{\mathscr{E}}_i$$
$$-\mathscr{M}_i\overset{*}{B}_i+\mathscr{E}\cdot\mathscr{J}-\frac{\tilde{\mathbf{q}}}{\theta}\cdot\nabla\theta\geqslant 0\,. \tag{4.55}$$

On account of the second equation of (4.31) and after some decompositions, Eq. (4.51) yields

$$-\rho\left(\frac{\hat{d}\hat{\Psi}}{dt}+\eta\,\frac{\hat{d}\hat{\theta}}{dt}\right)+\hat{\sigma}_{ij}\hat{\hat{D}}_{ij}+\langle\mathscr{E}\rangle\cdot\hat{\mathscr{J}}-\frac{\hat{\mathbf{q}}}{\hat{\theta}}\cdot\hat{\nabla}\hat{\theta}$$
$$+\hat{\theta}\,\langle\tilde{q}_{(\hat{n})}\rangle[1/\theta]+\hat{\theta}\,[\tilde{q}_{(\hat{n})}](\langle 1/\theta\rangle-1/\hat{\theta})$$
$$-\langle{}^T\tau_{(\hat{n})}\rangle\cdot[\mathbf{v}]+[{}^T\tau_{(\hat{n})}]\cdot(\langle\mathbf{v}\rangle-\hat{\mathbf{v}})\geqslant 0\,, \tag{4.56}$$

where we have set

$$\hat{\hat{D}}_{ij}=\hat{\nabla}_{(j}\hat{v}_{i)}\,,$$
$$q_{(\hat{n})}=\mathbf{q}\cdot\hat{\mathbf{n}}\,,\quad {}^T\tau_{(\hat{n})}={}^T\boldsymbol{\tau}\cdot\hat{\mathbf{n}}\,, \tag{4.57}$$

and

$${}^T\tau_{ij}=t_{ij}+\tau^{\rm em}_{ij}-\Delta\delta_{ij}\,, \tag{4.58}$$
$$\Delta=\rho\{\eta(\theta-\hat{\theta})+(\Psi-\hat{\Psi})+\tfrac{1}{2}(\mathbf{v}-\hat{\mathbf{v}})^2\}\,. \tag{4.59}$$

If we consider the case where the only quantities attached to the interface are of electric nature such as $\hat{q}_f$, $\hat{\mathscr{J}}$, on setting $\hat{\mathbf{v}}=\boldsymbol{\nu}$ and $\hat{\theta}^{-1}=\langle\theta^{-1}\rangle$, see Ref. 4, we have the following balance laws: Conservation of mass,

$$\frac{d\rho}{dt}+\rho\nabla\cdot\mathbf{v}=0\quad\text{in } D\text{-}\Sigma\,, \tag{4.60}$$
$$[\rho(\mathbf{v}-\boldsymbol{\nu})]\cdot\hat{\mathbf{n}}=0\quad\text{across } \Sigma(t); \tag{4.61}$$

balance of momentum,

$$\rho\,\frac{dv_i}{dt}=t_{ij,j}+f_i+f^{\rm em}_i\quad\text{in } D\text{-}\Sigma\,, \tag{4.62}$$
$$[\rho v_i(v_j-\nu_j)-t_{ij}-(t^{\rm em}_{ij}+G_i\nu_j)]\hat{n}_j=0\quad\text{across } \Sigma(t); \tag{4.63}$$

conservation of energy,

$$\rho\,\frac{de}{dt}=t_{ij}v_{i,j}+\rho\mathscr{E}\cdot\frac{d\boldsymbol{\pi}}{dt}+\rho\mathbf{B}\cdot\frac{d\boldsymbol{\mu}}{dt}$$
$$+\mathscr{J}\cdot\mathscr{E}-\nabla\cdot\tilde{\mathbf{q}}+\rho h\quad\text{in } D\text{-}\Sigma\,, \tag{4.64}$$
$$[\{\tfrac{1}{2}\rho v^2+\rho e+\tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2-2\mathscr{M}\cdot\mathbf{B})\}\{v_j-\nu_j\}$$
$$-(t_{ij}+t^{\rm em}_{ij}+G_i\nu_j)v_i+\mathscr{S}_j+\tilde{q}_j]\hat{n}_j=0\quad\text{across } \Sigma\,; \tag{4.65}$$

and entropy inequality

$$\rho\,\frac{d\eta}{dt}\geqslant\frac{1}{\theta}\,(\rho h-\nabla\cdot\tilde{\mathbf{q}})-\tilde{\mathbf{q}}\cdot\nabla\left(\frac{1}{\theta}\right)\quad\text{in } D\text{-}\Sigma\,, \tag{4.66}$$
$$[\rho\eta(v_j-\nu_j)+(1/\theta)\tilde{q}_j]\hat{n}_j\geqslant 0\quad\text{across } \Sigma\,. \tag{4.67}$$

The above-mentioned balance laws are the ones obtained by Maugin and Eringen from the vectorial approach [see Eqs. (5.8)–(5.16) in Ref. 5]. In order to have the same notation, we must replace e by $e+\boldsymbol{\mu}\cdot\mathbf{B}$ and $\tilde{\mathbf{q}}$ by $-\mathbf{q}$.

V. CONSTITUTIVE EQUATIONS

For illustration purposes and further comparison with other works it is salient to consider the special case of electromagnetic fluids, where we assume *a priori* that all dependent functions $\tilde{\Psi}$, η, t_{ij}, P_i, $\mathscr{M}_i$, $\mathscr{J}_i$, and $\tilde{q}_i$ may depend on the *same set* of variables (Ref. 6)

$$\rho^{-1},D_{ij},\theta,\nabla\theta,\mathscr{E}_i,B_i\,, \tag{5.1}$$

A similar assumption is used at the interface where the set (5.1) is replaced by

$$\hat{\rho}^{-1},\hat{\hat{D}}_{ij},\hat{\theta},\hat{\nabla}\hat{\theta}\,. \tag{5.2}$$

Thus $\tilde{\Psi}$ is assumed to depend on the set (5.1) while $\hat{\Psi}$ depends on the set (5.2). On computing $d\tilde{\Psi}/dt$ and $\hat{d}\hat{\Psi}/dt$ and carrying the results in Eqs. (4.54) and (4.56) that must be satisfied for all independent thermodynamical processes $[d\theta/dt,\ \Omega_{ij},\ (d/dt)\ (D_{ij}),\ (d/dt)\ (\nabla\theta),\ (d/dt)\ \mathscr{E}_i,\ (d/dt)\ B_j]$, we obtain the following.

(i) In D-Σ

$$\frac{\partial\tilde{\Psi}}{\partial(D_{ij})}=0,\quad\frac{\partial\tilde{\Psi}}{\partial(\nabla\theta)}=0,\quad t_{[ij]}=0\,, \tag{5.3}$$

and

$$\eta=-\frac{\partial\tilde{\Psi}}{\partial\theta},\quad P_i=-\rho\,\frac{\partial\tilde{\Psi}}{\partial\mathscr{E}_i},\quad\mathscr{M}_i=-\rho\,\frac{\partial\tilde{\Psi}}{\partial B_i}, \tag{5.4}$$

while there remains the following dissipation inequality:

$$\Gamma\equiv{}^Dt_{ij}D_{ij}-(1/\theta)\tilde{\mathbf{q}}\cdot\nabla\theta+\mathscr{J}\cdot\mathscr{E}\geqslant 0\,, \tag{5.5}$$

with

$${}^Dt_{ij}=t_{ij}+p\delta_{ij},\quad p=-\frac{\partial\tilde{\Psi}}{\partial\rho^{-1}}\,. \tag{5.6}$$

From the first two equations of (5.3) we see that $\tilde{\Psi}$ is independent of $\mathbf{D}$ and $\nabla\theta$, while the third one states that the stress tensor must be symmetric. If we recall the definition of t_{ij} we see that we must have

$$P_{[i}\mathscr{E}_{j]} + \mathscr{M}_{[i}B_{j]} = 0. \quad (5.7)$$

(ii) On Σ

$$\frac{\partial\hat{\Psi}}{\partial(\hat{\hat{D}}_{ij})} = 0, \quad \frac{\partial\hat{\Psi}}{\partial(\hat{\nabla}\hat{\theta})} = 0, \quad \hat{\eta} = -\frac{\partial\hat{\Psi}}{\partial\hat{\theta}}, \quad (5.8)$$

and

$$\hat{\Gamma} \equiv {}^D\hat{\sigma}_{ij}\hat{\hat{D}}_{ij} - (1/\hat{\theta})\,\hat{\mathbf{q}}\cdot\hat{\nabla}\hat{\theta} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + \hat{\theta}\{\langle\tilde{q}_{(\hat{n})}\rangle[1/\theta] + [\tilde{\mathbf{q}}_{(\hat{n})}](\langle 1/\theta\rangle - 1/\hat{\theta})\} + \langle{}^T\tilde{\boldsymbol{\tau}}_{(\hat{n})}\rangle\cdot[\mathbf{v}] + [{}^T\tilde{\boldsymbol{\tau}}_{(\hat{n})}](\langle\mathbf{v}\rangle - \hat{\mathbf{v}}) \geqslant 0, \quad (5.9)$$

with

$${}^D\hat{\sigma}_{ij} = \hat{\sigma}_{ij} + \hat{p}P_{ij}, \quad \hat{p} = -\frac{\partial\hat{\Psi}}{\partial\hat{\rho}^{-1}}, \quad (5.10)$$

$${}^T\tilde{\tau}_{ij} = {}^T\tau_{ij} + (\rho/\hat{\rho})\,\hat{p}\delta_{ij}. \quad (5.11)$$

After decomposing Eqs. (5.5) and (5.9) into traceless tensors, vectors, and scalars, we are led to

$$\Gamma \equiv {}^D_0t_{ij\,0}D_{ij} - (1/\theta)\tilde{\mathbf{q}}\cdot\nabla\theta + \mathscr{J}\cdot\mathscr{E} + {}^DtD_{kk} \geqslant 0, \quad (5.12)$$

$$\hat{\Gamma} \equiv {}^D_0\hat{\sigma}_{ij\,0}\hat{\hat{D}}_{ij} - (1/\hat{\theta})\,\hat{\mathbf{q}}\cdot\hat{\nabla}\hat{\theta} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + {}^D\hat{\sigma}\hat{\hat{D}}_{kk} + \langle{}^T\tilde{\boldsymbol{\tau}}_{(\hat{n})}\rangle_{\tan}\cdot[\mathbf{v}]_{\tan} + [{}^T\tilde{\boldsymbol{\tau}}_{(\hat{n})}]_{\tan}\cdot(\langle\mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan} + \langle\tau_{(\hat{n}\hat{n})}\rangle\cdot[v_{(\hat{n})}] + [\tau_{(\hat{n}\hat{n})}]\cdot(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}) + \hat{\theta}\{\langle\tilde{q}_{(\hat{n})}\rangle[1/\theta] + [\tilde{q}_{(\hat{n})}](\langle 1/\theta\rangle - 1/\hat{\theta})\} \geqslant 0, \quad (5.13)$$

with

$$A_{ij} = {}_0A_{ij} + A\delta_{ij}, \quad {}_0A_{kk} = 0, \quad A = \tfrac{1}{3}A_{kk},$$
$$\hat{A}_{ij} = {}_0\hat{A}_{ij} + \hat{A}P_{ij}, \quad {}_0\hat{A}_{kk} = 0, \quad \hat{A} = \tfrac{1}{2}\hat{A}_{kk}, \quad (5.14)$$

$$\hat{\mathbf{A}} = \hat{\mathbf{A}}_{\tan} + \hat{A}_{(\hat{n})}\hat{\mathbf{n}}, \quad (5.15)$$

where the subscript tan indicates the tangential component and

$$\tau_{(\hat{n}\hat{n})} \equiv \tau_{ij}\hat{n}_i\hat{n}_j. \quad (5.16)$$

For simplicity we consider linear relations between the generalized fluxes and thermodynamical forces occurring in the entropy production. Taking also into account the tensorial nature of the various quantities, the two-dimensional isotropy on the surface, and the isotropy in the bulk we obtain the following linear laws.

In D-Σ,

$${}^D_0t_{ij} = 2\mu_0D_{ij}, \quad {}^Dt = kD_{kk}, \quad (5.17)$$

$$-\tilde{\mathbf{q}} = K_1\nabla\theta + K_2\mathscr{E}, \quad (5.18)$$

$$\mathscr{J} = \sigma_1\mathscr{E} + \sigma_2\nabla\theta. \quad (5.19)$$

On using the first equation of (5.14) with (5.17) we obtain a more conventional form for the dissipative stress ${}^D\mathbf{t}$ as follows:

$${}^Dt_{ij} = \lambda D_{kk}\delta_{ij} + 2\mu D_{ij}, \quad (5.20)$$

with the following relation:

$$k = 3\lambda + 2\mu. \quad (5.21)$$

From the entropy inequality (5.12) it is clear that

$$K_1 \geqslant 0, \quad \sigma_1 \geqslant 0, \quad 4K_1\sigma_1\theta^{-1} - (K_2\theta^{-1} + \sigma_2)^2 \geqslant 0,$$
$$k = 3\lambda + 2\mu \geqslant 0, \quad \mu \geqslant 0. \quad (5.22)$$

On Σ we have (i) the traceless part of the tensorial quantities

$${}_0\hat{\sigma}_{ij} = 2\hat{\mu}_0\hat{\hat{D}}_{ij}; \quad (5.23)$$

(ii) the surface vectorial quantities

$$-\hat{\mathbf{q}} = \hat{K}_1\hat{\nabla}\hat{\theta} + \hat{K}_2\langle\mathscr{E}\rangle_{\tan} + \hat{K}_3[\mathbf{v}]_{\tan} + \hat{K}_4(\langle\mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan},$$
$$\hat{\mathscr{J}} = \hat{\sigma}_1\langle\mathscr{E}\rangle_{\tan} + \hat{\sigma}_2\hat{\nabla}\hat{\theta} + \hat{\sigma}_3[\mathbf{v}]_{\tan} + \hat{\sigma}_4(\langle\mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan},$$
$$\langle{}^T\tilde{\boldsymbol{\tau}}_{(\hat{n})}\rangle = \hat{\alpha}_1[\mathbf{v}]_{\tan} + \hat{\alpha}_2\hat{\nabla}\hat{\theta} + \hat{\alpha}_3\langle\mathscr{E}\rangle_{\tan} + \hat{\alpha}_4(\langle\mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan},$$
$$[{}^T\tilde{\boldsymbol{\tau}}_{(\hat{n})}] = \hat{\beta}_1(\langle\mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan} + \hat{\beta}_2\hat{\nabla}\hat{\theta} + \hat{\beta}_3\langle\mathscr{E}\rangle_{\tan} + \hat{\beta}_4[\mathbf{v}]_{\tan}; \quad (5.24)$$

and (iii) the scalar quantities,

$${}^D\hat{\sigma} = \hat{k}_1\hat{\hat{D}}_{kk} + \hat{k}_2[1/\theta] + \hat{k}_3(\langle 1/\theta\rangle - 1/\hat{\theta}) + \hat{k}_4[v_{(\hat{n})}] + \hat{k}_5(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}),$$
$$\langle\tilde{q}_{(\hat{n})}\rangle = \hat{l}_1[1/\theta] + \hat{l}_2\hat{\hat{D}}_{kk} + \hat{l}_3(\langle 1/\theta\rangle - 1/\hat{\theta}) + \hat{l}_4[v_{(\hat{n})}] + \hat{l}_5(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}),$$
$$[\tilde{q}_{(\hat{n})}] = \hat{m}_1(\langle 1/\theta\rangle - 1/\hat{\theta}) + \hat{m}_2\hat{\hat{D}}_{kk} + \hat{m}_3[1/\theta] + \hat{m}_4[v_{(\hat{n})}] + \hat{m}_5(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}), \quad (5.25)$$
$$\langle{}^T\tilde{\tau}_{(\hat{n}\hat{n})}\rangle = \hat{r}_1[v_{(\hat{n})}] + \hat{r}_2\hat{\hat{D}}_{kk} + \hat{r}_3[1/\theta] + \hat{r}_4(\langle 1/\theta\rangle - 1/\hat{\theta}) + \hat{r}_5(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}),$$
$$[{}^T\tilde{\tau}_{(\hat{n}\hat{n})}] = \hat{s}_1(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}) + \hat{s}_2\hat{\hat{D}}_{kk} + \hat{s}_3[1/\theta] + \hat{s}_4(\langle 1/\theta\rangle - 1/\hat{\theta}) + \hat{s}_5[v_{(\hat{n})}].$$

On combining the first equation of (5.25) and accounting for the second equation of (5.14) we get

$${}^D\hat{\sigma}_{ij} = \hat{\lambda}\hat{\hat{D}}_{kk}P_{ij} + 2\hat{\mu}\hat{\hat{D}}_{ij} + \hat{k}_2[1/\theta] + k_3(\langle 1/\theta\rangle - 1/\hat{\theta}) + \hat{k}_4[v_{(\hat{n})}] + \hat{k}_5(\langle v_{(\hat{n})}\rangle - \hat{v}_{(\hat{n})}), \quad (5.26)$$

with

$$\hat{k}_1 = \hat{\lambda} + \hat{\mu}. \quad (5.27)$$

In the same manner as in the bulk, the entropy inequality (5.13) places restrictions, which we do not discuss here, on the scalar coefficients introduced in (5.23)–(5.25).

In some particular situations such as when we have no surface mass density ($\hat{\rho} = 0$), it is necessary to define the energy and entropy per unit surface, hence on setting

$${}^\Sigma\Psi = \hat{\rho}\hat{\Psi}, \quad {}^\Sigma\eta = \hat{\rho}\hat{\eta}, \quad (5.28)$$

and accounting for mass conservation we get the following transformation:

$$\hat{\rho}\frac{\hat{d}\hat{\Psi}}{dt} - [m]\hat{\Psi} = \frac{\hat{d}\,{}^\Sigma\Psi}{dt} + {}^\Sigma\Psi P_{ij}\hat{D}_{ij} = \frac{\hat{d}\,{}^\Sigma\Psi}{dt} + {}^\Sigma\Psi\hat{\hat{D}}_{kk}, \quad (5.29)$$

and the Clausius–Duhem inequalities (4.51) and (4.56) take on the following alternate forms:

$$-\left(\frac{\hat{d}\,{}^\Sigma\Psi}{dt} + {}^\Sigma\eta\frac{\hat{d}\hat{\theta}}{dt}\right) + (\hat{\sigma}_{ij} - {}^\Sigma\Psi P_{ij})\hat{D}_{ij} + \langle\mathscr{E}\rangle\cdot\hat{\mathscr{J}} - \hat{\boldsymbol{\phi}}\cdot\hat{\nabla}\hat{\theta}$$

$$+ [(t_{ij} + \tau_{ij}^{em})(v_i - \hat{v}_i) - \phi_j(\theta - \hat{\theta})]\hat{n}_j$$
$$- [m\{\eta(\theta - \hat{\theta}) + \Psi + \tfrac{1}{2}(\mathbf{v} - \mathbf{v})^2\}] \geqslant 0 \quad (5.30)$$

and

$$-\left(\frac{\hat{d}\,{}^{\Sigma}\Psi}{dt} + {}^{\Sigma}\eta\,\frac{\hat{d}\hat{\theta}}{dt}\right) + \hat{\sigma}_{ij}\hat{D}_{ij} + \langle \mathscr{E} \rangle\cdot\hat{\mathscr{J}} - (\hat{\mathbf{q}}/\hat{\theta})\cdot\hat{\nabla}\hat{\theta}$$
$$+ \hat{\theta}\,\langle \tilde{q}_{(\hat{n})}\rangle[1/\theta] + \theta\,[\tilde{q}_{(\hat{n})}](\langle 1/\theta \rangle - 1/\hat{\theta})$$
$$\langle {}^{T}\bar{\boldsymbol{\tau}}_{(\hat{n})}\rangle\cdot[\mathbf{v}] + [{}^{T}\bar{\boldsymbol{\tau}}_{(\hat{n})}]\cdot(\langle \mathbf{v}\rangle - \hat{\mathbf{v}}) \geqslant 0\,, \quad (5.31)$$

$${}^{d}\hat{\sigma}_{ij} = \hat{\sigma}_{ij} + {}^{\Sigma}pP_{ij}, \quad {}^{\Sigma}p = -\,\hat{\rho}\hat{\Psi} = -\,{}^{\Sigma}\Psi\,, \quad (5.32)$$

$${}^{T}\bar{\tau}_{ij} = {}^{T}\tau_{ij} + \frac{\rho}{\hat{\rho}}\,{}^{\Sigma}p\delta_{ij}\,. \quad (5.33)$$

Notice the perfect symmetry between (5.32) and (5.33) and (5.10) and (5.11). Although Eq. (5.33) seems to depend on $\hat{\rho}$, if we decompose ${}^{T}\tau_{ij}$ with the help of Eqs. (4.58) and (4.59) and the first equation of (5.32), we immediately notice that ${}^{T}\bar{\tau}_{ij}$ is independent of $\hat{\rho}$. Equation (5.33) is given in the above-mentioned form only to allow for the analogy with (5.11).

As to Eqs. (5.2), (5.8), and (5.9), they will be replaced, respectively, by

$$\hat{\hat{D}}_{ij}, \quad \hat{\theta}, \quad \hat{\nabla}\hat{\theta}\,, \quad (5.34)$$

$$\frac{\partial\,{}^{\Sigma}\Psi}{\partial\hat{\hat{D}}_{ij}} = 0, \quad \frac{\partial\,{}^{\Sigma}\Psi}{\partial(\hat{\nabla}\hat{\theta})} = 0, \quad {}^{\Sigma}\eta = -\frac{\partial\,{}^{\Sigma}\Psi}{\partial\hat{\theta}}\,, \quad (5.35)$$

and

$$\hat{\gamma} \equiv {}^{d}\hat{\sigma}_{ij}\hat{\hat{D}}_{ij} - (\hat{\mathbf{q}}/\hat{\theta})\cdot\hat{\nabla}\hat{\theta} + \hat{\mathscr{J}}\cdot\langle \mathscr{E}\rangle + \hat{\theta}\,\langle q_{(\hat{n})}\rangle[1/\theta]$$
$$+ \hat{\theta}\,[q_{(\hat{n})}](\langle 1/\theta \rangle - 1/\hat{\theta})$$
$$+ \langle {}^{T}\bar{\boldsymbol{\tau}}_{(\hat{n})}\rangle\cdot[\mathbf{v}] + [{}^{T}\bar{\boldsymbol{\tau}}_{(\hat{n})}]\cdot(\langle \mathbf{v}\rangle - \hat{\mathbf{v}}) \geqslant 0\,. \quad (5.36)$$

Notice that Eqs. (5.35) and (5.36) are similar to (5.8) and (5.9) where $\hat{\Psi}$, ${}^{D}\hat{\sigma}_{ij}$, and ${}^{T}\hat{\tau}_{ij}$ have been replaced, respectively, by ${}^{\Sigma}\Psi$, ${}^{d}\hat{\sigma}_{ij}$, and ${}^{T}\bar{\tau}_{ij}$. The subsequent transformations and the construction of linear dissipative constitutive laws lead formally to the same results as (5.23)–(5.27), and hence to the same conclusions. This new formulation allows one to treat the case where we have no surface mass density, without the ambiguity concerning the definition of the surface energy density. However, we have lost some symmetry between the equations obtained in the bulk and at the interface.

Comparison with previous works: The bulk constitutive equations (5.17)–(5.22) are exactly the same as the ones derived in Eqs. (10.24.5)–(10.24.8) and (10.24.13) and (10.24.14) of Ref. 6, while the surface dissipative equations (5.23)–(5.27) generalize the expressions obtained for a purely mechanical interface with no mass transfer in Ref. 4.

Indeed, if we disregard electromagnetic effects and mass transfer ($m = 0$, i.e., $v^{+}_{(\hat{n})} = \hat{v}_{(\hat{n})} = v^{-}_{(\hat{n})}$), which would be the case of an immiscible single-component nonelectromagnetic fluid, the dissipative part of the Clausius–Duhem inequality at the interface reduces to

$$\hat{\gamma} \equiv {}^{d}_{0}\hat{\sigma}_{ij\,0}\hat{\hat{D}}_{ij} - (\hat{q}/\hat{\theta})\cdot\hat{\nabla}\hat{\theta} + \langle \boldsymbol{\sigma}_{(\hat{n})}\rangle_{\tan}\cdot[\mathbf{v}]_{\tan}$$
$$+ [\boldsymbol{\sigma}_{(\hat{n})}]_{\tan}\cdot(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan} + {}^{d}\hat{\sigma}\hat{\hat{D}}_{kk}$$
$$+ \hat{\theta}\,\langle q_{(\hat{n})}\rangle[1/\theta] + \hat{\theta}[q_{(\hat{n})}](\langle 1/\theta \rangle - 1/\hat{\theta}) \geqslant 0\,. \quad (5.37)$$

where we have set

$${}^{d}_{0}\hat{\sigma}_{ij} = {}^{d}\hat{\sigma}_{ij} - {}^{d}\hat{\sigma}P_{ij}, \quad {}^{d}\hat{\sigma} = \tfrac{1}{2}\,(\mathrm{tr}\,{}^{d}\hat{\boldsymbol{\sigma}}) \quad (5.38)$$

and

$$\boldsymbol{\sigma}_{(\hat{n})} \equiv \boldsymbol{\sigma}\cdot\hat{\mathbf{n}}\,.$$

On setting up linear relations between the fluxes and thermodynamical forces occurring in the entropy production (5.37) we get

$${}^{d}\hat{\sigma}_{ij} = 2\,{}^{\Sigma}\mu_0\hat{\hat{D}}_{ij}\,, \quad (5.39)$$

$$-\,\hat{\mathbf{q}} = {}^{\Sigma}K_1\hat{\nabla}\hat{\theta} + {}^{\Sigma}K_2[\mathbf{v}]_{\tan} + {}^{\Sigma}K_3(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan}\,,$$
$$\langle \boldsymbol{\sigma}_{(\hat{n})}\rangle_{\tan} = {}^{\Sigma}\alpha_1[\mathbf{v}]_{\tan} + {}^{\Sigma}\alpha_2\hat{\nabla}\hat{\theta} + {}^{\Sigma}\alpha_3(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan}\,, \quad (5.40)$$
$$[\boldsymbol{\sigma}_{(\hat{n})}]_{\tan} = {}^{\Sigma}\beta_1(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan} + {}^{\Sigma}\beta_2\hat{\nabla}\hat{\theta} + {}^{\Sigma}\beta_3[\mathbf{v}]_{\tan}\,,$$

and

$${}^{d}\hat{\sigma} = {}^{\Sigma}k_1\hat{\nabla}\cdot\hat{\mathbf{v}} + {}^{\Sigma}k_2[1/\theta] + {}^{\Sigma}k_3(\langle 1/\theta \rangle - 1/\hat{\theta})\,,$$
$$\langle q_{(\hat{n})}\rangle = {}^{\Sigma}l_1[1/\theta] + {}^{\Sigma}l_2\hat{\nabla}\cdot\hat{\mathbf{v}} + {}^{\Sigma}l_3(\langle 1/\theta \rangle - 1/\hat{\theta})\,, \quad (5.41)$$
$$[q_{(\hat{n})}] = {}^{\Sigma}m_1(\langle 1/\theta \rangle - 1/\hat{\theta}) + {}^{\Sigma}m_2\hat{\nabla}\cdot\hat{\mathbf{v}} + {}^{\Sigma}m_3[1/\theta]\,.$$

On account of (5.39)–(5.41), Eq. (5.37) yields

$$2\hat{\mu}_0\hat{d}_{ij\,0}\hat{d}_{ij} + \frac{{}^{\Sigma}K_1}{\hat{\theta}}(\hat{\nabla}\hat{\theta})^2 + {}^{\Sigma}\alpha_1([\mathbf{v}]_{\tan})^2 + {}^{\Sigma}\beta_1(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})^2_{\tan} + \left(\frac{{}^{\Sigma}K_2}{\hat{\theta}} + {}^{\Sigma}\alpha_2\right)([\mathbf{v}]_{\tan}\cdot\hat{\nabla}\hat{\theta})$$
$$+ \left(\frac{{}^{\Sigma}K_3}{\hat{\theta}} + {}^{\Sigma}\beta_2\right)[(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan}\cdot\hat{\nabla}\hat{\theta}] + ({}^{\Sigma}\alpha_3 + {}^{\Sigma}\beta_3)\,[[\mathbf{v}]_{\tan}\cdot(\langle \mathbf{v}\rangle - \hat{\mathbf{v}})_{\tan}] + {}^{\Sigma}k_1(\hat{\nabla}\cdot\hat{\mathbf{v}})^2 + {}^{\Sigma}l_1([1/\theta])^2$$
$$+ {}^{\Sigma}m_1\left(\left\langle\frac{1}{\theta}\right\rangle - \frac{1}{\hat{\theta}}\right)^2 + ({}^{\Sigma}k_2 + {}^{\Sigma}l_2)(\hat{\nabla}\cdot\hat{\mathbf{v}})[1/\theta]$$
$$+ ({}^{\Sigma}k_3 + {}^{\Sigma}m_2)(\hat{\nabla}\cdot\hat{\mathbf{v}})(\langle 1/\theta \rangle - 1/\hat{\theta}) + ({}^{\Sigma}l_3 + {}^{\Sigma}m_3)[1/\theta](\langle 1/\theta \rangle - 1/\hat{\theta}) \geqslant 0\,. \quad (5.42)$$

This inequality must be invarient under time reversal.

Since the quantities $[\mathbf{v}]$, $\langle \mathbf{v}\rangle - \hat{\mathbf{v}}$, and $\hat{\nabla}\cdot\hat{\mathbf{v}}$ are odd under a change of time direction and $\hat{\nabla}\hat{\theta}$, $[1/\theta]$, and $\langle 1/\theta \rangle - 1/\hat{\theta}$ are even, we can easily notice that we must have the following restrictions:

$${}^{\Sigma}\alpha_2 = -\,{}^{\Sigma}K_2/\hat{\theta}, \quad {}^{\Sigma}\beta_2 = -\,{}^{\Sigma}K_3/\hat{\theta},$$
$${}^{\Sigma}k_2 = -\,{}^{\Sigma}l_2, \quad {}^{\Sigma}k_3 = -\,{}^{\Sigma}m_2\,. \quad (5.43)$$

Other restrictions that we do not develop here are imposed through Eq. (5.42), more particularly on the sign of the scalar coefficients. At this point it is interesting to notice that Eqs. (5.39)–(5.41) subject to (5.43) are exactly the same as the one derived in Ref. 7.

Notice that if we neglect surface tension and surface forces ($\hat{\sigma}_{ij}=0, \hat{f}_i=0$) in the present case the surface equation of motion reduces to the usual jump relation

$$[\sigma_{ij}]\hat{n}_j=0 \tag{5.44}$$

and the third equation of (5.40) may be written as

$$\hat{\mathbf{v}}=\langle\mathbf{v}\rangle+(1/{}^{\Sigma}\beta_1)\{{}^{\Sigma}\beta_2\hat{\nabla}\hat{\theta}+{}^{\Sigma}\beta_3[\mathbf{v}]\}. \tag{5.45}$$

Since $\hat{v}_{(\hat{n})}=v^{\pm}_{(\hat{n})}$ and the terms within the curly brackets are linear in the thermodynamic forces and therefore small close to equilibrium, at a zeroth-order approximation one may therefore use

$$\hat{\mathbf{v}}=\langle\mathbf{v}\rangle. \tag{5.46}$$

This justifies the choice (5.46) in some particular situations. By the way, if we consider the electromagnetic case with the above-mentioned simplifications that lead to (5.44) we obtain from the equation of motion at Σ,

$$[t_{ij}+t^{\mathrm{em}}_{ij}+G_i\hat{v}_j]\hat{n}_j=0, \tag{5.47}$$

so that

$$[{}^{T}\tau_{(\hat{n})}]\cdot(\langle\mathbf{v}\rangle-\hat{\mathbf{v}})=-\hat{q}_f\langle\mathscr{E}\rangle\cdot(\langle\mathbf{v}\rangle-\hat{\mathbf{v}}) \tag{5.48}$$

and the Clausius–Duhem inequality (5.31) transforms to

$$-\left(\frac{\hat{d}\,{}^{\Sigma}\Psi}{dt}+{}^{\Sigma}\eta\frac{\hat{d}\hat{\theta}}{dt}\right)+(\hat{\sigma}_{ij}-{}^{\Sigma}\Psi P_{ij})\hat{D}_{ij}-(\hat{\mathbf{q}}/\hat{\theta})\cdot\hat{\nabla}\hat{\theta}$$
$$+\hat{\theta}\langle q_{(\hat{n})}\rangle[1/\theta]+\hat{\theta}[q_{(\hat{n})}](\langle 1/\theta\rangle-1/\hat{\theta})$$
$$+\langle{}^{T}\tau_{(\hat{n})}\rangle\cdot[\mathbf{v}]+\langle\mathscr{E}\rangle_{\tan}\cdot{}^{\Sigma}\mathscr{J}\geqslant 0, \tag{5.49}$$

where

$${}^{\Sigma}\mathscr{J}=\hat{\mathscr{J}}-\hat{q}_f(\langle\mathbf{v}\rangle-\hat{\mathbf{v}})=\hat{\mathbf{J}}-\hat{q}_f\langle\mathbf{v}\rangle. \tag{5.50}$$

We easily notice that Eq. (5.46) is included naturally in the Clausius–Duhem inequality (5.49), through the definition (5.50). This proves the validity of the approximation (5.46), for electromagnetic phenomena, this time, and directly from the Clausius–Duhem inequality and not borrowed from thermodynamical arguments through linear laws.

APPENDIX A: EFFECTIVE LORENTZ FORCE, PONDEROMOTIVE FORCE, AND ENERGY OF ELECTROMAGNETIC FORCES IN THE PRESENCE OF A SINGULAR SURFACE

1. Before evaluating the electromagnetic forces and their energy it is necessary to introduce the following identities, definitions, and mathematical properties

(i) On account of (3.5), (3.10), (3.18), (3.29), and (3.30) and after some calculations we obtain the following identities:

$$q_f\mathbf{E}+(1/c)\,\mathbf{J}\times\mathbf{B}\equiv q_f\mathscr{E}+(1/c)\mathscr{J}\times\mathbf{B}, \tag{A1}$$

$$\hat{q}_f\langle\mathbf{E}\rangle+(1/c)\,\hat{\mathbf{J}}\times\langle\mathbf{B}\rangle\equiv\hat{q}_f\hat{\mathscr{E}}+(1/c)\hat{\mathscr{J}}\times\langle\mathbf{B}\rangle, \tag{A2}$$

$$q^{\mathrm{eff}}\mathbf{E}+(1/c)\mathbf{J}^{\mathrm{eff}}\times\mathbf{B}\equiv q^{\mathrm{eff}}\mathscr{E}+(1/c)\,\mathscr{J}^{\mathrm{eff}}\times\mathbf{B}, \tag{A3}$$

$$\hat{q}^{\mathrm{eff}}\langle\mathbf{E}\rangle+(1/c)\hat{\mathbf{J}}^{\mathrm{eff}}\times\langle\mathbf{B}\rangle\equiv\hat{q}^{\mathrm{eff}}\hat{\mathscr{E}}+(1/c)\,\hat{\mathscr{J}}^{\mathrm{eff}}\times\langle\mathbf{B}\rangle, \tag{A4}$$

and

$$[q_f\mathbf{E}+(1/c)\,\mathbf{J}\times\mathbf{B}]\cdot\mathbf{v}$$
$$=[q_f\mathscr{E}+(1/c)\,\mathscr{J}\times\mathbf{B}]\cdot\mathbf{v}\equiv\mathbf{J}\cdot\mathbf{E}-\mathscr{J}\cdot\mathscr{E}, \tag{A5}$$

$$[\hat{q}_f\langle\mathbf{E}\rangle+(1/c)\,\hat{\mathbf{J}}\times\langle\mathbf{B}\rangle]\cdot\mathbf{v}$$
$$=[\hat{q}_f\hat{\mathscr{E}}+(1/c)\,\hat{\mathscr{J}}\times\langle\mathbf{B}\rangle]\cdot\mathbf{v}\equiv\hat{\mathbf{J}}\cdot\langle\mathbf{E}\rangle-\hat{\mathscr{J}}\cdot\hat{\mathscr{E}}, \tag{A6}$$

where $\hat{\mathscr{E}}$ has been defined by

$$\hat{\mathscr{E}}=\langle\mathbf{E}\rangle+(1/c)\hat{\mathbf{v}}\times\langle\mathbf{B}\rangle. \tag{A7}$$

It is useful to notice the following correspondence between quantities in the bulk and at the interface:

$$q_f\to\hat{q}_f,\quad \mathbf{J}\to\hat{\mathbf{J}},\quad \mathscr{J}\to\hat{\mathscr{J}},\quad q^{\mathrm{eff}}\to\hat{q}^{\mathrm{eff}},\quad \mathbf{J}^{\mathrm{eff}}\to\hat{\mathbf{J}}^{\mathrm{eff}},$$
$$\mathscr{J}^{\mathrm{eff}}\to\hat{\mathscr{J}}^{\mathrm{eff}},\quad \mathbf{v}\to\hat{\mathbf{v}},\quad \mathbf{B}\to\langle\mathbf{B}\rangle,\quad \mathbf{E}\to\langle\mathbf{E}\rangle. \tag{A8}$$

(ii) Some mathematical properties are as follows:

$$\begin{aligned}&(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D})=(\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D})-(\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}),\\&(\mathbf{A}\times\mathbf{B})\times\mathbf{C}=(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{B}\cdot\mathbf{C})\mathbf{A},\\&\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\cdot\mathbf{C})\mathbf{B}-(\mathbf{A}\cdot\mathbf{B})\mathbf{C},\\&(\mathbf{A}\times\mathbf{B})\cdot\mathbf{C}=\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}),\end{aligned} \tag{A9}$$

and

$$\begin{aligned}&[a_ib_j]=[a_i]\langle b_j\rangle+\langle a_i\rangle[b_j],\\&\langle a_ib_j\rangle=\langle a_i\rangle\langle b_j\rangle+\tfrac{1}{4}[a_i][b_j],\\&[a_ib_jc_k]=[a_i]\langle b_j\rangle\langle c_k\rangle+\langle a_i\rangle[b_j]\langle c_k\rangle\\&\qquad+\langle a_i\rangle\langle b_j\rangle[c_k]+\tfrac{1}{4}[a_i][b_j][c_k],\\&\langle a_ib_jc_k\rangle=+\tfrac{1}{4}\{\langle a_i\rangle[b_j][c_k]+[a_i]\langle b_j\rangle[c_k]\\&\qquad+[a_i][b_j]\langle c_k\rangle\}+\langle a_i\rangle\langle b_j\rangle\langle c_k\rangle.\end{aligned} \tag{A10}$$

2. Effective Lorentz force

(i) The expression for the effective Lorentz force in the bulk (reminder) [see Eq. (3.16), Ref. 5] is

$${}_{L}\,\mathbf{f}\equiv\operatorname{div}\mathbf{t}^{F}-\frac{\partial\mathbf{G}}{\partial t}=q^{\mathrm{eff}}\mathscr{E}+\frac{1}{c}\mathscr{J}^{\mathrm{eff}}\times\mathbf{B}, \tag{A11}$$

with

$$t^{F}_{ij}=E_iE_j+B_iB_j-\tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)\delta_{ij},\quad \mathbf{G}=(1/c)\mathbf{E}\times\mathbf{B}. \tag{A12}$$

(ii) We evaluate the effective Lorentz force at the interface. On account of Maxwell's equations at the interface of Eq. (A7) and the first equation of (A10) and after some long and rather tricky calculations, we obtain

$${}_{L}\hat{f}\equiv[\mathbf{t}^{F}+\mathbf{G}\otimes\hat{\mathbf{v}}]\cdot\hat{\mathbf{n}}=\hat{q}^{\mathrm{eff}}\hat{\mathscr{E}}+(1/c)\,\hat{\mathscr{J}}^{\mathrm{eff}}\times\langle\mathbf{B}\rangle. \tag{A13}$$

3. Expression of the ponderomotive force in the bulk and at the interface

On account of (A11) and (A13) and of the identities

$$\mathbf{f}^{\mathrm{em}}={}_{L}\,\mathbf{f}+\operatorname{div}\bar{\mathbf{t}}^{\mathrm{em}},\quad \hat{\mathbf{f}}^{\mathrm{em}}\equiv{}_{L}\,\hat{\mathbf{f}}+[\bar{\mathbf{t}}^{\mathrm{em}}]\cdot\hat{\mathbf{n}}, \tag{A14}$$

where

$$\bar{t}^{\mathrm{em}}_{ij}=\mathscr{E}_iP_j-\mathscr{M}_iB_j+\mathscr{M}\cdot\mathbf{B}\delta_{ij}, \tag{A15}$$

we are led to

$$\mathbf{f}^{\mathrm{em}}=q^{\mathrm{eff}}\mathscr{E}+(1/c)(\mathscr{J}^{\mathrm{eff}}-c\nabla\times\mathscr{M})\times\mathbf{B}$$
$$+\operatorname{div}(\mathscr{E}\otimes\mathbf{P})+(\nabla\mathbf{B})\cdot\mathscr{M}, \tag{A16}$$

$$\hat{\mathbf{f}}^{\mathrm{em}} = \hat{q}^{\mathrm{eff}}\hat{\mathscr{E}} + (1/c)(\hat{\mathscr{J}}^{\mathrm{eff}} - c\hat{\mathbf{n}}\times[\mathscr{M}])\times\langle\mathbf{B}\rangle + [\mathscr{E}\otimes\hat{\mathbf{P}}]\cdot\hat{\mathbf{n}} + (\langle\mathscr{M}\rangle\cdot[\mathbf{B}])\hat{\mathbf{n}}, \quad (A17)$$

or equivalently

$$\mathbf{f}^{\mathrm{em}} = q^{\mathrm{eff}}\mathbf{E} + (1/c)(\mathbf{J}^{\mathrm{eff}} - c\nabla\times\mathbf{M})\times\mathbf{B} + \mathrm{div}(\mathbf{E}\otimes\mathbf{P}) + (\nabla\mathbf{B})\cdot\mathbf{M} + {}^{R}\mathbf{f}^{\mathrm{em}}, \quad (A18)$$

$$\hat{\mathbf{f}}^{\mathrm{em}} = \hat{q}^{\mathrm{eff}}\langle\mathbf{E}\rangle + (1/c)(\hat{\mathbf{J}}^{\mathrm{eff}} - c\hat{n}\times[\mathbf{M}])\times\langle\mathbf{B}\rangle + [\mathbf{E}\otimes\mathbf{P}]\cdot\hat{\mathbf{n}} + (\langle\mathbf{M}\rangle\cdot[\mathbf{B}])\hat{\mathbf{n}} + {}^{R}\hat{\mathbf{f}}^{\mathrm{em}}, \quad (A19)$$

with

$${}^{R}\mathbf{f}^{\mathrm{em}} = (1/c)\mathrm{div}\{(\mathbf{v}\times\mathbf{B})\otimes\mathbf{P} - (\mathbf{v}\times\mathbf{P})\otimes\mathbf{B} + [(\mathbf{v}\times\mathbf{P})\cdot\mathbf{B}]I\}, \quad (A20)$$

$${}^{R}\hat{\mathbf{f}}^{\mathrm{em}} = (1/c)[(\mathbf{v}\times\mathbf{B})\otimes\mathbf{P} - (\mathbf{v}\times\mathbf{P})\otimes\mathbf{B} + [(\mathbf{v}\times\mathbf{P})\cdot\mathbf{B}]I]\cdot\hat{\mathbf{n}}. \quad (A21)$$

Remarks: (a) Note that Eqs. (A16) and (A17), where the electric field, the effective current, and magnetization are expressed in the comoving frame $R_c(x,t)$, are more convenient to deal with than the equivalent Eqs. (A18) and (A19) where the above-mentioned fields are expressed in the Galilean frame R_G. This justifies the use of (A16) and (A17) in the present paper instead of (A18) and (A19).

(b) It is easy to notice that Eqs. (A11), (A13), and (A16)–(A21) do not violate the correspondences given in (A8), so we have an absolute symmetry between the bulk and the interface.

(c) If we decompose the *effective* quantities and combine them with $\bar{t}_{ij}^{\mathrm{em}}$ in Eqs. (A14) we obtain

$$\mathbf{f}^{\mathrm{em}} = q_f\mathscr{E} + (1/c)(\mathscr{J} + \overset{*}{\mathbf{P}})\times\mathbf{B} + (\mathbf{P}\cdot\nabla)\mathscr{E} + (\nabla\mathbf{B})\cdot\mathscr{M}, \quad (A22)$$

$$\hat{\mathbf{f}}^{\mathrm{em}} = \hat{q}_f\hat{\mathscr{E}} + (1/c)(\hat{\mathscr{J}} + \hat{\mathbf{n}}\times[\mathbf{P}\times(\mathbf{v}-\hat{\mathbf{v}})])\times\langle\mathbf{B}\rangle + [\mathscr{E}](\langle\mathbf{P}\rangle\cdot\hat{\mathbf{n}}) + (\langle\mathscr{M}\rangle\cdot[\mathbf{B}])\hat{\mathbf{n}} + \boldsymbol{\delta}([\mathbf{P}]\cdot\hat{\mathbf{n}}), \quad (A23)$$

with

$$\boldsymbol{\delta} = \langle\mathscr{E}\rangle - \hat{\mathscr{E}} = (1/c)(\langle\mathbf{v}-\hat{\mathbf{v}}\rangle\times\mathbf{B}\rangle = (1/c)\{(\langle\mathbf{v}\rangle - \hat{\mathbf{v}})\times\langle\mathbf{B}\rangle + \tfrac{1}{4}[\mathbf{v}]\times[\mathbf{B}]\}. \quad (A24)$$

Equations (A22) and (A23) show the existence of additional terms at the interface, which vanish only if $\mathbf{v} = \hat{\mathbf{v}}$. Hence, in general, the absolute symmetry is lost, and this will complicate the evaluation of the electromagnetic energy identity, as will be noticed later on.

4. Energy of electromagnetic forces

We consider the case where the electric quadrupoles are neglected as well as surface polarization and magnetization. The calculation follows the same line as the much simpler one performed in the previous section or Refs. 5 and 6. This cannot be reproduced because of lack of space.

On account of (A9), (A10), (A14), and Maxwell's equations, it is possible to evaluate the right-hand side of the following equation [obtained in Ref. 5 from Eqs. (3.58) with a notation similar to that used in (4.2)]

$$({}_{L}\mathscr{f}\cdot\mathbf{v} + \mathscr{J}^{\mathrm{eff}}\cdot\mathscr{E})\hat{} \equiv [(t_{ij}^{F} + G_i\hat{v}_j)v_i - c^2\mathscr{G}_j - \tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)(v_j - \hat{v}_j)]\hat{n}_j. \quad (A25)$$

After long and rather tricky calculations we obtain

$$({}_{L}\mathscr{f}\cdot\mathbf{v} + \mathscr{J}^{\mathrm{eff}}\cdot\mathscr{E})\hat{} = {}_{L}\hat{\mathscr{f}}\hat{\mathbf{v}} + \hat{\mathscr{J}}^{\mathrm{eff}}\cdot\langle\mathscr{E}\rangle + \hat{r}, \quad (A26)$$

with

$$\hat{r} = [G_i\hat{v}_j(v_i - \hat{v}_i) - 2\{(\mathbf{E} - \langle\mathscr{E}\rangle)_i E_{[i}(\mathbf{v}-\hat{\mathbf{v}})_{j]} + (\mathbf{B} - \langle\mathscr{B}\rangle)_i B_{[i}(\mathbf{v}-\hat{\mathbf{v}})_{j]}\}]\hat{n}_j, \quad (A27a)$$

or equivalently

$$\hat{r} = [((\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})G_j - 2\{(\mathbf{E} - \langle\mathscr{E}\rangle)_{[i}\hat{n}_{j]}E_i + (\mathbf{B} - \langle\mathscr{B}\rangle)_{[i}\hat{n}_{j]}B_i\})(v_j - \hat{v}_j)], \quad (A27b)$$

where we have used the following property:

$$[A_i\langle B_j\rangle] = [A_i]\langle B_j\rangle \Leftrightarrow [\langle\mathbf{B}\rangle] = 0, \quad \langle\langle\mathbf{B}\rangle\rangle \equiv \langle\mathbf{B}\rangle. \quad (A27c)$$

Noting that

$$(E_i - \mathscr{E}_i)E_i = (B_i - \mathscr{B}_i)B_i = \mathbf{G}\cdot\mathbf{v},$$
$$(E_i - \mathscr{E}_i)v_i = (B_i - \mathscr{B}_i)v_i = 0,$$

and

$$\{v_i - \hat{v}_i\}\{(E_i - \mathscr{E}_i)E_j + (B_i - \mathscr{B}_i)B_j\} = (1/c)\{-(\mathbf{v}\times\mathbf{B})_i E_j + (\mathbf{v}\times\mathbf{E})_i B_j\}\{v_i - \hat{v}_i\} = -(2/c)(\hat{\mathbf{v}}\times\mathbf{v})_i E_{[i}B_{j]}, \quad (A27d)$$

we are led to

$$\hat{r} = 2\{[(\tfrac{1}{2}G_i\hat{v}_j - \mathbf{G}\cdot\mathbf{v}\delta_{ij})(v_i - \hat{v}_i) - (1/c)(\hat{\mathbf{v}}\times\mathbf{v})E_{[i}B_{j]}]\hat{n}_j + \langle E_i(v_j - \hat{v}_j)\rangle \times[\mathscr{E}_{[i}\hat{n}_{j]}] + \langle B_i(v_j - \hat{v}_j)\rangle[\mathscr{B}_{[i}\hat{n}_{j]}]\}, \quad (A27e)$$

or equivalently,

$$\hat{r} = K_i(\langle v_i\rangle - \hat{v}_i) + L_i[v_i], \quad (A27f)$$

with

$$K_i = [K_{ij}]\hat{n}_j + \langle E_j\rangle[\mathscr{E}_{[j}\hat{n}_{i]}] + \langle B_j\rangle[\mathscr{B}_{[j}\hat{n}_{i]}],$$
$$L_i = \langle K_{ij}\rangle\hat{n}_j + \tfrac{1}{4}\{[E_j][\mathscr{E}_{[j}\hat{n}_{i]}] + [B_j][\mathscr{B}_{[j}\hat{n}_{i]}]\},$$
$$K_{ij} = G_i\hat{v}_j - 2\mathbf{G}\cdot\mathbf{v}\delta_{ij} + (1/c)\{(\mathbf{v}\times\mathbf{E})_i B_j - (\mathbf{v}\times\mathbf{B})_i E_j\}. \quad (A27g)$$

We easily notice that $\hat{r}$ vanishes for $\mathbf{v} = \hat{\mathbf{v}}$.

Now if we combine (A25) and (A26) with the Eq. (4.8) of Ref. 5,

$$\hat{\omega}^{\mathrm{em}} \equiv [(t_{ij}^{\mathrm{em}} + G_i\hat{v}_j)v_i - \mathscr{S}_j - \tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)(v_j - \hat{v}_j)]\hat{n}_j, \quad (A28)$$

we get

$$\hat{\omega}^{\mathrm{em}} = {}_{L}\hat{\mathbf{f}}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}^{\mathrm{eff}}\cdot\langle\mathscr{E}\rangle + [\bar{t}_{ij}^{\mathrm{em}}v_i + c(\mathscr{E}\times\mathscr{M})_j]\hat{n}_j + \hat{r}. \quad (A29)$$

On account of (A14) and (3.25), Eq. (A29) transforms to

$$\hat{\omega}^{\mathrm{em}} = \hat{\mathbf{f}}^{\mathrm{em}}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + \langle\mathscr{E}\rangle\{\hat{\mathbf{n}}\times[\mathbf{P}\times(\mathbf{v}-\hat{\mathbf{v}})]\} + c\langle\mathscr{M}\rangle\cdot(\hat{\mathbf{n}}\times[\mathscr{E}]) + \langle\bar{t}_{ij}^{\mathrm{em}}\rangle[v_i]\hat{n}_j + \hat{r} + [\bar{t}_{ij}^{\mathrm{em}}](\langle v_i\rangle - \hat{v}_i)\hat{n}_j. \quad (A30)$$

Notice the analogy with the following volume equations [(3.42) and (3.41) of Ref. 5]:

$$\omega^{\rm em} = {}_L\, \mathbf{f}\cdot\mathbf{v} + \mathscr{J}^{\rm eff}\cdot\mathscr{E} + (\bar{t}^{\rm em}_{ij} v_i + c(\mathscr{E}\times\mathscr{M})_j)_{,j}\,, \tag{A31}$$

and

$$\omega^{\rm em} = \mathbf{f}^{\rm em}\cdot\mathbf{v} + \mathscr{J}\cdot\mathscr{E} + \mathscr{E}\cdot\overset{*}{\mathbf{P}} + c\mathscr{M}\cdot(\nabla\times\mathscr{E}) + \bar{t}^{\rm em}_{ij} v_{i,j}\,. \tag{A32}$$

Expressing $\nabla\times\mathscr{E}$ and $\mathbf{n}\times[\mathscr{E}]$ with the help of (3.12) and (3.23), decomposing $v_{i,j}$ and $[v_i]\hat{n}_j$ into symmetric and skew symmetric parts, introducing the vorticity vector $\boldsymbol{\omega} = \frac{1}{2}\nabla\times\mathbf{v}$ and its surface counterpart $\hat{\boldsymbol{\omega}} = \frac{1}{2}\hat{\mathbf{n}}\times[\mathbf{v}]$, and taking account of (3.22) of Ref. 4, we obtain

$$\omega^{\rm em} = \mathbf{f}^{\rm em}\cdot\mathbf{v} + \mathbf{c}^{\rm em}\cdot\boldsymbol{\omega} + \rho h^{\rm em} \tag{A33}$$

and

$$\hat{\omega}^{\rm em} = \hat{\mathbf{f}}^{\rm em}\cdot\hat{\mathbf{v}} + \langle c^{\rm em}\rangle\cdot\hat{\boldsymbol{\omega}} + (\rho h^{\rm em})\hat{} + \hat{r} + [\bar{t}^{\rm em}_{ij}](\langle v_i\rangle - \hat{v}_i)\hat{n}_j\,, \tag{A34}$$

where

$$\rho h^{\rm em} = \mathscr{J}\cdot\mathscr{E} + \mathscr{E}\cdot\overset{*}{\mathbf{P}} - \mathscr{M}\cdot\overset{*}{\mathbf{B}} + v_{(i,j)}\bar{t}^{\rm em}_{ij}\,, \tag{A35}$$

$$(\rho h^{\rm em})\hat{} = \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + \langle\mathscr{E}\rangle\cdot\{\hat{\mathbf{n}}\times[\mathbf{P}\times(\mathbf{v}-\hat{\mathbf{v}})]\} - \langle\mathscr{M}\rangle\cdot\{\hat{\mathbf{n}}\times[\mathbf{B}\times(\mathbf{v}-\hat{\mathbf{v}})]\} + [v_{(i}]\hat{n}_{j)}\langle\bar{t}^{\rm em}_{ij}\rangle\,. \tag{A36}$$

Thus $\omega^{\rm em}$ and $\hat{\omega}^{\rm em}$ are made of the powers developed by the ponderomotive force and couple, of a volume and surface contributions $\rho h^{\rm em}$ and $(\rho h^{\rm em})\hat{}$, which result solely from the fact that the material is assumed to be electrically polarized, magnetized, and is a conductor, and of additional surface terms which vanish for $\mathbf{v} = \hat{\mathbf{v}}$.

Another useful form may be obtained on decomposing $\overset{*}{\mathbf{P}}, \overset{*}{\mathbf{B}}$, and introducing $\mathbf{P} = \rho\boldsymbol{\pi}$. After many manipulations we are led to

$$\omega^{\rm em} = \mathbf{f}^{\rm em}\cdot\mathbf{v} + \mathscr{J}\cdot\mathscr{E} + \rho\mathscr{E}\cdot\frac{d\boldsymbol{\pi}}{dt} - \mathscr{M}\cdot\frac{d\mathbf{B}}{dt}, \tag{A37}$$

$$\begin{aligned}\hat{\omega}^{\rm em} = {}& \hat{\mathbf{f}}^{\rm em}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + \langle\mathscr{E}\rangle\cdot[\mathbf{P}(v_j - \hat{v}_j)]\hat{n}_j \\ & + [\mathscr{E}_i]\langle P_j(v_i - \hat{v}_i)\rangle\hat{n}_j - [\mathscr{M}_i]\langle B_j(v_i-\hat{v}_i)\rangle\hat{n}_j \\ & + [\mathscr{M}]\cdot\langle\mathbf{B}(v_j - \hat{v}_j)\rangle\hat{n}_j + \hat{r}\,,\end{aligned} \tag{A38}$$

or equivalently

$$\hat{\omega}^{\rm em} = \hat{\mathbf{f}}^{\rm em}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + X_i(\langle v_i\rangle - \hat{v}_i) + Y_i[v_i]\,, \tag{A39}$$

with

$$\begin{aligned} & X_i = K_i + X_{ij}\hat{n}_j, \quad Y_i = L_i + Y_{ij}\hat{n}_j\,, \\ & X_{ij} = (\langle\mathscr{E}\rangle\cdot[\mathbf{P}] + [\mathscr{M}]\cdot\langle\mathbf{B}\rangle)\delta_{ij} + [\mathscr{E}_i]\langle P_j\rangle - [\mathscr{M}_i]\langle B_j\rangle\,, \\ & Y_{ij} = (\langle\mathscr{E}\rangle\cdot\langle\mathbf{P}\rangle + \tfrac{1}{4}[\mathscr{M}]\cdot[\mathbf{B}])\delta_{ij} + \tfrac{1}{4}[\mathscr{E}_i][P_j]\,.\end{aligned} \tag{A40}$$

An alternate expression of $\hat{\omega}^{\rm em}$ may be obtained, directly from Eq. (A28), which, on account of Maxwell's equations and (A13) and (A14), leads to

$$\hat{\omega}^{\rm em} = \hat{\mathbf{f}}^{\rm em}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle - {}^{\Sigma}\mathscr{S}^d\cdot\hat{\mathbf{n}} + [(t^{\rm em}_{ij} + G_i\hat{v}_j)(v_i - \hat{v}_i) - \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2)(v_j - \hat{v}_j)]\hat{n}_j\,, \tag{A41}$$

with

$${}^{\Sigma}\mathscr{S}^d\cdot\hat{\mathbf{n}} = [\mathscr{S}]\cdot\hat{\mathbf{n}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle \tag{A42}$$

and

$$\begin{aligned}{}^{\Sigma}\mathscr{S}^d_j & = [\langle\mathscr{E}\rangle\times[\mathbf{D}\times(\mathbf{v}-\hat{\mathbf{v}})]]_j + [\langle\mathscr{H}\rangle\times[\mathbf{B}\times(\mathbf{v}-\hat{\mathbf{v}})]]_j \\ & = [{}^{\Sigma}\mathscr{S}^d_{ij}(v_i - \hat{v}_i)]\,,\end{aligned} \tag{A43a}$$

with

$${}^{\Sigma}\mathscr{S}^d_{ij} = \langle\mathscr{E}_i\rangle D_j = \langle\mathscr{H}_i\rangle B_j - (\langle\mathscr{E}\rangle\cdot\mathbf{D} + \langle\mathscr{H}\rangle\cdot\mathbf{B})\delta_{ij}\,. \tag{A43b}$$

The superscript d stands for dielectric, since, in the absence of conduction ($\hat{\mathscr{J}} = \mathbf{0}$) ${}^{\Sigma}\mathscr{S}^d$ reduces to the jump of Poynting's vectors. It is easy to notice that for a material surface ($\mathbf{v} = \hat{\mathbf{v}} = \mathbf{v}$), (i) ${}^{\Sigma}\mathscr{S}^d$ vanishes and $-[\mathscr{S}]\cdot\hat{\mathbf{n}}$ is nothing but the Joule effect at the interface and (ii) the various expressions of $\hat{\omega}^{\rm em}$ reduce to

$$\hat{\omega}^{\rm em} = \hat{\mathbf{f}}^{\rm em}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle\,. \tag{A44}$$

The decomposition of (A41)–(A44) allows one to write

$$\begin{aligned}\hat{\omega}^{\rm em} = {}& \hat{\mathbf{f}}^{\rm em}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + ([t^{\rm em}_{ij} + G_i\hat{v}_j - \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2)\delta_{ij}] \\ & \times\hat{n}_j - {}^{M}\mathscr{S}_i)(\langle v_i\rangle - \hat{v}_i) \\ & + (\langle t^{\rm em}_{ij} + G_j\hat{v}_j - \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2)\delta_{ij}\rangle\hat{n}_j - {}^{J}\mathscr{S}_i)[v_i]\,,\end{aligned} \tag{A45}$$

where we have set

$${}^{M}\mathscr{S}_i = +\hat{q}_f\langle\mathscr{E}_i\rangle - (\langle\mathscr{E}\rangle\cdot[\mathbf{D}] + \langle\mathscr{H}\rangle\cdot[\mathbf{B}])\hat{n}_i\,, \tag{A46}$$

$${}^{J}\mathscr{S}_i = +\{(\hat{\mathbf{n}}\cdot\langle\mathbf{D}\rangle)\langle\mathscr{E}_i\rangle - (\hat{\mathbf{n}}\cdot\langle\mathbf{B}\rangle)\langle\mathscr{H}_i\rangle\} - (\langle\mathscr{E}\rangle\cdot\langle\mathbf{D}\rangle + \langle\mathscr{H}\rangle\cdot\langle\mathbf{B}\rangle)\hat{n}_i\,. \tag{A47}$$

Equations (A26) with (A27f), (A39), and (A45) give directly the expressions required for the treatment of dissipative constitutive equations and they allow one to consider easily particular cases such as the absence of mass transfer, in which case we may assume $\hat{\mathbf{v}} = \langle\mathbf{v}\rangle$ in a first approximation (see Ref. 7) while $[\mathbf{v}] \neq \mathbf{0}$.

Before ending this appendix let us notice that the combination of (A28) with (3.54) of Ref. 5 leads to the electromagnetic energy written in the Galilean frame as follows:

$$\hat{\omega}^{\rm em} = (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})[\tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2)] - [\mathbf{S} - \mathbf{v}(\mathbf{E}\cdot\mathbf{P})]\cdot\hat{\mathbf{n}}\,. \tag{A48}$$

After many calculations, using Maxwell's equations in R_G we are led to

$$\hat{\omega}^{\rm em} = \hat{\mathbf{J}}\cdot\langle\mathbf{E}\rangle - (\langle\mathbf{E}\rangle\cdot[\mathbf{P}] - \langle\mathbf{M}\rangle\cdot[\mathbf{B}])(\hat{\mathbf{v}}\cdot\hat{\mathbf{n}}) + [(\mathbf{E}\cdot\mathbf{P})v_i]\hat{n}_i\,. \tag{A49}$$

Equations (A48) and (A49) are the surface counter parts of the following volume equations obtained in Ref. 5:

$$\omega^{\rm em} = -\frac{\partial}{\partial t}\frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2) - \nabla\cdot[\mathbf{S} - \mathbf{v}(\mathbf{E}\cdot\mathbf{P})] \tag{A50}$$

and

$$\omega^{em} = \mathbf{J}\cdot\mathbf{E} + \mathbf{E}\cdot\frac{\partial \mathbf{P}}{\partial t} - \mathbf{M}\cdot\frac{\partial \mathbf{B}}{\partial t} + \{(\mathbf{E}\cdot\mathbf{P})v_i\}_{,i}\,, \quad \text{(A51)}$$

Equation (A49) may be of interest to deal with in the case of a fixed surface $\hat{\mathbf{v}} = 0$; hence we obtain

$$\hat{\omega}^{em} = \hat{\mathbf{J}}\cdot\langle\mathbf{E}\rangle + [\![\mathbf{E}\cdot\mathbf{P}]\!]\langle\mathbf{v}\rangle\cdot\hat{\mathbf{n}} + \langle\mathbf{E}\cdot\mathbf{P}\rangle[\mathbf{v}]\!]\cdot\hat{\mathbf{n}}\,. \quad \text{(A52)}$$

On using the identity (A6), we obtain an alternate form for $\hat{\omega}^{em}$, related to $\hat{\mathscr{J}}\cdot\hat{\mathscr{E}}$ instead of $\hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle$, and to $\hat{\mathbf{f}}^{em}\cdot\hat{\mathbf{v}}$ through $[\hat{q}_f\hat{\mathscr{E}} + (1/c)\hat{\mathscr{J}}\times\langle\mathbf{B}\rangle]\cdot\hat{\mathbf{v}}$ when we account for (A23).

APPENDIX B: ELECTROMAGNETIC ENERGY IDENTITY IN THE PRESENCE OF A SINGULAR SURFACE

The aim of this appendix is the evaluation of the electromagnetic energy identity when the medium is spanned by a discontinuity surface. We shall show the power developed by surface electromagnetic forces and account explicitly for the Joule effect in the same way as in the bulk. We recall that we consider neither surface polarization nor surface magnetization. To that purpose, let us write the following identity:

$$\begin{aligned}\frac{d}{dt}\int_{D-\Sigma}\frac{1}{2}(E^2+B^2)dv &= -\int_{D-\Sigma}\left(\mathbf{f}^{em}\cdot\mathbf{v} + \rho\mathscr{E}\cdot\frac{d\boldsymbol{\pi}}{dt} - \mathscr{M}\cdot\frac{d\mathbf{B}}{dt} + \mathscr{J}\cdot\mathscr{E}\right)dv\\ &\quad - \int_{\partial D-\Sigma}(\mathscr{S}_j - (t_{ij}^{em} + G_i v_j)v_i)n_j da\\ &\quad - \int_\Sigma [\![(t_{ij}^{em} + G_i\hat{v}_j)v_i - \tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)\\ &\quad \times(v_j - \hat{v}_j) - \mathscr{S}_j]\!]\hat{n}_j\, da\,, \end{aligned} \quad \text{(B1)}$$

which is obtained on combining (A37) and (4.5) of Ref. 5, and accounting for $\hat{\mathbf{v}}\cdot\hat{\mathbf{n}} = \mathbf{v}\cdot\hat{\mathbf{n}}$.

On account of (B1), the second equation of (4.18), (4.14), (4.22), (4.23), and (2.14) and the transport theorems we can write

$$\begin{aligned}\frac{d}{dt}U^{em} &= -\int_{D-\Sigma}\left(\mathbf{f}^{em}\cdot\mathbf{v} + \rho\mathscr{E}\cdot\frac{d\boldsymbol{\pi}}{dt} + \rho\mathbf{B}\cdot\frac{d\boldsymbol{\mu}}{dt} + \mathscr{J}\cdot\mathscr{E}\right)\\ &\quad\times dv - \int_{\partial D-\Sigma}(\mathbf{T}^{em}\cdot\mathbf{v} + \mathscr{S}\cdot\mathbf{n})da\\ &\quad - \int_\Sigma [\![(t_{ij}^{em} + G_i\hat{v}_j - \tfrac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2 - 2\mathscr{M}\cdot\mathbf{B})\delta_{ij})\\ &\quad\times(v_i - \hat{v}_i) - \mathscr{S}_j]\!]\hat{n}_j\, da\,. \end{aligned} \quad \text{(B2)}$$

At this point, only volume Maxwell's equations have been employed. In order to place in evidence the Joule effect at the interface, we use surface Maxwell's equations (3.21) and (3.22), thus obtaining the electromagnetic energy identity as follows [compare with (A41)–(A47)]:

$$\begin{aligned}\frac{d}{dt}U^{em} &= -\int_{D-\Sigma}\left(\mathbf{f}^{em}\cdot\mathbf{v} + \rho\mathscr{E}\cdot\frac{d\boldsymbol{\pi}}{dt} + \rho\mathbf{B}\cdot\frac{d\boldsymbol{\mu}}{dt} + \mathscr{J}\cdot\mathscr{E}\right)dv\\ &\quad - \int_{\partial D-\Sigma}(\mathbf{T}^{em}\cdot\mathbf{v} + \mathscr{S}\cdot\mathbf{n})da\\ &\quad - \int_\Sigma(\hat{\mathbf{f}}^{em}\cdot\hat{\mathbf{v}} + \hat{\mathscr{J}}\cdot\langle\mathscr{E}\rangle + [\![\tau_{ij}^{em}(v_i - \hat{v}_i)]\!]\hat{n}_j)da\,, \end{aligned} \quad \text{(B3)}$$

with

$$\tau_{ij}^{em} = t_{ij}^{em} + G_i\hat{v}_j - {}^{\Sigma}\mathscr{S}_{ij}^d - \left(\frac{\mathbf{E}^2 + \mathbf{B}^2}{2} - \mathscr{M}\cdot\mathbf{B}\right)\delta_{ij}\,, \quad \text{(B4)}$$

$${}^{\Sigma}\mathscr{S}_{ij}^d = \langle\mathscr{E}_i\rangle D_j + \langle\mathscr{H}_i\rangle B_j - (\langle\mathscr{E}\rangle\cdot\mathbf{D} + \langle\mathscr{H}\rangle\cdot\mathbf{B})\delta_{ij}\,. \quad \text{(B5)}$$

APPENDIX C: STATEMENT OF THE PRINCIPLE OF VIRTUAL POWER IN THE PRESENCE OF A SINGULAR SURFACE AND INCLUDING ELECTROMAGNETIC PHENOMENA

The principle of virtual power may be extended to account for electromagnetic phenomena, in the presence of a singular surface, by the use of the so-called first gradient theory discussed at length for electromagnetic continuous media in Ref. 3 and for purely mechanical media presenting a singular surface in Ref. 4. In the present approach we consider neither electromagnetic ordering such as in ferromagnetism or ferroelectricity nor surface polarization and magnetization.

On account of the systematic procedures presented in Refs. 3 and 4 and of the above-mentioned restrictions we get the following expressions

$${}^{\prime}P_i^* - \int_{D-\Sigma} p_i^*\, dv - \int_\Sigma({}^{\Sigma}p_i^* + \hat{p}_i^*)da\,, \quad \text{(C1)}$$

$$\begin{aligned}{}^{\prime}P_d^* &= \int_{D-\Sigma}\left[f_i v_i^* - t_{ij}^{em} v_{i,j}^* + \rho\mathscr{E}_i\left(\frac{d\pi_i}{dt}\right)^*\right.\\ &\quad\left. + \rho B_i\left(\frac{d\mu_i}{dt}\right)^*\right]dv + \int_\Sigma \hat{f}_i\hat{v}_i^*\, da\,, \end{aligned} \quad \text{(C2)}$$

$${}^{\prime}P_c^* = \int_{\partial D-\Sigma} T_i v_i^*\, da + \int_{\partial\Sigma}\hat{T}_i\hat{v}_i^*\, dl\,, \quad \text{(C3)}$$

and

$$\begin{aligned}{}^{\prime}P_a^* &= \int_{D-\Sigma}\left(\rho\frac{dv_i}{dt} + \frac{\partial G_i}{\partial t}\right)v_i^*\, dv\\ &\quad + \int_\Sigma\left(\hat{\rho}\frac{\hat{d}\hat{v}_i}{dt} + [\![m(v_i - \hat{v}_i) - (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})G_i]\!]\right)\hat{v}_i^*\, da\\ &\quad + \int_{\partial D-\Sigma} G_i v_j n_j v_i^*\, da\,, \end{aligned} \quad \text{(C4)}$$

where we have set

$$p_i^* = \sigma_{ij}D_{ij}^* - \rho^L E_i(D_J\,\pi)_i^* - \rho^L B_i(D_J\,\mu)_i^*\,, \quad \text{(C5)}$$

$${}^{\Sigma}p_i = [\![\mathscr{T}_i(v_i - \hat{v}_i)^*]\!], \quad \hat{p}_i = \hat{\sigma}_{ij}\hat{D}_{ij}^*\,, \quad \text{(C6)}$$

and which must satisfy the following statement:

$${}^{\prime}P_a^* = {}^{\prime}P_i^* + {}^{\prime}P_d^* + {}^{\prime}P_c^* \quad \text{(C7)}$$

for any virtual field and any element of volume, surface, and line. (The only new point concerning the construction of ${}^{\prime}P_a^*$ is that when Σ coincides with the boundary, $[\![m(v_i - \hat{v}_i) - (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}})G_i]\!]$ reduces to $G_i v_j n_j$. This term vanishes only if the surface is fixed ($\hat{\mathbf{v}} = \mathbf{v} = \mathbf{0}$).) After use of the generalized Stokes' theorem accounting for the presence of discontinuity surface, we are led to the following local equations:

$$\rho \frac{dv_i}{dt} + \frac{\partial G_i}{\partial t} = (t_{ij} + t_{ij}^{\mathrm{em}})_{,j} + f_i \quad \text{in } D\text{-}\Sigma\,, \tag{C8}$$

$$T_i = (t_{ij} + t_{ij}^{\mathrm{em}} + G_i v_j) n_j \quad \text{on } \partial D\text{-}\Sigma\,, \tag{C9}$$

$$\hat{\rho} \frac{\hat{d}\hat{v}_i}{dt} + [m(v_i - \hat{v}_i) - (\hat{\mathbf{v}}\cdot\hat{\mathbf{n}}) G_i]$$

$$= [\mathscr{T}_i] + (\hat{\nabla}_j + 2\Omega \hat{n}_j)\hat{\sigma}_{ij} + \hat{f}_i \quad \text{on } \Sigma\,, \tag{C10}$$

$$\hat{T}_i = \hat{\sigma}_{ij}\tau_j \quad \text{along } \partial\Sigma\,, \tag{C11}$$

$$\mathscr{T}_i^{\pm} = (t_{ij} + t_{ij}^{\mathrm{em}})^{\pm} \hat{n}_j \quad \text{on } \Sigma^{\pm}, \quad \hat{\sigma}_{ij}\hat{n}_j = 0 \quad \text{on } \Sigma\,, \tag{C12}$$

$$^{L}E_i + \mathscr{E}_i = 0, \quad {}^{L}B_i + B_i = 0 \quad \text{in } D\text{-}\Sigma\,. \tag{C13}$$

On combining (C10) with (C12) we obtain the more conventional form

$$\hat{\rho} \frac{\hat{d}\hat{v}_i}{dt} + [m(v_i - \hat{v}_i)] = [t_{ij} + t_{ij}^{\mathrm{em}} + G_i \hat{v}_j]\hat{n}_j + \hat{\nabla}_j \hat{\sigma}_{ij} + \hat{f}_i\,. \tag{C14}$$

On using the concept of ponderomotive forces $\mathbf{f}^{\mathrm{em}}$ and $\hat{\mathbf{f}}^{\mathrm{em}}$ instead of the electromagnetic tensor t_{ij}^{em} and momentum $\mathbf{G}$, through the volume and surface identities (4.21) and (4.22), we may state the principle of virtual power in a simpler but equivalent form. The latter is given by Eqs. (4.1) and (4.4)–(4.7).

It is easy to notice that the local equations (C8)–(C13) are exactly the same as (4.27)–(4.32) combined with (4.21) and (4.22).

APPENDIX D: GENERALIZED DIVERGENCE AND TRANSPORT THEOREMS

The transport theorems for volumes and surfaces are

$$\frac{d}{dt}\int_{D\text{-}\Sigma} \phi\, dv = \int_{D\text{-}\Sigma} \left\{ \frac{d\phi}{dt} + \phi(\nabla\cdot\mathbf{v}) \right\} dv + \int_\Sigma [\phi(\mathbf{v} - \boldsymbol{\nu})]\cdot\hat{\mathbf{n}}\, da\,, \tag{D1}$$

$$\frac{\hat{d}}{dt}\int_\Sigma \hat{\phi}\, da = \int_\Sigma \left\{ \frac{\hat{d}\hat{\phi}}{dt} + \hat{\phi}(\hat{\nabla}\cdot\hat{\mathbf{v}}) \right\} da\,. \tag{D2}$$

The divergence theorems in volume and on a surface are

$$\int_{D\text{-}\Sigma} \nabla\cdot\mathbf{A}\, dv + \int_\Sigma [\mathbf{A}]\cdot\hat{\mathbf{n}}\, da = \int_{\partial D\text{-}\Sigma} \mathbf{A}\cdot\mathbf{n}\, da\,, \tag{D3}$$

$$\int_\Sigma (\hat{\nabla} + 2\Omega\hat{\mathbf{n}})\cdot\hat{\mathbf{A}}\, da = \int_{\partial\Sigma} \hat{\mathbf{A}}\cdot\boldsymbol{\tau}\, dl\,, \tag{D4}$$

where

$$\hat{\nabla}_i = P_{ij}\nabla_j, \quad P_{ij} = \delta_{ij} - \hat{n}_i\hat{n}_j, \quad 2\Omega = -\nabla\cdot\hat{\mathbf{n}}\,. \tag{D5}$$

We also note that

$$\frac{\partial}{\partial t}\int_{D\text{-}\Sigma} \phi\, dv = \int_{D\text{-}\Sigma} \frac{\partial\phi}{\partial t} dv - \int_\Sigma [\![(\boldsymbol{\nu}\cdot\hat{\mathbf{n}})\phi]\!] da\,. \tag{D6}$$

[1] P. Germain, "La méthode des puissances virtuelles en mécanique des milieux continus: théorie du second gradient," J. Mécanique **12**, 235 (1974).
[2] P. Germain, "The method of virtual power in continuum mechanics, Part 2: Microstructure," SIAM J. Appl. Math. **25**, 556 (1973).
[3] G. A. Maugin, "The method of virtual power in continuum mechanics. Application to coupled fields," Acta Mech. **35**, 1 (1980).
[4] N. Daher and G. A. Maugin, "The method of virtual power in continuum mechanics. Application to media presenting singular surfaces and interfaces," Acta Mech. **60**, 217 (1986).
[5] G. A. Maugin and A. C. Eringen, "On the equations of the electrodynamics of deformable bodies of finite extent," J. Mécanique **16**, 101 (1977).
[6] A. C. Eringen, *Mechanics of Continua* (Krieger, New York, 1980), 2nd ed.
[7] P. A. Wolf and A. M. Albano, "Non-equilibrium thermodynamics of interfaces including electromagnetic effects," Physica (Utrecht) **98A**, 491 (1979).
[8] N. Daher and G. A. Maugin, "Deformable semiconductors with interfaces: Basic equations" (submitted for publication).
[9] D. Bedeaux, A. M. Albano, and P. Mazur, "Boundary conditions and non-equilibrium thermodynamics," Physica (Utrecht) **82A**, 438 (1975).
[10] G. A. Maugin, "On Maxwell's covariant equations in matter," J. Franklin Inst. **305**, 11 (1978).

[11] *Note*: *The primary* concepts in Eqs. (4.21) and (4.22) (see their derivation in Ref. 5) are those of volume and surface forces. The second parts in these equations involve *secondary* notions such as those of stresses $\mathbf{t}^{\mathrm{em}}$ and electromagnetic momentum. There exist several forms [by Abraham, Minkowski, and others; see G. A. Maugin, "Comments on the equivalence of Abraham's, Minkowski's, and others electrodynamics," Can. J. Phys. **58**, 1163 (1980)] for the latter but all these cannot be discussed without noticing that in a *nonlinear* theory of continua, whatever the choice of G—which must be consistent with that of $\mathbf{t}^{\mathrm{em}}$—adjustments occur automatically through coupled nonlinear electromagnetomechanical constitutive equations (e.g., for $^{L}\mathbf{E}$ and $^{L}\mathbf{B}$ in the bulk of the material) since neither the electromagnetic system, nor the mechanical one, considered separately, constitute closed thermodynamical systems. This is better explained in the relativistic framework: see G. A. Maugin, "On the covariant equations of the relativistic electrodynamics of continua I, II, III, IV, J. Math. Phys. **19**, 1198, 1206, 1212, 1220 (1978). The same remark holds when dealing with surface polarization and magnetization as is shown in a forthcoming work (Ref. 8, above) where nonlinear constitutive equations can be formulated for *surface* local fields akin to $^{L}\mathbf{E}$ and $^{L}\mathbf{B}$. The latter work, however, is more concerned with surface *charge* and *conduction* effects than with surface polarization and magnetization because of the nature of the studied electromagnetic materials (semiconductors).

A bimodular representation of ten-dimensional fermions

A. J. Davies and G. C. Joshi
School of Physics, University of Melbourne, Parkville, Victoria 3052, Australia

(Received 13 March 1986; accepted for publication 28 July 1986)

The fermionic and supersymmetric octonionic bimodular representations are constructed. It is shown that an octonionic formulation of the fermion leads to an intrinsically ten-dimensional world with two independent representations. Consistency conditions for an octonionic supersymmetric algebra are also discussed.

I. INTRODUCTION

There has been a surge of interest in the superstring theories of unified fundamental interaction. It is well established that such theories can be consistently formulated only in space-time dimension 10 for superstring theories, so it is of considerable interest to understand why this dimension should be special.

In the past few years the association of division algebras with supersymmetry for various N and space-time dimensions D has been studied by several authors.[1,2] Kugo and Townsend[2] made a systematic study of this problem. They showed that the sequence of association $D=3$ with real numbers, $D=4$ with complex numbers, and $D=6$ with quaternions could be understood by the isomorphism $\overline{\mathrm{SO}}(2,1)\sim \mathrm{SL}(2;\mathbf{R})$, $\overline{\mathrm{SO}}(3,1)\sim \mathrm{SL}(2;C)$, and $\overline{\mathrm{SO}}(5,1)\sim \mathrm{SL}(2;\mathbf{H})$, where $\overline{\mathrm{SO}}(s,t)$ is the universal group of the Lorentz group. For octonians, however, no such association exists as the algebra is *nonassociative*. Although several authors[2,3] had anticipated that octonions may be related to the interesting case of $D=10$, this could not be proved because of the nonassociativity of the octonions. In this paper we attempt to prove this association. We get around the problems of nonassociativity by using an alternative bimodular representation for octonions. The plan of the paper is as follows. We first present the octonionic bimodules. This formulation is then used to study (1) fermions and (2) the supersymmetric representation, which is valid under certain constraints on the octonionic Grassman number. This may provide a natural prescription for compactification along certain octonionic direction.

We introduce an octonion

$$a=a_0e_0+a_1e_1+\cdots+a_7e_7\equiv a_0e_0+a_ie_i, \tag{1}$$

where the algebra of the octonion units is given by

$$\begin{aligned} &e_0e_0=e_0,\\ &e_0e_i=e_ie_0=e_i,\\ &e_ie_j=-\delta_{ij}+\epsilon_{ijk}e_k. \end{aligned} \tag{2}$$

To completely specify the algebra we need to give the cycles for which the antisymmetric $\epsilon_{ijk}=+1$. There are eight possible variations available; we use for definiteness Cayley's original variant, with $\epsilon_{ijk}=+1$ for $(ijk)=(123)$, (145), (176), (246), (257), (347), and (365). Note that the algebra (2) implies that multiplication is nonassociative.

We define octonionic conjugation as the involution

$$\bar{a}=a_0-a_ie_i.$$

It follows that

$$a\bar{a}=\bar{a}a=a_0^2+a_1^2+a_2^2+\cdots+a_7^2\equiv|a|^2, \tag{3}$$

where the quantity $|a|$ is called the norm of a. Clearly $|a|=0$ if, and only if, $a=0$. As a consequence we may define the multiplicative inverse of a:

$$a^{-1}=\bar{a}/|a|,\quad a\neq 0. \tag{4}$$

With the properties of the norm given above, and the result $|ab|=|a|\cdot|b|$, the octonions form a division algebra. In fact, they are the most general such algebra, the others being the reals, the complex numbers, and Hamilton's quaternions.

We must be careful in defining division over the octonions. We define x to be the left quotient of a divided by b if $bx=a$, and write $x=b/a$. Similarly we define a right quotient by $x=a/b$ if x satisfies $a=xb$. In general the left quotient of two octonions will not equal the right quotient. The associator of three octonions is defined by

$$\{a,b,c\}\equiv(ab)c-a(bc). \tag{5}$$

With this definition we can write down left and right division tables for the octonionic units e_0, e_i. From these tables we can extract *two* sets of 8×8 matrices which together form the bimodule representation of the octonions.[4,5]

From the left division algebra we get the set

$$L_1=\begin{pmatrix} -i\sigma_2 & 0 & 0 & 0\\ 0 & -i\sigma_2 & 0 & 0\\ 0 & 0 & -i\sigma_2 & 0\\ 0 & 0 & 0 & i\sigma_2 \end{pmatrix},$$

$$L_2=\begin{pmatrix} 0 & -\sigma_3 & 0 & 0\\ \sigma_3 & 0 & 0 & 0\\ 0 & 0 & 0 & -\sigma_0\\ 0 & 0 & \sigma_0 & 0 \end{pmatrix},$$

$$L_3=\begin{pmatrix} 0 & -\sigma_1 & 0 & 0\\ \sigma_1 & 0 & 0 & 0\\ 0 & 0 & 0 & -i\sigma_2\\ 0 & 0 & -i\sigma_2 & 0 \end{pmatrix},$$

$$L_4=\begin{pmatrix} 0 & 0 & -\sigma_3 & 0\\ 0 & 0 & 0 & \sigma_0\\ \sigma_3 & 0 & 0 & 0\\ 0 & -\sigma_0 & 0 & 0 \end{pmatrix}, \tag{6a}$$

0022-2488/86/123036-04$02.50

$$L_5=\begin{pmatrix}0&0&-\sigma_1&0\\0&0&0&i\sigma_2\\\sigma_1&0&0&0\\0&i\sigma_2&0&0\end{pmatrix},$$

$$L_6=\begin{pmatrix}0&0&0&-\sigma_0\\0&0&-\sigma_3&0\\0&\sigma_3&0&0\\\sigma_0&0&0&0\end{pmatrix},$$

$$L_7=\begin{pmatrix}0&0&0&-i\sigma_2\\0&0&-\sigma_1&0\\0&\sigma_1&0&0\\-i\sigma_2&0&0&0\end{pmatrix},$$

$$L_0=\mathbf{1}.$$

The σ matrices are the familiar Pauli matrices

$$\sigma_0=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$
$$\sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}. \tag{6b}$$

There is a similar representation for the R matrices. Together they satisfy the relations

$$\begin{aligned}&\{L_i,L_j\}=-2\delta_{ij}L_0,\\&\{R_i,R_j\}=-2\delta_{ij}R_0,\\&R_iR_j+L_jL_i=C^l_{ij}(L_l+R_l),\end{aligned}\tag{6c}$$

where $i,j,l=1,2,\dots,7$ and C^l_{ij} are structure constants.

It is amusing to note that

$$\det(a_0L_0+a_iL_i)=|a|^8=(a_0^2+a_1^2+\cdots+a_7^2)^4$$

and, since the L_i matrices are orthogonal, we see explicitly how the octonions are related to the group SO(8).[2,4]

II. THE FERMIONS

To derive our Dirac matrices we introduce two octonionic spinors

$$\phi=\begin{pmatrix}a\\b\end{pmatrix},\quad \chi=\begin{pmatrix}b\\-a\end{pmatrix}. \tag{7}$$

In lower dimensions over the other division algebras these correspond to chiral and antichiral spinors, respectively.[2] Note that $\chi=\epsilon\phi$, with

$$\epsilon=\begin{pmatrix}0&1\\-1&0\end{pmatrix}.$$

We form the product

$$\phi\phi^+=\begin{pmatrix}a\bar a&a\bar b\\b\bar a&b\bar b\end{pmatrix}\equiv\begin{pmatrix}V_0+V_9&\bar V\\V&V_0-V_9\end{pmatrix}\equiv V_\mu\gamma^\mu,$$

where $V=V_1e_0+V_2e_1+\cdots+V_8e_7$ and V_0,V_9 are real. We thus obtain a ten-dimensional space with the set of matrices given by

$$\gamma^\mu=\begin{pmatrix}1&0\\0&1\end{pmatrix},\ \begin{pmatrix}0&1\\1&0\end{pmatrix},$$
$$\begin{pmatrix}0&-e_i\\e_i&0\end{pmatrix},\ \begin{pmatrix}1&0\\0&-1\end{pmatrix},\quad i=1,2,\dots,7. \tag{8a}$$

Similarly, setting $\chi\chi^+\equiv V_\mu\tilde\gamma^\mu$ we obtain

$$\tilde\gamma^\mu=\begin{pmatrix}1&0\\0&1\end{pmatrix},\ \begin{pmatrix}0&-1\\-1&0\end{pmatrix},$$
$$\begin{pmatrix}0&e_i\\-e_i&0\end{pmatrix},\ \begin{pmatrix}-1&0\\0&1\end{pmatrix}. \tag{8b}$$

The sets $\gamma^\mu,\tilde\gamma^\mu$ are the ten-dimensional analog of $\sigma^\mu,\bar\sigma^\mu$ in four dimensions.[2] It is easily checked that

$$\gamma^\mu\tilde\gamma^\nu+\gamma^\nu\tilde\gamma^\mu=\tilde\gamma^\mu\gamma^\nu+\tilde\gamma^\nu\gamma^\mu=2\eta^{\mu\nu}\mathbf{1}, \tag{8c}$$

where

$$\eta^{\mu\nu}=\begin{cases}+1, & \mu=\nu=0,\\ -1, & \mu=\nu\neq 0,\\ 0, & \text{otherwise.}\end{cases}$$

If we now substitute the L_i *for the* e_i in $\gamma^\mu,\tilde\gamma^\mu$ we obtain two sets of 16×16 matrices which we use to write our Dirac set of 32×32 matrices:

$$\Gamma^\mu=\begin{pmatrix}0&\gamma^\mu\\\tilde\gamma^\mu&0\end{pmatrix}. \tag{9}$$

By virtue of (8) these satisfy

$$\Gamma^\mu\Gamma^\nu+\Gamma^\nu\Gamma^\mu=2\eta^{\mu\nu}\mathbf{1},$$

and hence

$$\mathrm{tr}(\Gamma^\mu\Gamma^\nu)=32\eta^{\mu\nu}. \tag{10}$$

Of course, there is another set, obtained by replacing e_i by R_i, which also obey the Dirac algebra (10). We now turn to some simple applications of the set defined in (9).

We write the Dirac equation as

$$(i\Gamma^\mu\partial_\mu-m)\Psi=0,$$

with Ψ a 32-component complex spinor.

The matrices representing P, C, and T are obtained in the same manner as those in four dimensions, with the proviso that complex conjugation be replaced by octonionic conjugation, and that this process takes $L_i\to-L_i$. We then obtain (up to phases):

$$\begin{aligned}&P=\Gamma^0,\\&C=\Gamma^0\Gamma^1\Gamma^9,\\&T=\Gamma^0\Gamma^2\Gamma^3\Gamma^4\Gamma^5\Gamma^6\Gamma^7\Gamma^8.\end{aligned}\tag{11}$$

We also have the generalization of γ^5 in $D=4$:

$$\Gamma^{11}=\Gamma^0\Gamma^1\cdots\Gamma^9=\begin{pmatrix}1&0\\0&-1\end{pmatrix},$$

and we can therefore form helicity projection operators

$$P_\pm=\tfrac12(1\pm\Gamma^{11}).$$

To illustrate some of this consider the rest frame Dirac equation in momentum space:

$$(\not p - m)\Psi = 0. \tag{12}$$

Since $p^\mu = (E;0,...,0)$ in the rest frame, (12) becomes

$$\begin{pmatrix} -m & E \\ E & -m \end{pmatrix}\Psi = 0. \tag{13}$$

The solutions of (13) are

(i) $E = +m$: $\Psi_+ = \begin{pmatrix}\psi \\ \psi\end{pmatrix}$; $\Gamma^0\Psi_+ = (+1)\Psi_+$,

(ii) $E = -m$: $\Psi_- = \begin{pmatrix}\psi \\ -\psi\end{pmatrix}$; $\Gamma^0\Psi_- = (-1)\Psi_-$.

We see that particle and antiparticle solutions have the opposite parity. Similar applications may be demonstrated in analogy with the well-known four-dimensional Dirac theory.

III. SUPERSYMMETRY

We show here that in constructing the simplest ten-dimensional supersymmetry, an on-shell model with one octonionic scalar field, and an octonionic spinor, we encounter an intriguing result—the supersymmetry parameters cannot be the most general type of octonions.

We begin with the ten-dimensional representations Eqs. (8a)–(8c) where $e_i \to L_i$ (and R_i) for $i = 1,...,7$. The action is given by

$$S = \int d^{10}x \operatorname{Tr}\{-\mathbf{A}^*\Box\mathbf{A} + i\lambda^+ i\tilde{\not\partial}\lambda\}, \tag{14}$$

where $\mathbf{A}$ is an octonionic scalar field, λ is an octonionic spinor, and the differential operators are given by $(\mu = 1-10)$

$$\begin{aligned} &\Box \mathbf{1} = \not\partial\tilde{\not\partial} = \tilde{\not\partial}\not\partial = \partial^\mu\partial_\mu, \\ &\not\partial = \gamma^\mu\partial_\mu, \\ &\tilde{\not\partial} = \tilde{\gamma}^\mu\partial_\mu. \end{aligned} \tag{15}$$

The on-shell field equations are

$$\Box\mathbf{A} = \not\partial\lambda = 0. \tag{16}$$

And the symmetry transformations are

$$\begin{aligned} &\delta\mathbf{A} = \epsilon^+\lambda, \\ &\delta\lambda = -i\gamma^\mu\epsilon\partial_\mu\mathbf{A}, \end{aligned} \tag{17}$$

where

$$\epsilon = \begin{pmatrix}\mathbf{a} \\ \mathbf{b}\end{pmatrix}$$

is an octonionic spinor with $\mathbf{a} = e_i a_i$ and $\mathbf{b} = e_i b_i$ and a_i and b_i are complex Grassmann numbers. It should be noted that in general $\epsilon^2 = 0$ for complex Grassmann numbers, $\epsilon^4 = 0$ for quaternionic Grassmann numbers, and $\epsilon^8 = 0$ for octonionic Grassmann numbers.

The commutator of two supersymmetric transformations on $\mathbf{A}$ is given by (in the notation of Ref. 2):

$$[\delta_1,\delta_2]\mathbf{A} = 2\xi^\alpha\partial_\alpha\mathbf{A},$$

with

$$\xi^\alpha = (i/2)(\epsilon_1^+\gamma^\alpha\epsilon_2 - \epsilon_2^+\gamma^\alpha\epsilon_1). \tag{18}$$

It can be shown by direction calculation that ξ^α is not an octonion—that is, all but the coefficient of e_0 in ξ^α vanish; for definiteness we take

$$\epsilon_1 = \begin{pmatrix}\mathbf{a} \\ \mathbf{b}\end{pmatrix}, \quad \epsilon_2 = \begin{pmatrix}\mathbf{c} \\ \mathbf{d}\end{pmatrix}.$$

The commutator of two supersymmetric transformations on λ yields

$$[\delta_1,\delta_2]\lambda = -i\gamma^\beta(\epsilon_2\epsilon_1^+ - \epsilon_1\epsilon_2^+)\partial_\beta\lambda. \tag{19}$$

In the $D = 6$ case with quaternionic fields the identity

$$\epsilon_2\epsilon_1^+ - \epsilon_1\epsilon_2^+ = -i\xi^\alpha\tilde{\gamma}_\alpha \tag{20}$$

can be used to close the algebra.[2] In our $D = 10$ case the identity (20) holds if, and only if,

$$(a_i d_j - a_j d_i) = 0, \quad i \neq j = 1,2,...,7. \tag{21}$$

We now discuss two types of cases where the above conditions are satisfied:

(a) The octonionic entries in the ϵ spinors choose a "special" direction in the octonionic space. For definiteness we take the e_7 direction and demanding that the entries of ϵ are of the form

$$\epsilon_1 = \begin{pmatrix} a_0e_0 + a_7e_7 \\ b_0e_0 + b_7e_7\end{pmatrix}, \quad \epsilon_2 = \begin{pmatrix} c_0e_0 + c_7e_7 \\ d_0e_0 + d_7e_7\end{pmatrix}, \tag{22}$$

leads to a closed algebra on shell, as Eq. (20) is finally satisfied. It is amusing to note that as supersymmetry chooses an octonionic direction one can split octonions along this direction, i.e.,

$$\begin{aligned} &u_j = \tfrac{1}{2}(e_j + ie_{j+3}), \quad u_j^* = \tfrac{1}{2}(e_j - ie_{j+3}), \quad j = 1,2,3, \\ &u_0 = \tfrac{1}{2}(e_0 + ie_7), \quad u_0^* = \tfrac{1}{2}(1 - ie_7). \end{aligned} \tag{23}$$

This gives us a natural prescription of compactification along e_7 direction.

(b) The other way to close the on-shell algebra Eqs. (19) and (20) is to write the supersymmetric transformations as

$$\epsilon_1 = \begin{pmatrix}\mathbf{a} \\ 0\end{pmatrix}, \quad \epsilon_2 = \begin{pmatrix}\mathbf{c} \\ 0\end{pmatrix}. \tag{24}$$

For these truncated ϵ's, $\xi^\nu = 0$ if $\nu \neq 0,9$. This give us

$$-i\xi^\nu\tilde{\gamma}_\nu = \tfrac{1}{2}(-\bar{\mathbf{a}}\mathbf{c} + \bar{\mathbf{c}}\mathbf{a})(\tilde{\gamma}_0 + \tilde{\gamma}_9). \tag{25}$$

From Eqs. (8a) and (8b) we have

$$\frac{1}{2}(\tilde{\gamma}_0 + \tilde{\gamma}_9) = \begin{pmatrix}1 & 0 \\ 0 & 0\end{pmatrix}. \tag{26}$$

And Eq. (20) is satisfied:

$$-i\xi^\nu\tilde{\gamma}_\nu = \epsilon_1\epsilon_2^+ - \epsilon_2\epsilon_1^+.$$

These types of solutions have recently been studied by Adler,[6] in connection with the quaternionic field theories.

Finally, we note that if we start with an octonionic vector and follow a procedure as in Eqs. (8a) and (8b), we find this is associated with a 26-dimensional space via Jordan matrices $[M_3^8$ with $\operatorname{Tr}(M) = 1]$.[7]

ACKNOWLEDGMENTS

One of the authors (G.C.J.) would like to thank Professor R. Blankenbecker and the Theory Group at SLAC for their warm hospitality.

He (G.C.J.) would also like to thank the Australia–America foundation for a Senior Fulbright Grant.

[1]E. A. Ivanov, Phys. Lett. B **117**, 59 (1982); V. I. Ogievetsky and E. Sokatchev, *ibid.* **79**, 222 (1978); A. S. Galperin, E. A. Ivanov, and V. I. Ogievetsky, JETP Lett. **33**, 168 (1981); R. Coquereaux, Phys. Lett. B **115**, 389 (1982); J. Lukierski and A. Nowicki, Fortschr. Phys. **30**, 35 (1982); J. Lukierski and P. Minnaert, Phys. Lett. B **129**, 392 (1983); J. Lukierski, ICTP preprint IC/83/188, 1983.

[2]T. Kugo and P. Townsend, Nucl. Phys. B **221**, 357 (1983).

[3]T. L. Curtright, Phys. Rev. Lett. **48**, 1704 (1982).

[4]L Sorgsepp and J. Löhmus, Had. J. Phys. **2**, 1388 (1979).

[5]G. C. Joshi, Melbourne University preprint UM-P-85/83, 1985.

[6]S. L. Adler, Princeton IAS preprints 85-0502 and 85-0646, 1985.

[7]P. Jordan, Z. Phys. **80**, 285 (1933); A. Gamba, J. Math. Phys. **8**, 775 (1967); L. C. Biedenharn and P. Truini, SLAC-PUB-2802, 1981; P. Gürsey, in *International Symposium on Mathematical Problems in Theoretical Physics*, edited by A. Araki (Springer, New York, 1975).

Anomalies in supersymmetric nonlinear σ models based on E_l-type groups

Takeshi Moriya
Department of Physics, Tohoku University, Sendai 980, Japan
Yukinori Yasui
Department of Physics, Osaka City University, Osaka 558, Japan

(Received 6 January 1986; accepted for publication 14 August 1986)

The anomalies in exceptional-type supersymmetric nonlinear σ models are investigated. The cohomology rings of Kähler manifolds $E_6/\mathrm{Spin}(10)\times U(1)$, $E_7/SU(5)\times SU(3)\times U(1)$, and $E_8/SO(10)\times SU(3)\times U(1)$ and the Chern classes of the tangent bundles of these manifolds are calculated explicitly. It is shown that the $E_6/\mathrm{Spin}(10)\times U(1)$ model is anomaly-free but the $E_7/SU(5)\times SU(3)\times U(1)$ and $E_8/SO(10)\times SU(3)\times U(1)$ models are anomalous.

I. INTRODUCTION

Supersymmetric nonlinear σ models based on exceptional groups[1–3] have some interesting properties especially in that they can accommodate quarks and leptons required in the standard GUT models as quasi-Nambu–Goldstone fermions[4] in a quite natural manner. In fact, it can be shown group theoretically[1–4] that the nonlinear realizations on Kähler manifolds $E_6/\mathrm{Spin}(10)\times U(1)$, $E_7/SU(5)\times SU(3)\times U(1)$, and $E_8/SO(10)\times SU(3)\times U(1)$ admit, respectively, one, three, and four families[5] of quarks and leptons. The closed forms of Lagrangians also have been derived recently.[1–3] So, it is worthwhile to investigate them in more detail as to whether they are good candidates for realistic models at least as effective theories at GUT level.

Recently it was shown by Moore and Nelson[6] that several nonlinear σ models with Weyl fermions suffer from a certain anomaly analogous to that of non-Abelian gauge theories. If the models of the exceptional type we are considering contain this anomaly, they are unsuitable for describing dynamics even as effective theories, because the anomaly makes the theories ill-defined dynamically.

In this paper it will be shown that the anomaly pointed out by Moore and Nelson does not exist in the $E_6/\mathrm{Spin}(10)\times U(1)$ model but does exist in the $E_7/SU(5)\times SU(3)\times U(1)$ and $E_8/SO(10)\times SU(3)\times U(1)$ models. As was clarified by Moore and Nelson,[6] this kind of anomaly originates from the nontrivial topology of the space of mappings from the space-time to the coset space G/H in which the scalar fields take their values. In four-dimensional space-time and with Weyl fermions, as was pointed out by Zumino,[7] supersymmetry requires the coset space G/H to admit a complex (Kähler) structure (the choice of a complex structure is determined by the fermion representation). This designates one complex structure on the tangent bundle $T(G/H)$. By the Atiyah–Singer family index theorem,[8] the anomaly is related to the characteristic class of $T(G/H)$. The result of Ref. 6 is

$$\text{Anomaly} = \int_{S^2\times(\text{space-time})} \hat{\varphi}^*\,\mathrm{ch}_3(T(G/H)), \qquad (1.1)$$

where $\hat{\varphi}$: $S^2\times(\text{space-time})\to G/H$ and $\hat{\varphi}^*$ is the pullback induced by $\hat{\varphi}$. In Sec. II, we briefly review this anomaly and also interpret it relating to the group cocycle introduced by Faddeev[9] in the study of the non-Abelian gauge anomaly.

Our motivation as physicists lies in the investigation of whether there is an anomaly in supersymmetric nonlinear σ models based on the E_l-type group, but our main task is a purely mathematical one; to calculate $\mathrm{ch}_3(T(G/H))$, an element of $H^6(G/H)$, and to evaluate its pullback to H^6(space-time$\times S^2$).

So, let us roughly outline our way of doing mathematics. Our calculation of $\mathrm{ch}_3(T(G/H))$ is essentially based on the general work of Borel[10] and Borel and Hirzenbruch.[11] In this paper we will do, at every stage of our investigation, some necessary and introductory review of their work. After that we will present our results of the calculation.

When a $2n$-dimensional real manifold G/H admits a complex structure (it is in our case), its tangent space at any point can be seen as an n-dimensional complex vector space. The subgroup H of G acts on G/H as an isotropy group and on tangent space as a linear unitary transformation. This is called a linear unitary isotropy representation of H which we denote

$$\iota_C:\ H\to\overline{H}\subset U(n). \qquad (1.2)$$

It is proved in Ref. 11 that the Chern class of the tangent bundle $T(G/H)$ is equivalent to that of a principal $U(n)$ bundle over G/H, which is induced, by the map ι_C: $H\to\overline{H}$, from the bundle η (total space G, base space G/H, structure group H) and is denoted by η_C. The Chern class of the bundle η_C is given by a certain pullback of that of the universal $U(n)$ bundle. The authors of Ref. 11 analyze the nature of some relevant bundle mappings and pullbacks induced by them, and finally conclude

$$c(T(G/H)) = c(\eta_C) \approx \prod_{j=1}^{n}(1+\omega_j), \quad \mathrm{mod}\, I^+(G), \qquad (1.3)$$

where ω_j ($j=1$–n) are the so-called "complementary roots" and $I^+(G)$ is the ring of Weyl invariant polynomials[12] of the root of G without the constant term.

The complementary roots are the roots of G that characterize a G-invariant complex structure on G/H. We regard a $2n$-dimensional real manifold G/H as an n-dimensional complex manifold by identifying G/H with G^C/P, where G^C is the complexification of G obtained by complexifying the Lie algebra of G and P is a closed subgroup of G^C such

 0022-2488/86/123040-11$02.50

that $P \cap G = H$. Then, the complementary roots are such that

{all roots of G} $-$ {roots relevant to P};

they are the weights of the linear isotropy unitary representation of H.

Thus our task is to calculate the complementary roots and find $I^+(G)$, both represented in a suitable manner conveniently to evaluate the pullback of $c(T(G/H))$ to $S^2 \times$ (space-time).

The construction of the sections are as follows. In Sec. III, we present the roots and invariant polynomials[12] of the Weyl groups for the Lie algebras of E_l for G and those of A_l and D_l for H. In Sec. IV, we calculate the cohomology rings of G/H ($G = E_l$). We represent them by several sets of coordinates of the maximal torus, which are suited to see the manifest invariance under the Weyl group of H. This parametrization is necessary to evaluate the pullback of the Chern character of $T(G/H)$, which will be done finally in Sec. VII. In Sec. V, the complementary roots ω_j for E_l/H are calculated, with the introduction to complex structure. In Sec. VI, Eq. (1.3), especially $\mathrm{ch}_3(T(G/H))$, is calculated using the results of Secs. III and V, with a brief review of the relevant part of Ref. 11. Finally in Sec. VII the pullback of the Chern character into $H^*(S^2 \times$ space-time) is evaluated using the homotopy argument and the presence or absence of the anomaly is concluded to each model with some physical considerations. In the Appendix the roots of E_8^C are given, from which the roots of E_7^C and E_6^C are easily derived.[13]

II. ANOMALY

Before detailed discussions of the anomalies in exceptional type nonlinear σ models, we briefly describe the generalities of the anomaly in a nonlinear σ model following Ref. 6. Let $\mathrm{Map}(X,G/H)$ be the set of differentiable mappings from the space-time X, which we suppose to be compactified to S^4, to a homogeneous space G/H admitting some complex structure where G is a compact Lie group and H some subgroup of G. Here $\mathrm{Map}(X,G/H)$ becomes a topological space with the appropriate topology. Given a $\varphi \in \mathrm{Map}(X,G/H)$, the coordinates $\{y^i\}$ of G/H are related to the coordinates $\{x^\mu\}$ of X, i.e., the image points of x^μ are $y^i = \varphi^i(x)$, which are differentiable functions.

Let $g_{ij}(y)$ be the metric tensor of G/H. The σ model action is then

$$S_b = \int d^4x\, g_{ij^*}\, \partial_\mu \varphi^i \partial^\mu \varphi^{j^*}. \qquad (2.1)$$

Weyl fermions ψ^i coupled to the fields φ^i are regarded as sections of the tensor product fiber bundle $E_\varphi^\pm \equiv S^\pm \otimes \varphi^*(T(G/H))$, where $S^\pm$ are the spin bundles of chirality ± 1 and $\varphi^*(T(G/H))$ is the pullback of the tangent bundle $T(G/H)$. Each $\varphi \in \mathrm{Map}(X,G/H)$ gives a Dirac operator $\not{D}_\varphi$: $\Gamma(E_\varphi^+) \to \Gamma(E_\varphi^-)$ (Γ denotes the vector space of sections). In local coordinates

$$(\not{D}_\varphi \psi)^i = \gamma^\mu(\partial_\mu \delta^i_k + \Gamma^i_{jk}(\varphi)\partial_\mu \varphi^j)[(1+\gamma_5)/2]\, \psi^k, \qquad (2.2)$$

where Γ^i_{jk} is the Riemann connection of G/H. The invariant action for $\psi^i \in \Gamma(E_\varphi^+)$ is

$$S_f = \int d^4x\, g_{ij^*}\, \bar{\psi}^i \sqrt{-1}\,(\not{D}_\varphi \psi)^j \qquad (2.3)$$

and the effective action is formally written as

$$Z(\varphi) = \int [d\bar{\psi}]\,[d\psi] \exp(-S_f)$$

$$\equiv \det(\sqrt{-1}\,\not{D}_\varphi). \qquad (2.4)$$

The result of Ref. 6 is that we should not regard $Z(\varphi)$ as an ordinary function on $\mathrm{Map}(X,G/H)$ and if we stubbornly insist on it, the function generally has a singularity. It is a section of the complex line bundle L over $\mathrm{Map}(X,G/H)$. So we must cover $\mathrm{Map}(X,G/H)$ by patches $\{\rho_\alpha\}$, and define the effective action $Z^\alpha(\varphi)$ patchwise: for $\varphi \in \rho_\alpha \cap \rho_\beta$, $Z^\alpha(\varphi)$ must be such that

$$Z^\alpha(\varphi) = g_{\alpha\beta}(\varphi) Z^\beta(\varphi), \qquad (2.5)$$

where $g_{\alpha\beta}(\varphi)$ is a continuous map

$$\rho_\alpha \cap \rho_\beta \to \mathrm{GL}(1,\mathbb{C}) = \mathbb{C}^* = \mathbb{C} - \{0\}, \qquad (2.6)$$

called the transition function.

Since the structure group $\mathrm{GL}(1,\mathbb{C})$ of L is reducible to $\mathrm{U}(1)$, we can put $g_{\alpha\beta}(\varphi) = e^{\sqrt{-1}\,\alpha_1(\varphi)}$. Thus we may regard Eq. (2.5) as a unitary ray representation of G that acts on $\mathrm{Map}(X,G/H)$ and α_1 as a one-cocycle of G, i.e., $\alpha_1 \in H^1(\mathrm{Map}(X,G/H);\, \mathbb{C}^*)$.[9]

For the theory to have a well-defined quantum behavior, $Z(\varphi)$ must be an ordinary function on $\mathrm{Map}(X,G/H)$, so the anomaly is defined as the nontriviality of L, i.e., the nonvanishing "twist" of L. The twist of L is determined by the first-Chern class $c_1(L)$ or α_1, which is an element of

$$H^2(\mathrm{Map}(X,G/H); Z) = H^1(\mathrm{Map}(X,G/H);\, \mathbb{C}^*). \qquad (2.7)$$

This equality[9] is derived from the exact sequence,

$$0 \to Z \underset{\text{inclusion}}{\to} \mathbb{C} \underset{\text{exp}}{\to} \mathbb{C}^* \to 0. \qquad (2.8)$$

Thus

$$\mathrm{Anomaly} = \int_{S^2} \mathrm{ch}_1(L)$$

$$= \int_{S^2 \times X} \hat{\varphi}^*(\mathrm{ch}_3(T(G/H)))\hat{A}(X), \qquad (2.9)$$

where S^2 is a noncontractible two-sphere imbedded in $\mathrm{Map}(X,G/H)$, $\hat{\varphi}$ is a mapping $S^2 \times X \to G/H$, and $\hat{A}(X)$ is the Dirac genus. The last equality comes from the family index theorem.[8] Taking $X = S^4$ the above expression becomes

$$\int_{S^2 \times S^4} \hat{\varphi}^*\, \mathrm{ch}_3(T(G/H)). \qquad (2.10)$$

In the following sections we evaluate this formula using the roots of G.

III. ROOTS AND INVARIANT POLYNOMIALS OF THE WEYL GROUP

Let $\mathscr{G}$ be the Lie algebra of a semisimple group G with the rank l and the dimension $\mathfrak{d} = l + 2m$ and $\mathfrak{h}$ be a Cartan subalgebra of it, i.e., the Lie algebra of a maximal torus of G. The roots of $\mathscr{G}$ with respect to $\mathfrak{h}$, $\pm a_i$ $(1<i<m)$, define real-valued linear forms on $\mathfrak{h}$. We have then $\mathscr{G}^C$, the complexification of $\mathscr{G}$, and $\mathfrak{h}^C$, a Cartan subalgebra of $\mathscr{G}^C$, such that $\mathfrak{h}^C \cap \mathscr{G} = \mathfrak{h}$. We denote the roots of $\mathscr{G}^C$ with respect to $\mathfrak{h}^C$ by $\pm 2\pi\sqrt{-1}\, a_i$ $(1<i<m)$ and choose, as the basis of $\mathscr{G}^C$, $h_i \in \mathfrak{h}^C$ $(1<i<l)$ and $e_{\pm i}$ $(1<i<m)$ satisfying

$$\begin{aligned}
&[h_i, h_j] = 0,\\
&[h_i, e_{\pm j}] = \pm 2\pi\sqrt{-1}\, a_j(h_i) e_{\pm j},\\
&[e_j, e_k] = N_{j,k}\, e_{j+k},\\
&[e_{-i}, e_{+i}] = \text{linear combination of } h_i\text{'s},
\end{aligned} \tag{3.1}$$

where each $N_{j,k}$ is a real number that does not vanish if $a_j + a_k$ is a root again. Here $\mathscr{G}$ is then spanned by h_i $(1<i<l)$, $e_j + e_{-j}$, and $\sqrt{-1}(e_j - e_{-j})$ $(1<j<m)$ over the reals.

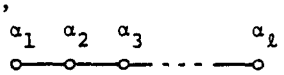
The scalar product on $\mathfrak{h}^*$ (dual space of $\mathfrak{h}$) is induced by the Killing form (see the Appendix), from now on denoted (,), and the Weyl chamber W is defined by $\{y \in \mathfrak{h}^* | (a_k, y) > 0, a_k\ (1<k<l) \text{ are simple roots}\}$. The Weyl group $W(G)$ is generated by

$$\begin{aligned}
&\{\sigma_{a_i} | a_i\ (1<i<l) \text{ are simple roots}\},\\
&\sigma_{a_i}(\lambda) = \lambda - 2\frac{(\lambda, a_i)}{(a_i, a_i)} a_i, \quad \lambda \in \mathfrak{h}^*.
\end{aligned} \tag{3.2}$$

Especially for a, $b \in$root, $\sigma_a(b)$ is a root again.

We denote by $I(G)$ the ring of $W(G)$-invariant polynomials[12] on $\mathfrak{h}^*$ with real coefficients, by $S(y_i, \ldots, y_n)$ the ring of symmetric polynomials of y_i's, by $S_i(y_1, \ldots, y_n)$ the ith elementary symmetric polynomial. We present here some results for later use.

(1) $\mathscr{G} = A_l$ $(l > 1)$,

Dynkin diagram,  ;

simple roots, $\alpha_i = x_i - x_{i+1}$ $(1<i<l)$;

roots, $\pm(x_i - x_j)$ $(i \neq j,\ 1<i<j<l+1)$, $\sum_i x_i = 0$ (3.3)

(x_i's denote the coordinates of the maximal torus)

$$\begin{aligned}
&\sigma_{\alpha_i}\ (i = 1\text{–}l) \text{ generate } W(\mathrm{SU}(l+1)),\\
&\sigma_{\alpha_i}:\ x_i \leftrightarrow x_{i+1},
\end{aligned} \tag{3.4}$$

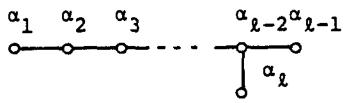
where $\leftrightarrow$ means the permutation

$$I(A_l) = S(x_1, \ldots, x_{l+1}) \tag{3.5}$$

and

$\{S_i(x_1, \ldots, x_{l+1});\ i = 2\text{–}l+1\}$ are generators of $I(A_l)$.

(2) $\mathscr{G} = D_l$ $(l > 4)$,

Dynkin diagram,  ;

simple roots, $\alpha_i = x_i - x_{i+1}$ $(1<i<l-1)$,

$\alpha_l = x_{l-1} + x_l$;

roots, $\pm(x_i + x_j)$, $\pm(x_i - x_j)$ $(1<i<j<l)$; (3.6)

σ_{α_i} $(i = 1\text{–}l)$ generate $W(\mathrm{SO}(2l))$,

σ_{α_i}: $x_i \leftrightarrow x_{i+1}$ $(1<i<l-1)$

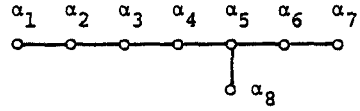
and

$$\sigma_{\alpha_l}:\ x_{l-1} \leftrightarrow -x_l. \tag{3.7}$$

$$I(D_l) = S(x_1^2, \ldots, x_l^2) \times \mathbf{R}\, x_1 x_2 x_3 \cdots x_l. \tag{3.8}$$

$I(D_l)$ is generated by $\{S_i(x_1^2, \ldots, x_l^2)\}$ $(l<i<l-1)$, $S_l(x) = x_1 x_2 \cdots x_l\}$.

(3) $\mathscr{G} = \mathrm{E}_8$ (see the Appendix),

Dynkin diagram,  ;

simple roots, $\alpha_i = x_i - x_{i+1}$ $(1<i<7)$,

$\alpha_8 = x_6 + x_7 + x_8$;

$$\begin{aligned}
\text{roots,}\quad &\pm(x_i - x_j) \quad (1<i<j<9),\\
&\pm(x_i + x_j + x_k) \quad (1<i<j<k<9),\ \sum_i x_i = 0;
\end{aligned} \tag{3.9}$$

σ_{α_i} $(i = 1\text{–}8)$ generate $W(\mathrm{E}_8)$,

$$\begin{aligned}
&\sigma_{\alpha_i}:\ x_i \leftrightarrow x_{i+1} \quad (1<i<7),\\
&\sigma_{\alpha_8}: \begin{cases} x_i \to x_i + \frac{1}{3}(x_6 + x_7 + x_8) & (1<i<5),\\ x_i \to x_i - \frac{2}{3}(x_6 + x_7 + x_8) & (6<i<8), \end{cases}
\end{aligned} \tag{3.10}$$

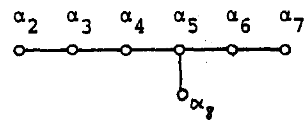
It is known[12] that $I(\mathrm{E}_8)$ is generated by

$\{I_i;\ i = 2,8,12,14,18,20,24,30\}$,

where the index denotes the degree of the polynomial, especially in our parametrization $\{x_i\}$,

$$I_2 = \mathbf{R}\left(\sum_{i=1}^{8} x_i^2 + \left(\sum_{i=1}^{8} x_i\right)^2\right). \tag{3.11}$$

(4) $\mathscr{G} = \mathrm{E}_7$. We may replace the diagram of E_7 with that of E_8 when α_1 is eliminated.

Dynkin diagram,  ;

simple roots, α_i $(i = 2\text{–}8)$;

$$\begin{aligned}
\text{roots,}\quad &\pm(x_i - x_j) \quad (2<i<j<8),\\
&\pm(x_i + x_j + x_k) \quad (2<i<j<k<8),\\
&\pm(x_2 + x_3 + \cdots x_8 - x_i) \quad (2<i<8);
\end{aligned} \tag{3.12}$$

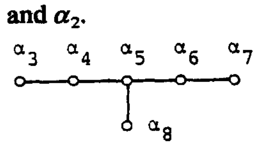
σ_{α_i} $(i = 2\text{–}8)$ generate $W(\mathrm{E}_7)$ and $I(\mathrm{E}_7)$ is generated[12] by

$\{I_i;\ i = 2,6,8,10,12,14,18\}$,

especially

$$I_2 = \mathbf{R}\left(2\sum_{i=2}^{8} x_i^2 + \left(\sum_{i=2}^{8} x_i\right)^2\right). \tag{3.13}$$

(5) $\mathscr{G} = \mathrm{E}_6$. We may also replace the diagram with that of E_8 eliminated by α_1 and α_2.

Dynkin diagram,  ;

simple roots, $\alpha_i \quad (i=3\text{–}8)$;

roots, $\pm(x_i - x_j) \quad (3 \leqslant i < j \leqslant 8)$,

$$\pm(x_i + x_j + x_k) \quad (3 \leqslant i < j < k \leqslant 8),$$

$$\pm(x_3 + x_4 + \cdots x_8); \tag{3.14}$$

σ_{α_i} $(i = 3\text{–}8)$ generate $W(\mathrm{E}_6)$ and $I(\mathrm{E}_6)$ is generated[12] by

$$\{I_i;\quad i = 2,5,6,8,9,12\},$$

especially

$$I_2 = \mathbf{R}\left(3\sum_{i=3}^{8} x_i^2 + \left(\sum_{i=3}^{8} x_i\right)^2\right). \tag{3.15}$$

IV. COHOMOLOGY RING OF G/H

The Chern class $c(T(G/H))$ is an element of the cohomology ring $H^*(G/H)$. We first determine $H^*(G/H)$ for the exceptional type G/H by using roots of G. Let $H^i(X)$ be the ith cohomology group of X with real coefficients and $H^*(X)$ be a direct sum of all $H^i(x)$'s. Then $H^*(X)$ becomes a ring under the cup product. A map $f: X \to Y$ induces a map between cohomology rings, $f^*: H^*(Y) \to H^*(X)$. We denote a fiber bundle schematically by

$$F \xrightarrow{i} E \xrightarrow{\pi} B \quad \text{or} \quad (E,\pi,B,F), \tag{4.1}$$

where F, E, and B are the fiber, the total space, and the base space, respectively, and i and π are the inclusion and the projection. Let H be a subgroup of G with the same rank as G and $T^l = (\mathrm{U}(1))^l$ be the maximal torus of G (and H). In order to evaluate $H^*(G/H)$, one may use the cohomology ring of the classifying space. We denote the universal bundles for the groups G, H, and T^l by (EG, π, BG, G), (EH, π', BH, H), and (ET^l, π'', BT^l, T^l), where BG, BH, and BT^l are classifying spaces. Since T^l acts on EH and EH is contractible by the definition, $(EH, p, EH/T^l, T^l)$ is the universal $(\mathrm{U}(1))^l$ bundle, i.e., $EH/T^l = BT^l$. One may thus consider

$$\begin{tikzcd} & EH \arrow[d, "p"] \\ H/T^l \arrow[r, "i_1"] & EH/T^l = BT^l \arrow[d, "\rho(T^l,H)"] \\ G/H \arrow[r, "i_2"] & BH \arrow[d, "\rho(H,G)"] \\ & BG \end{tikzcd} \tag{4.2}$$

where i_1 and i_2 are inclusions and p, $\rho(T^l, H)$, and $\rho(H,G)$ are projections. It is known[10] that i_2^* is surjective, $\ker(i_2^*) = \mathrm{Im}\,\rho^*(H,G)$, $\rho^*(H,G)$, and $\rho^*(T^l, H)$ are injective, and

$$\rho^*(T^l, H)H^*(BH) = H^*(BT^l)^{W(H)}, \tag{4.3}$$

where $H^*(BT^l)^{W(H)}$ is the $W(H)$-invariant subring of $H^*(BT^l)$. Therefore one has from the diagram,

$$\begin{aligned} H^*(G/H) &= H^*(BH)/\rho^*(H,G)H^+(BG) \\ &= \rho^*(T^l,H)H^*(BH) \\ &\quad \times [\rho^*(T^l,H) \circ \rho^*(H,G)H^+(BG)]^{-1} \\ &= H^*(BT^l)^{W(H)}/H^+(BT^l)^{W(G)}, \end{aligned} \tag{4.4}$$

where $H^+(\text{——}) = H^*(\text{——}) - H^0(\text{——})$.

For the cohomology rings of T^l and BT^l one has

$$H^*(T^l) = \Lambda_R\,[x_1,\ldots,x_l], \quad x_i \in H^1(T^l), \tag{4.5}$$

and

$$H^*(BT^l) = R\,[t_1,\ldots,t_l], \quad t_i \in H^2(BT^l), \tag{4.6}$$

where $\Lambda_R[\ \ ]$ is the exterior algebra and $R[\ \ ]$ is the polynomial ring, both over the reals. The transgression τ defined[10] by

$$\begin{aligned} &H^{n-1}(T^l) \xrightarrow{\delta^*} H^n(ET^l, T^l) \xleftarrow{q^*} H^n(BT^l), \\ &\tau:\ \delta^{*-1}(\mathrm{Im}\,q^*) \to H^*(BT^l)/\ker q^*, \end{aligned} \tag{4.7}$$

gives an isomorphism,

$$\tau:\ H^1(T^l) \simeq H^2(BT^l), \text{ i.e., } \tau(x_i) = t_i \quad (1 \leqslant i \leqslant l). \tag{4.8}$$

Since $\mathfrak{h}^*$ is identified in a well-known way to $H^1(T^l)$,

$$H^2(BT^l) \simeq H^1(T^l) \simeq \mathfrak{h}^*. \tag{4.9}$$

Thus one may identify

$$H^*(BT^l)^{W(H)} \simeq I(H) \quad \text{and} \quad H^*(BT^l)^{W(G)} \simeq I(G), \tag{4.10}$$

obtaining

$$H^*(G/H) \simeq I(H)/I^+(G), \tag{4.11}$$

where $I^+(G)$ denotes a Weyl invariant polynomial ring without constant terms.

We shall explicitly evaluate the cohomology groups of the exceptional type G/H at lower orders relevant to the anomaly argument. The evaluation will be done by using Eq. (4.11) and the results of Sec. III.

(1) $\mathrm{E}_8/\mathrm{SO}(10) \times \mathrm{SU}(3) \times \mathrm{U}(1)$. We choose the subalgebra of the Lie algebra of E_8 as illustrated schematically in Fig. 1. To find $I(\mathrm{SO}(10) \times \mathrm{SU}(3) \times \mathrm{U}(1))$ in a manifestly invariant form under $W(\mathrm{SO}(10)) \times W(\mathrm{SU}(3)) \times W(\mathrm{U}(1))$, we introduce new parameters $\{c_1, e_i\ (i = 1\text{–}3, \sum_{i=1}^{3} e_i = 0), y_a\ (a = 1\text{–}5)\}$ instead of $\{x_i\ (i = 1\text{–}8)\}$ used in Sec. III such that

$$\begin{aligned} &(c_1, \alpha_j) = 0, \quad \text{except for } j = 3, \\ &\alpha_1 = e_1 - e_2, \quad \alpha_2 = e_2 - e_3, \quad \sum_{i=1}^{3} e_i = 0, \\ &\alpha_{a+3} = y_{a+1} - y_a \quad (a = 1\text{–}4), \end{aligned}$$

and

$$\alpha_8 = y_1 + y_2. \tag{4.12}$$

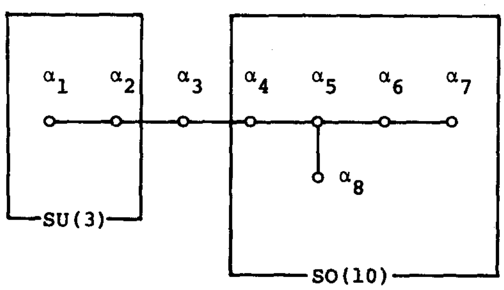


FIG. 1. Subalgebra of the Lie algebra of E_8.

It is shown in the later sections that c_1 is the coordinate corresponding to that of U(1) and is proportional to the first Chern class. (We normalize it identically.) Written in old parameters,

$$c_1 = -14\left(4\sum_{i=1}^{3} x_i + 3\sum_{j=4}^{8} x_j\right) = c_1(T(\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1))),$$

$$e_i = x_i - \frac{1}{3}\sum_{j=1}^{3} x_j,$$

and

$$y_a = -x_{a+3} + \frac{1}{2}\sum_{j=4}^{8} x_j. \tag{4.13}$$

They are transformed under σ_{α_i} $(i = 1,2,4\text{–}8)$ as follows:

c_1 is invariant under σ_{α_i},

e_i's are invariant under σ_{α_i} $(i = 4\text{–}8)$, which generate $W(\mathrm{SO}(10))$, $\sigma_{\alpha_i}: e_i \leftrightarrow e_{i+1}$ $(i = 1,2)$ (this transformation property coincides with that of x_i of A_2 in Sec. III),

y_a's are invariant under σ_{α_i} $(i = 1,2)$, which generate $W(\mathrm{SU}(3))$,

$$\sigma_{\alpha_{a+3}}: y_a \leftrightarrow y_{a+1} \ (a = 1\text{–}4),$$

and

$$\sigma_{\alpha_8}: \ y_1 \leftrightarrow -y_2, \tag{4.14}$$

coinciding with the transformation property of x_i of D_5 in Sec. III. Therefore we have

$$I(\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1)) = S(e_1,e_2,e_3)\times S(y_1^2, y_2^2,\ldots, y_5^2) \times \mathbb{R}\, y_1 y_2 y_3 y_4 y_5 \times \mathbb{R}\, c_1, \tag{4.15}$$

where $S(\cdots)$ is a symmetric polynomial defined in Sec. III. The explicit forms of $H^*(\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1))$ at lower orders are

$$H^2 = \mathbb{R}\, c_1,$$

$$H^4 = \mathbb{R}\, c_1^2 + \mathbb{R}\sum_i e_i^2 + \mathbb{R}\sum_a y_a^2, \quad \mathrm{mod}\, I_2^+(E_8), \tag{4.16}$$

$$H^6 = \mathbb{R}\, c_1^3 + \mathbb{R}\, c_1\sum_i e_i^2 + \mathbb{R}\, c_1\sum_a y_a^2 + \mathbb{R}\sum_i e_i^3, \quad \mathrm{mod}\, I_2^+(E_8),$$

where

$$I_2^+(E_8) = \mathbb{R}\left(\sum_{i=1}^{8} x_i^2 + \left(\sum_{i=1}^{8} x_i\right)^2\right) = \mathbb{R}\left(12\left(\frac{c_1}{168}\right)^2 + \sum_i e_i^2 + \sum_a y_a^2\right). \tag{4.17}$$

(2) $\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)$. The subalgebra of the Lie algebra of E_7 are chosen similarly (see Fig. 2). New parameters are also introduced,

$$c_1 = -5\left(5\sum_{i=2}^{4} x_i + 3\sum_{j=5}^{8} x_j\right) = c_1(T(\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1))),$$

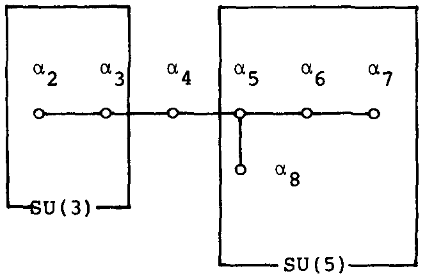


FIG. 2. Subalgebra of the Lie algebra of E_7.

$$e_i = x_{i+1} - \frac{1}{3}\sum_{j=2}^{4} x_j \quad \left(i = 1\text{–}3, \ \sum_i e_i = 0\right),$$

$$y_1 = \frac{3}{5}\sum_{j=5}^{8} x_j,$$

and

$$y_a = x_{a+3} - \frac{2}{5}\sum_{j=5}^{8} x_j, \quad \left(a = 2\text{–}5, \ \sum_{a=1}^{5} y_a = 0\right). \tag{4.18}$$

They are transformed under σ_{α_i} $(i = 2,3,5\text{–}8)$ as follows:

c_1 is invariant under σ_{α_i},

e_i's are invariant under σ_{α_i} $(i = 5\text{–}8)$, which generate $W(\mathrm{SU}(5))$,

$$\sigma_{\alpha_{i+1}}: \ e_i \leftrightarrow e_{i+1} \ (i = 1,2),$$

y_a's are invariant under σ_{α_i} $(i = 2,3)$, which generate $W(\mathrm{SU}(3))$,

$$\sigma_{\alpha_8}: y_1 \leftrightarrow y_2,$$

and

$$\sigma_{\alpha_{a+3}}: \ y_a \leftrightarrow y_{a+1} \ (a = 2\text{–}4). \tag{4.19}$$

The explicit forms of $H^*(\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1))$ at lower orders are similarly evaluated:

$$H^2 = \mathbb{R}\, c_1,$$

$$H^4 = \mathbb{R}\, c_1^2 + \mathbb{R}\sum_i e_i^2 + \mathbb{R}\sum_a y_a^2, \quad \mathrm{mod}\, I_2^+(E_7), \tag{4.20}$$

$$H^6 = \mathbb{R}\, c_1^3 + \mathbb{R}\sum_i e_i^3 + \mathbb{R}\sum_a y_a^3 + \mathbb{R}\, c_1\sum_i e_i^2 + \mathbb{R}\, c_1\sum_a y_a^2, \quad \mathrm{mod}\, I_2^+(E_7),$$

where

$$I_2^+(\mathrm{E}_7) = \mathbb{R}\left(2\sum_{i=2}^{8} x_i^2 + \left(\sum_{i=2}^{8} x_i\right)^2\right) = \mathbb{R}\left(c_1^2 + 750\sum_i e_i^2 + 750\sum_a y_a^2\right). \tag{4.21}$$

(3) $\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)$. We choose the subalgebra as in Fig. 3. New parameters are

$$c_1 = -4\left(4x_3 + \sum_{j=4}^{8} x_j\right) = c_1(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)))$$

and

$$y_a = -x_{a+3} + \frac{1}{2}\sum_{j=4}^{8} x_j \quad (a = 1\text{–}5). \tag{4.22}$$

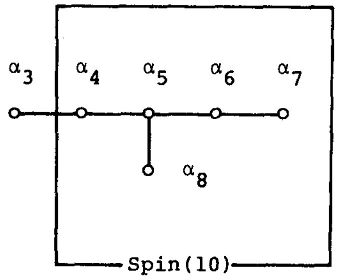


FIG. 3. Subalgebra of the Lie algebra of E_6.

Their transformation properties under σ_{α_i} $(i=4\text{–}8)$, which generate $W(\mathrm{Spin}(10))$, are the same as in the case of $E_8/SO(10)\times SU(3)\times U(1)$. Cohomology groups at lower orders are

$$H^2=\mathbf{R}\,c_1,$$

$$H^4=\mathbf{R}\,c_1^2+\mathbf{R}\sum_a y_a^2,\quad \mathrm{mod}\, I_2^+(E_6), \tag{4.23}$$

$$H^6=\mathbf{R}\,c_1^3+\mathbf{R}\,c_1\sum_a y_a^2,\quad \mathrm{mod}\, I_2^+(E_6),$$

where

$$I_2^+(\mathrm{E}_6)=\mathbf{R}\left(3\sum_{i=3}^{8}x_i^2+\left(\sum_{i=3}^{8}x_i\right)^2\right)$$

$$=\mathbf{R}\left(\frac{1}{2}\left(\frac{c_1}{4}\right)^2+6\sum_{a=1}^{5}y_a^2\right). \tag{4.24}$$

V. COMPLEX STRUCTURE OF G/H AND COMPLEMENTARY ROOTS

As has been mentioned in Sec. I, in four-dimensional supersymmetric nonlinear σ models with Weyl fermions, the homogeneous space G/H must be such that it admits a complex structure, and if it is Kählerian it is sufficient for the theory to be supersymmetric—this is as in our case. The possible complex structure in our case is unique (up to the complex conjugation), which we shall state in this section.

The subgroup H that we are considering is the centralizer of a torus $U(1)$ of a compact simple group G. Then by a Borel's theorem[14] G/H is homogeneous, Kählerian, and algebraic. The invariant complex structure J of such a manifold G/H can be introduced uniquely by identifying G/H with G^C/P, where G^C is the complexification of G and P is a closed complex subgroup of G^C such that $P\cap G=H$.[11,14] The Lie algebra of P is generated[14] by h_j $(j=1\text{–}l)\in\mathfrak{h}^C$, e_i $(i=1\text{–}m)$, and e_{-k}, such that a_i's relevant to e_i's are all positive roots and a_{-k}'s relevant to e_{-k}'s are (negative) simple roots satisfying $(a_k,b)=0$, where b is an element of the Weyl chamber and the centralizer of T_b (see the Appendix) is the Lie algebra of H. The roots of G^C complementary to P are defined as

$$\{\text{all roots of } G^C-\text{roots relevant to } P\}.$$

Then, a invariant complex structure on G/H corresponds to a root system ψ, which is also a set of weights of the linear isotropy unitary representation of H.[11] [See Eq. (6.1).]

Comments: As for the G-invariant Kähler metric, it is constructed by means of Maurer–Cartan forms. Let ω_k be the left-invariant one-forms on G^C whose restriction to $\mathscr{G}^C$ is annihilated by $\mathfrak{h}^C$ and $\omega_k(e_j)=\delta_{k,j}$. Then the Kähler metric[11,14] is

$$ds^2=\sum_{j=1}^{m}(b,a_j)\omega_j\,\omega_j^*, \tag{5.1}$$

where $\omega_j^*\equiv\omega_{-j}$ and ω_j^* is the complex conjugate of ω_j. In Refs. 1 and 3, one of the authors (Y. Y.) constructed the closed forms of the Kähler metrics for $E_6/\mathrm{Spin}(10)\times U(1)$ and $E_8/SO(10)\times SU(3)\times U(1)$ using the fact that these manifolds are embedded into Grassmann manifolds.

We present here the roots of G^C complementary to P in the case of the exceptional type G/H. These are used to calculate the Chern class in the next section.

(1) $E_6/\mathrm{Spin}(10)\times U(1)$,

$$b\in\mathbf{R}(4\alpha_3+5\alpha_4+6\alpha_5+4\alpha_6+2\alpha_7+3\alpha_8)$$
$$=\mathbf{R}\,(4x_3+x_4+x_5+x_6+x_7+x_8),$$

and

$$(\alpha_k,b)=0\quad\text{for } k=4\text{–}8. \tag{5.2}$$

Thus the Lie algebra of P is generated by $\mathfrak{h}^C_{E_6}$, e_i $(i=1\text{–}36)$ (a_i are all positive roots) and e_{-k} $(k=4\text{–}8)$. Using the new parameters we introduced in the previous section, the roots of E_6^C complementary to P are written as follows:

$$\frac{c_1}{16}+y_a+y_b-\frac{1}{2}\sum_{c=1}^{5}y_c\quad(1\leqslant a<b\leqslant 5),$$
$$\frac{c_1}{16}-y_a+\frac{1}{2}\sum_{c=1}^{5}y_c\quad(1\leqslant a\leqslant 5), \tag{5.3}$$
$$\frac{c_1}{16}-\frac{1}{2}\sum_{c=1}^{5}y_c.$$

(2) $E_7/SU(5)\times SU(3)\times U(1)$,

$$b\in\mathbf{R}\,(5\alpha_2+10\alpha_3+15\alpha_4+18\alpha_5+12\alpha_6+6\alpha_7+9\alpha_8)$$
$$=\mathbf{R}\,(5(x_2+x_3+x_4)+3(x_5+x_6+x_7+x_8)),$$

and

$$(\alpha_k,b)=0\quad\text{for } k=2,3,5\text{–}8. \tag{5.4}$$

The Lie algebra of P is generated by $\mathfrak{h}^C_{E_7}$, e_i $(i=1\text{–}63)$ (a_i are all positive roots) and e_{-k} $(k=2,3,5\text{–}8)$. The roots of E_7^C complementary to P are

$$c_1/75-e_i+y_a+y_b\quad(1\leqslant i\leqslant 3,\quad 1\leqslant a<b\leqslant 5),$$
$$2(c_1/75)+e_i-y_a\quad(1\leqslant i\leqslant 3,\quad 1\leqslant a\leqslant 5), \tag{5.5}$$
$$3(c_1/75)+y_a\quad(1\leqslant a\leqslant 5).$$

(3) $E_8/SO(10)\times SU(3)\times U(1)$,

$$b\in\mathbf{R}\,(4\alpha_1+8\alpha_2+12\alpha_3+15\alpha_4$$
$$+18\alpha_5+12\alpha_6+6\alpha_7+9\alpha_8)$$
$$=\mathbf{R}\,(4(x_1+x_2+x_3)+3(x_4+x_5+x_6+x_7+x_8))$$

and

$$(\alpha_k,b)=0\quad\text{for } k=1,2,4\text{–}8. \tag{5.6}$$

The Lie algebra of P is generated by $\mathfrak{h}^C_{E_8}$, e_i $(i=1\text{–}120)$ (a_i are all positive roots) and e_{-k} $(k=1,2,4\text{–}8)$. Complementary roots are

$$\frac{c_1}{168} - e_i + \frac{1}{2}\sum_{a=1}^{5} \epsilon_a y_a \quad (1<i<3),$$
$$2(c_1/168) - (e_i + e_j) \pm y_a \quad (1<i<j<3, \quad 1<a<5),$$
$$3\left(\frac{c_1}{168}\right) - \frac{1}{2}\sum_{a=1}^{5} \epsilon_a y_a,$$
$$4(c_1/168) - e_i \quad (1<i<3), \tag{5.7}$$
where $\epsilon_a = \pm 1$ and $\Pi_{a=1}^{5} \epsilon_a = -1$.

VI. CHERN CHARACTER OF TG/H

We calculate the Chern character using the method investigated by Borel and Hirzebruch.[11] Left (right) translation by $g \in G$ induces a homeomorphism of G/H. If $h \in H$, it leaves $0 = \pi(e) \in G/H$ invariant and induces an automorphism $\bar{h}$: $(TG/H)_0 \to (TG/H)_0$. The homomorphism $\iota: h \to \bar{h}$ is called the isotropy representation. The complex structure on G/H defined in the previous section gives rise to a linear isotropy unitary representation of H,
$$\iota_C \colon H \to \bar{H} \subset U(n), \tag{6.1}$$
with $n =$ the complex dimension of G/H. This homomorphism ι_C induces the bundle map from $\eta = (G, \pi, G/H, H)$ to a principal $\mathrm{U}(n)$ bundle over G/H called the ι_C extension of η, which we denote by η_C. A theorem of Borel and Hirzebruch[11] asserts that η_C is equal to the principal bundle associated with the tangent bundle $T(G/H)$. Thus one has
$$c(T(G/H)) = c(\text{the principal bundle associated with } T(G/H)) = c(\eta_C). \tag{6.2}$$
We now consider the following commutative diagram of mappings:

$$\begin{tikzcd}
G \arrow[r] \arrow[d] & EH \arrow[r] \arrow[d] & E\mathrm{U}(n) \arrow[d] & G \times_H \mathrm{U}(n) \arrow[l] \arrow[dd, "\hat{\rho}"] \\
G/T^l \arrow[r, "\phi"] \arrow[d, "\rho"] & EH/T^l = BT^l \arrow[r, "\phi'"] \arrow[d, "\rho(T^l, H)"] & E\mathrm{U}(n)/T^n = BT^n \arrow[d, "\rho(T^n, U(n))"] & \\
G/H \arrow[r, "i_2"'] & BH \arrow[r, "\sigma(\iota_C)"'] \arrow[d, "\rho(H, G)"] & B\mathrm{U}(n) & G/H \arrow[l, "\sigma'"'] \\
 & BG & &
\end{tikzcd}, \tag{6.3}$$

where $(G \times_H \mathrm{U}(n), \hat{\rho}, G/H, \mathrm{U}(n)) = \eta_C$, $\sigma(\iota_C)$ is the map induced from the homomorphism ι_C called the characteristic map for ι_C extension of η and $\sigma' = \sigma(\iota_C) \circ i_2$. The Chern class of η_C is given by
$$c(\eta_C) = \sigma'^* c \in H^*(G/H), \tag{6.4}$$
where $c \in H^*(BU(n))$ is the Chern class of the universal $U(n)$ bundle, which is defined by the equation,
$$\rho^*(T^n, U(n))c = \prod_{j=1}^{n} (1 + \tau'(x_j')) \tag{6.5}$$
with τ' the transgression of $H^1(T^n)$ to $H^2(BT^n)$ and x_j' the standard coordinate of T^n identified with the element of $H^1(T^n)$. By the equation $\sigma' = \sigma(\iota_C) \circ i_2$ and the commutativity of the diagram,
$$\begin{aligned} \rho^* c(\eta_C) &= \rho^* \circ i_2^* \circ \sigma(\iota_C)^* c \\ &= \phi^* \circ \phi'^* \circ \rho^*(T^n, U(n))c \\ &= \phi^* \circ \phi'^* \prod_{j=1}^{n} (1 + \tau'(x_j')), \end{aligned} \tag{6.6}$$
where ρ^* is applicable since $c(\eta_C)$ belongs to $H^*(G/H)$. Because ϕ'^* commutes with τ',
$$\phi'^* \prod_{j=1}^{n} (1 + \tau'(x_j')) = \prod_{j=1}^{n} (1 + \tau(\omega_j)), \tag{6.7}$$
where τ is the transgression of $H^1(T^l)$ to $H^2(BT^l)$ and $\omega_j(x) = \phi'^*(x_j')$ are the weights of ι_C by the definition[11] and are also the roots of G^C complementary to P. We thus obtain
$$\begin{aligned} \rho^* c(\eta_C) &= \phi'^* \prod_{j=1}^{n} (1 + \tau(\omega_j)) \\ &= \prod_{j=1}^{n} (1 + \tilde{\tau}(\omega_j)), \end{aligned}$$
$$\mathrm{mod}\, I^+(G) = \rho^*(T, H) \circ \rho^*(H, G) H^*(BG), \tag{6.8}$$
with $\tilde{\tau}$ the transgression of $H^1(T^l)$ to $H^2(G/T^l)$. It is known[10,15] that when H is the centralizer of $\mathrm{U}(1)$, the spaces G/T^l and G/H have no torsion and therefore the homomorphism ρ^* induced by the projection ρ is injective. We may therefore identify
$$\begin{aligned} c(\eta_C) &\simeq \prod_{j=1}^{n} (1 + \tilde{\tau}(\omega_j)), \quad \mathrm{mod}\, I^+(G), \\ &\simeq \prod_{j=1}^{n} (1 + \omega_j), \quad \mathrm{mod}\, I^+(G), \end{aligned} \tag{6.9}$$
because the transgression $\tilde{\tau}$ in our case is isomorphic.

Using the results of Sec. V for ω_j and that of Sec. III for $I^+(G)$, we can explicitly evaluate Eq. (6.9). The results are as follows.

(1) $\mathrm{E}_8/\mathrm{SO}(10) \times \mathrm{SU}(3) \times \mathrm{U}(1)$,
$$c(T(E_8/\mathrm{SO}(10) \times \mathrm{SU}(3) \times \mathrm{U}(1))) = \prod_{\text{complementary roots}} \left(1 + \frac{c_1}{168} - e_i + \frac{1}{2}\sum_a \epsilon_a y_a\right) \times \left(1 + 2\left(\frac{c_1}{168}\right) - (e_j + e_k) \pm y_b\right)$$

$$\times\left(1+3\left(\frac{c_1}{168}\right)-\frac{1}{2}\sum_c \epsilon_c y_c\right)$$

$$\times\left(1+4\left(\frac{c_1}{168}\right)-e_l\right),$$

$$\operatorname{mod}\{I_i^+ \quad (i=2,8,12,14,18,20,24,30)\}. \tag{6.10}$$

Especially

$$c_1=\sum \text{(complementary roots)}$$

$$=-14\left(4\sum_{i=1}^{3}x_i+3\sum_{j=4}^{8}x_j\right),$$

$$c_2=139\,32\left(\frac{c_1}{168}\right)^2-11\sum_{a=1}^{5}y_a^2-\frac{27}{2}\sum_{i=1}^{3}e_i^2,$$

$$\operatorname{mod} I_2^+ = 12\left(\frac{c_1}{168}\right)^2+\sum_i e_i^2+\sum_a y_a^2,$$

$$c_3=760\,336\left(\frac{c_1}{168}\right)^3-1812\left(\frac{c_1}{168}\right)\sum_a y_a^2-\frac{7}{3}$$

$$\times\sum_i e_i^3-2228\left(\frac{c_1}{168}\right)\sum_i e_i^2, \quad \operatorname{mod} I_2^+. \tag{6.11}$$

Therefore we obtain

$$\mathrm{ch}_3(T(\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1)))$$

$$=1/6(c_1^3+3c_3-3c_1c_2)$$

$$=18\left(\frac{c_1}{168}\right)\sum_a y_a^2+20\left(\frac{c_1}{168}\right)\sum_i e_i^2+\left(\frac{7}{6}\right)$$

$$\times\sum_i e_i^3+152\left(\frac{c_1}{168}\right)^3, \quad \operatorname{mod} I_2^+. \tag{6.12}$$

(2) $\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)$,

$$c(T(E_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)))$$

$$=\prod_{\text{complementary roots}}\left(1+\frac{c_1}{75}-e_i+y_a+y_b\right)$$

$$\times\left(1+2\left(\frac{c_1}{75}\right)+e_j-y_c\right)\left(1+3\left(\frac{c_1}{75}\right)+y_d\right),$$

$$\operatorname{mod}\{I_i^+ \ (i=2,6,8,10,12,14,18)\}, \tag{6.13}$$

$$c_1=\sum \text{(complementary roots)}$$

$$=-5\left(5\sum_{i=2}^{4}x_i+3\sum_{j=5}^{8}x_j\right),$$

$$c_2=2745\left(\frac{c_1}{75}\right)^2-\left(\frac{15}{2}\right)\sum_i e_i^2-\left(\frac{13}{2}\right)\sum_a y_a^2,$$

$$\operatorname{mod} I_2^+ = 15\left(\frac{c_1}{75}\right)^2+2\sum_i e_i^2+2\sum_a y_a^2,$$

and

$$c_3=653\,45\left(\frac{c_1}{75}\right)^3+\left(\frac{1}{3}\right)\sum_a y_a^3-\left(\frac{5}{3}\right)$$

$$\times\sum_i e_i^3-\left(\frac{1085}{2}\right)\left(\frac{c_1}{75}\right)\sum_i e_i^2-\left(\frac{939}{2}\right)\left(\frac{c_1}{75}\right)$$

$$\times\sum_a y_a^2, \quad \operatorname{mod} I_2^+, \tag{6.14}$$

$$\mathrm{ch}_3(T(\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)))$$

$$=9\left(\frac{c_1}{75}\right)\sum_a y_a^2-\left(\frac{5}{6}\right)\sum_i e_i^3+\left(\frac{1}{6}\right)$$

$$\times\sum_a y_a^3+10\left(\frac{c_1}{75}\right)\sum_i e_i^2$$

$$+95\left(\frac{95}{2}\right)\left(\frac{c_1}{75}\right)^3, \quad \operatorname{mod} I_2^+. \tag{6.15}$$

(3) $\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)$

$$c(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)))$$

$$=\prod_{\text{complementary roots}}\left(1+\frac{c_1}{16}+y_a+y_b-\left(\frac{1}{2}\right)\sum_{c=1}^{5}y_c\right)$$

$$\times\left(1+\frac{c_1}{16}-y_d+\left(\frac{1}{2}\right)\sum_{e=1}^{5}y_e\right)$$

$$\times\left(1+\frac{c_1}{16}-\left(\frac{1}{2}\right)\sum_{f=1}^{5}y_f\right)$$

$$\operatorname{mod}\{I_i^+ \ (i=2,5,6,8,9,12)\}, \tag{6.16}$$

$$c_1=\sum \text{(complementary roots)}=-4\left(4x_3+\sum_{l=4}^{8}x_l\right),$$

$$c_2=(\tfrac{23}{3})(c_1/4)^2,$$

$$c_3=(\tfrac{28}{3})(c_1/4)^3,$$

and

$$\mathrm{ch}_3(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)))=0. \tag{6.17}$$

Here, we eliminated $\Sigma_{a=1}^5 y_a^2$ by using $\operatorname{mod} I_2^+$;

$$I_2^+=\mathbf{R}\left(\left(\frac{1}{2}\right)\left(\frac{c_1}{4}\right)^2+6\sum_a y_a^2\right). \tag{6.18}$$

VII. CONCLUSION

We now evaluate the anomalies of the exceptional-type nonlinear σ models; we show that

$$\int \hat{\varphi}^*\,\mathrm{ch}_3(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)))=0$$

for any $\hat{\varphi}: s^2\times s^4\to \mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)$ and $\int \hat{\varphi}^*\,\mathrm{ch}_3(TG/H)\neq 0$ for some $\hat{\varphi}: s^2\times s^4\to G/H$, where $G/H=\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)$ or $\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1)$.

(1) $\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)$. As seen in Eq. (6.17), the third Chern character vanishes in this case, and thus

$$\int_{s^2\times s^4}\hat{\varphi}^*\,\mathrm{ch}_3(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)))=0. \tag{7.1}$$

We thus conclude that the theory is anomaly-free. Note that this result depends on the dimension of the space-time; we are considering the theory in the four-dimensional space-time. In the case of two-dimensional space-time with Weyl fermions, we must consider

$$\int_{s^2\times s^2}\hat{\varphi}^*\,\mathrm{ch}_2(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1))), \tag{7.2}$$

where $\mathrm{ch}_2(T(\mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)))=\frac{1}{3}(c_1/4)^2$ and $\hat{\varphi}: s^2\times s^2\to \mathrm{E}_6/\mathrm{Spin}(10)\times\mathrm{U}(1)$. Because $\hat{\varphi}^*c_1$ belongs to $H^2(s^2\times s^2) = H^2(s^2)\otimes H^0(s^2)\oplus H^0(s^2)\otimes H^2(s^2)$, $(\hat{\varphi}^*c_1)^2\in H^2(s^2)\otimes H^2(s^2)$ is nonzero for some $\hat{\varphi}$ (Ref. 16) and thus the theory is anomalous. However, this is not the physically interesting case, because in two dimensions $N = 1$ supersymmetry does not require a complex structure nor does it usually involve the Weyl fermion, but it involves Majorana ones. This situation is also unaltered in the following cases.

(2) $E_7/SU(5)\times SU(3)\times U(1)$. Recall that the ith elementary symmetric polynomials $S_i(y_1,..., y_5)$ and $S_i(e_1,..., e_3)$ correspond to the ith Chern classes of the principal SU(5) [and SU(3)] bundles over $E_7/SU(5)\times SU(3)\times U(1)$, which we simply denote as

$$\frac{1}{2}\sum_{a=1}^{5} y_a^2 = c_2(\mathrm{SU}(5)), \quad \frac{1}{2}\sum_{i=1}^{3} e_i^2 = c_2(\mathrm{SU}(3)),$$

$$\frac{1}{3}\sum_{a=1}^{5} y_a^3 = c_3(\mathrm{SU}(5)),$$

and

$$\frac{1}{3}\sum_{i=1}^{3} e_i^3 = c_3(\mathrm{SU}(3)). \tag{7.3}$$

Therefore

$$\mathrm{ch}_3(T(\mathrm{E}_7/\mathrm{SU}(5)\times \mathrm{SU}(3)\times \mathrm{U}(1)))$$
$$= 18(c_1/75)c_2(\mathrm{SU}(5)) - \tfrac{5}{2}c_3(\mathrm{SU}(5)) + \tfrac{1}{2}c_3(\mathrm{SU}(3))$$
$$+ 20(c_1/75)c_2(\mathrm{SU}(3)) + \tfrac{25}{2}(c_1/75)^3, \quad \mathrm{mod}\, I_2^+. \tag{7.4}$$

We here consider a map $\hat{\varphi} \equiv \hat{\varphi}_2 \circ \hat{\varphi}_1 : s^2\times s^4 \to E_7/SU(5)\times SU(3)\times U(1)$ such that $\hat{\varphi}_1 : s^2\times s^4 \to s^6$ and $\hat{\varphi}_2 : s^6 \to E_7/SU(5)\times SU(3)\times U(1)$ and they belong to nonzero sectors of the homotopy classes[17]

$$[s^2\times s^4, s^6] \simeq Z,$$
$$[s^6, \mathrm{E}_7/\mathrm{SU}(5)\times \mathrm{SU}(3)\times \mathrm{U}(1)] \tag{7.5}$$
$$\simeq \pi_5(\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)) \simeq Z \oplus Z.$$

Then $\hat{\varphi}_2^* c_1 (\in H^2(s^6)) = 0$, $\hat{\varphi}_2^* c_2$ (SU(3)), and $\hat{\varphi}_2^* c_2$ (SU(5)) $(\in H^4(s^6)) = 0$, but $\hat{\varphi}_2^* c_3$ (SU(3)) and $\hat{\varphi}_2^* c_3$ (SU(5)) are nonzero elements of $H^6(s^6)$. By the pullback $\hat{\varphi}_1^*$ they still remain to be nonzero and belong to $H^6(s^2\times s^4)$. We thus obtain

$$\hat{\varphi}^* \mathrm{ch}_3(T(\mathrm{E}_7/\mathrm{SU}(5)\times\mathrm{SU}(3)\times\mathrm{U}(1)))$$
$$= \hat{\varphi}_1^* \circ \hat{\varphi}_2^*(-\tfrac{5}{2}c_3(\mathrm{SU}(3)) + \tfrac{1}{2}c_3(\mathrm{SU}(5))) \neq 0, \tag{7.6}$$

leading to the conclusion that the theory is anomalous.

This result is unchanged in the two-dimensional space-time.

(3) $E_8/SO(10)\times SU(3)\times U(1)$. The relevant elementary symmetric polynomials correspond to the characteristic classes as follows:

$$\frac{1}{2}\sum_{i=1}^{3} e_i^2 = c_2(\mathrm{SU}(3)), \quad \frac{1}{3}\sum_{i=1}^{3} e_i^3 = c_3(\mathrm{SU}(3)),$$

and

$$\sum_{a=1}^{5} y_a^2 = p_1(\mathrm{SO}(10)), \tag{7.7}$$

where $p_1(SO(10))$ is the first-Pontryagin class of the principal SO(10) bundle over $E_8/SO(10)\times SU(3)\times U(1)$. The third Chern character is then written by

$$\mathrm{ch}_3(T(\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1)))$$
$$= 18(c_1/168)\,p_1(\mathrm{SO}(10)) + \tfrac{7}{2}c_3(\mathrm{SU}(3))$$
$$+ 40(c_1/168)c_2(\mathrm{SU}(3)) + 152(c_1/168)^3,$$
$$\mathrm{mod}\, I_2^+, \tag{7.8}$$

and the pullback by $\hat{\varphi}^*$ is given by

$$\hat{\varphi}^* \mathrm{ch}_3(T(\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1)))$$
$$= \tfrac{7}{2}\hat{\varphi}^* c_3(\mathrm{SU}(3)) \in H^6(s^2\times s^4),$$

where $\hat{\varphi} = \hat{\varphi}_2 \circ \hat{\varphi}_1$ is the similar mapping to the case of $E_7/SU(5)\times SU(3)\times U(1)$; other terms are trivialized by $\hat{\varphi}_2^*$.

Thus there exist $\hat{\varphi}$ such that

$$\hat{\varphi}^* \mathrm{ch}_3(T(\mathrm{E}_8/\mathrm{SO}(10)\times\mathrm{SU}(3)\times\mathrm{U}(1))) \neq 0, \tag{7.9}$$

leading to the fact that the theory is anomalous. It is also anomalous in the case of two-dimensional space-time.

These are the conclusions. Our method of evaluating anomalies is also applicable[18] to other supersymmetric nonlinear σ models based on, in general, Kählerian coset spaces of the type G/H.

Note that our results are the same as those of non-Abelian anomalies in H-gauge theories. Whether this correspondence is accidental or indicating that these anomalies are related deeply is yet unknown to the authors.

Comment: Alvarez-Gaume and Ginsparg[19] have shown that $G/H = E_6/\mathrm{Spin}(10)\times U(1)$, $E_7/SU(5)\times SU(3)\times U(1)$, and $E_8/SO(10)\times SU(3)\times U(1)$ models are all "anomalous" in four-dimensional space-time because the 't Hooft anomaly matching condition cannot be satisfied in these models, i.e., the anomaly in the U(1) part of H is at least nonvanishing and can never be matched by any representation of E_l ($l = 6$–8), which have only an anomaly-free representation in four-dimensional space-time. However, our conclusion, especially that $E_6/\mathrm{Spin}(10)\times U(1)$ is anomaly-free, does not contradict with their result. Let us clarify this point.

As seen in Ref. 20, their statement is equivalent to that $\mathrm{Tr}\, R^3 \neq 0$ [R is the curvature form on $\hat{\varphi}^* T(G/H)$], but this does not necessarily mean the nonvanishing of the integration of $\mathrm{Tr}\, R^3$ over $S^2\times S^4$; the Chern character is evaluated at modulo exact form.

ACKNOWLEDGMENTS

We thank Professor H. Sato (Department of Mathematics, Tohoku University) for helpful discussions and informing us about some relevant mathematical papers.

APPENDIX: ROOTS OF E_8^C

In this Appendix we give roots of E_8^C and the definition of a scalar product on $\mathfrak{h}^*$, the real vector space spanned by the simple roots. The Lie algebra of E_8^C, $\mathscr{G}^C$, is constructed following Ref. 13.

Let $\mathfrak{l}$ be the Lie algebra of SL(9,C), V the complex valued antisymmetric tensor of the rank 3 in nine-dimensions, and V^* be its dual space. Thus $\dim_C \mathfrak{l} = 81 - 1 = 80$, $\dim_C V = \dim_C V^* = {}_9C_3 = 84$, and $\dim_C(\mathfrak{l} \oplus V \oplus V^*)$

$= 284$, which is equal to the dimension of the Lie algebra of E_8^C.

Now we introduce the commutation relation in the complex vector space $\mathfrak{l} \oplus V \oplus V^*$. Let

$$E_{ij} = \begin{array}{cl} \begin{bmatrix} \vdots & 1 & 0 \\ \vdots & 0 & \vdots \\ 0 & \vdots & 0 \end{bmatrix} & i\text{th row} \\ \quad j\text{th column} & \end{array}$$

e_{ijk}, and e^{ijk} be the base of $\mathfrak{l}$, V and V^*, respectively. Here each index runs from 1 to 9 and $e^{ijk}(e_{lmn}) = 3!\, \epsilon^{ijk}{}_{lmn'}$

$$\epsilon^{ijk}{}_{lmn} = \det \begin{bmatrix} \delta^i_l & \delta^i_m & \delta^i_n \\ \delta^j_l & \delta^j_m & \delta^j_n \\ \delta^k_l & \delta^k_m & \delta^k_n \end{bmatrix}. \tag{A1}$$

For $X = \Sigma X_{ij} E_{ij}$ with $\Sigma_i X_{ii} = 0$, $v = \frac{1}{3!} \Sigma v_{ijk} e_{ijk}$, and $v^* = \frac{1}{3!} \Sigma v^{ijk} e^{ijk}$, commutation relations

$$\begin{aligned} &[X, v]: \quad \mathfrak{l} \otimes V \to V, \\ &[X, v^*]: \quad \mathfrak{l} \otimes V^* \to V^*, \\ &[v, v']: \quad V \otimes V \to V^*, \\ &[v^*, v^{*\prime}]: \quad V^* \otimes V^* \to V, \\ &[v, v^*]: \quad V \otimes V^* \to \mathfrak{l}, \end{aligned} \tag{A2}$$

are defined as follows:

$$\begin{aligned} &[X, v]_{ijk} = X_{il}\, v_{ljk} + X_{jl}\, v_{ilk} + X_{kl}\, v_{ijl}, \\ &[X, v^*]^{ijk} = -X_{li}\, v^{ljk} - X_{lj}\, v^{ilk} - X_{lk}\, v^{ijl}, \\ &[v, v']^{opq} = \epsilon^{ijklmnopq}\, v_{ijk}\, v'_{lmn}, \\ &[v^*, v^{*\prime}]_{opq} = 1/(18 \times 6^3) \epsilon_{ijklmnopq}\, v^{ijk}\, v'^{lmn}, \\ &[v, v^*]_{ij} = \tfrac{1}{6}(v^{jkl}\, v_{ikl} - \tfrac{1}{9} \delta_{ij}\, v^{lmn}\, v_{lmn}. \end{aligned} \tag{A3}$$

The Cartan subalgebra $\mathfrak{h}^C$ of $\mathscr{G}^C$ is equal to that of $\mathfrak{l}$ and each $h \in \mathfrak{h}^C$ is written by using the coordinates of maximal torus:

$$\begin{aligned} h &= \mathrm{diag}(2\pi\sqrt{-1}\, x_1, \ldots, 2\pi\sqrt{-1}\, x_9), \\ &\sum_{i=1}^{9} x_i = 0, \quad x_i \in \mathbb{C}. \end{aligned} \tag{A4}$$

($\mathfrak{h} = \mathfrak{h}^C \cap \mathscr{G}$ is simply given by replacing $x_i \in \mathbb{C}$ with $x_i \in \mathbf{R}$.) Defining the mapping λ_i: $\mathfrak{h}^C \to \mathbb{C}$ by $\lambda_i(h) = x_i$ (the λ_i's are often written by x_i in this recognition),

$$h = \sum_{i=1}^{9} 2\pi\sqrt{-1}\, \lambda_i(h)\, E_{ii}, \quad \sum_{i=1}^{9} \lambda_i = 0. \tag{A5}$$

Note that the λ_i's are real valued on $\mathfrak{h}$ ($\lambda_i \in \mathfrak{h}^*$). Since

$$\begin{aligned} &[h, E_{ij}] = 2\pi\sqrt{-1}\, (\lambda_i - \lambda_j)(h)\, E_{ij} \quad (i \neq j), \\ &[h, e_{ijk}] = 2\pi\sqrt{-1}\, (\lambda_i + \lambda_j + \lambda_k)(h)\, e_{ijk}, \end{aligned} \tag{A6}$$

and

$$[h, e^{ijk}] = -2\pi\sqrt{-1}\, (\lambda_i + \lambda_j + \lambda_k)(h)\, e^{ijk},$$

all roots of $\mathscr{G}^C$ are

$$\begin{aligned} &\pm(\lambda_i - \lambda_j) \quad (1 \leqslant i < j \leqslant 9), \\ &\pm(\lambda_i + \lambda_j + \lambda_k) \quad (1 \leqslant i < j < k \leqslant 9), \end{aligned} \tag{A7}$$

and simple roots are

$$\begin{aligned} &\lambda_i - \lambda_{i+1} \quad (1 \leqslant i \leqslant 7), \\ &\lambda_6 + \lambda_7 + \lambda_8, \end{aligned} \tag{A8}$$

where we omitted $2\pi\sqrt{-1}$.

The scalar product (,) on $\mathfrak{h}^*$ is derived from the Killing form κ: $\mathscr{G}^C \otimes \mathscr{G}^C \to \mathbb{C}$. Here κ is defined by

$$\kappa(A, B) = \mathrm{Tr}((\mathrm{ad}\, A)(\mathrm{ad}\, B)) \quad \text{for each } A, B \in \mathscr{G}^C. \tag{A9}$$

This formula is reduced to

$$\begin{aligned} &\kappa(e^{ijk}, e_{lmn}) = -20\, \epsilon^{ijk}{}_{lmn}, \\ &\kappa(E_{ij}, E_{kl}) = 60 \delta_{il}\, \delta_{jk}, \end{aligned} \tag{A10}$$

otherwise zero.

For each $a \in \mathfrak{h}^*$, there exists a unique $T_a \in \mathfrak{h}$ satisfying the condition

$$a(h) = \kappa(T_a, h), \quad \text{for any } h \in \mathfrak{h}. \tag{A11}$$

Then the scalar product (a, b) is defined by

$$(a, b) = \kappa(T_a, T_b), \quad a, b \in \mathfrak{h}^*. \tag{A12}$$

For example,

$$\begin{aligned} &2\pi\sqrt{-1}\, T_{\lambda_i - \lambda_j} = \tfrac{1}{60}(E_{ii} - E_{jj}) \quad (i \neq j), \\ &(\lambda_i - \lambda_j, \lambda_i - \lambda_j) = (1/2\pi\sqrt{-1})^2\, \tfrac{1}{30}, \end{aligned}$$

and

$$\begin{aligned} &2\pi\sqrt{-1}\, T_{\lambda_i + \lambda_j + \lambda_k} = \tfrac{1}{60}(E_{ii} + E_{jj} + E_{kk} - \tfrac{1}{3} 1_9), \\ &(\lambda_i + \lambda_j + \lambda_k, \lambda_i + \lambda_j + \lambda_k) = (1/2\pi\sqrt{-1})^2\, \tfrac{1}{30}. \end{aligned} \tag{A13}$$

where 1_9 is the unit matrix of 9×9.

[1] As for the E_6 type, see Y. Yasui, Prog. Theor. Phys. **72**, 877 (1984); Y. Achiman, S. Aoyama, and J. W. van Holten, Preprint TH. 3814-CERN; WU B 84-29.

[2] As for the E_7 type, see T. Kugo and T. Yanagida, Phys. Lett. B **134**, 313 (1984).

[3] As for the E_8 type, see S. Irié and Y. Yasui, Z. Phys. C **29**, 123 (1985); T. Yanagida (unpublished); C. L. Ong, Phys. Rev. D **27**, 3044 (1983); **31**, 3271 (1985).

[4] W. Buchmüller, R. D. Peccei, and T. Yanagida, Phys. Lett. B **124**, 67 (1983); Nucl. Phys. B **227**, 503 (1983); R. Barbieri, A. Masiero, and G. Veneziano, Phys. Lett. B **128**, 179 (1983); O. W. Greenberg, R. N. Mohapatra, and M. Yasue, Phys. Rev. Lett. **51**, 1737 (1983).

[5] Three usual families and one mirror family, see Ref. 3.

[6] G. Moore and P. Nelson, Phys. Rev. Lett. **53**, 1519 (1984); Commun. Math. Phys. **100**, 83 (1985); A. Manohar, G. Moore, and P. Nelson, Phys. Lett. B **152**, 68 (1985); E. Cohen and C. Goméz, Nucl. Phys. B **254**, 235 (1985). As a work from a slightly different point of view, see P. Tataru-Mihai, Zentrm für Theoretische Physik preprint, March, 1985.

[7] B. Zumino, Phys. Lett. B **87**, 203 (1979).

[8] M. Atiyah and I. Singer, Ann. Math. **87**, 546 (1968).

[9] L. Faddeev, Phys. Lett. B **145**, 81 (1984); B. Grossman, *ibid.* **152**, 93 (1985); Y. S. Wu and A. Zee, *ibid.* **152**, 98 (1985); B. Zumino, Santa Barbara preprint NSF-ITP-84-150; R. Jackiw, MIT Preprint CTP-1209.

[10] A. Borel, Ann. Math. **57**, 115 (1953).

[11] A. Borel and F. Hirzebruch, Am. J. Math. **80**, 458 (1958).

[12] A. Borel and C. Chevally, Mem. Am. Math. Soc. **14**, 1 (1955).

[13] G. Takeuchi, *Lie algebra and Elementary Particle Physics* (Shyokabo, Tokyo, Japan, 1983), this text is written in Japanese, so we present some necessary content in the Appendix.

[14] A. Borel, Proc. Natl. Acad. Sci. USA **40**, 1147 (1954).

[15] R. Bott, Bull. Soc. Math. France **84**, 251 (1956).

[16] We can choose $\hat{\varphi}$, for example, to belong to a nonzero sector of $\Pi_2(G/H) \oplus \Pi_2(G/H) = \Pi_1(H) \oplus \Pi_1(H) = Z \oplus Z$, where $G = E_6$ and $H = \mathrm{Spin}(10) \times \mathrm{U}(1)$.

[17]These formulas are obtained from the exact sequences:

$$0 = [s^3 V s^5, s^6] \to [s^6, s^6] \to [s^2 \times s^4, s^6] \to [s^2 V s^4, s^6]$$
$$= 0 \text{ (Puppe exact sequence)}$$

and

$$0 = \Pi_6(G) \to \Pi_6(G/H) \to \Pi_5(H) \to \Pi_5(G) = 0,$$

where $G = E_7$ (or E_8) and $H = SU(5) \times SU(3) \times U(1)$ [or $SO(10) \times SU(3) \times U(1)$].

[18]T. Yanagida and Y. Yasui, Nucl. Phys. B **269**, 575 (1986); Y. Yasui, Prog. Theor. Phys. **76**, 238 (1986).
[19]L. Alvarez-Gaume and P. Ginsparg, Nucl. Phys. B **262**, 439 (1985).
[20]J. Bagger, D. Nemeschansky, and S. Yankielowicz, Nucl. Phys. B **262**, 478 (1985).

Gauge theory of a group of diffeomorphisms. II. The conformal and de Sitter groups

Eric A. Lord
Department of Applied Mathematics, Indian Institute of Science, Bangalore 560 012, India

(Received 12 March 1986; accepted for publication 14 August 1986)

The extension of Hehl's Poincaré gauge theory to more general groups that include space-time diffeomorphisms is worked out for two particular examples, one corresponding to the action of the conformal group on Minkowski space, and the other to the action of the de Sitter group on de Sitter space, and the effect of these groups on physical fields.

I. INTRODUCTION

In a recent work[1] (which we shall refer to as I) a scheme was developed for gauging a group that contains a group of space-time diffeomorphisms as well as (possibly), internal symmetry groups.

Let G be an $(N+M)$-parameter Lie group possessing an N-parameter subgroup H. Introduce, on an M-dimensional base manifold, a connection Γ_i associated with the group G, considered as a Yang–Mills group, and a set of physical fields ψ belonging to a linear representation of H. Under the simultaneous action of an infinitesimal diffeomorphism $x^i \to x^i - \xi^i$ on the base space, and an infinitesimal (local) action of H, we have

$$\delta\Gamma_i = \xi^j \partial_j \Gamma_i + \Gamma_j \partial_i \xi^j + \partial_i \bar{\epsilon} + [\bar{\epsilon}, \Gamma_i], \tag{1.1}$$

$$\delta\psi = \xi^j \partial_j \psi + \bar{\epsilon}\psi, \tag{1.2}$$

where $\bar{\epsilon}$ is an infinitesimal element of the Lie algebra of H, dependent on position on the base manifold. Of course, in (1.2) the representation of H provided by ψ is implied.

When the curvature

$$G_{ij} = \partial_i \Gamma_j - \partial_j \Gamma_i - [\Gamma_i, \Gamma_j] \tag{1.3}$$

vanishes, those transformations (1.1) and (1.2) that leave invariant a particular solution $\Gamma_i(x)$ of (1.3) constitute an $(N+M)$-parameter group of diffeomorphisms on the base space, isomorphic to G. The finite-dimensional linear representation of H corresponding to the action of H on ψ is thereby extended to the action on ψ of a group G of diffeomorphisms.

The purpose of the present work is to illustrate this idea by two particularly interesting special cases.

When $G = \mathrm{SO}(4,2)$, we obtain the action of the conformal group on Minkowski space together with the appropriate transformation laws for physical fields under the action of the conformal group.[2] When $G = \mathrm{SO}(4,1)$ we obtain the action of the de Sitter group on de Sitter space-time together with the appropriate transformation laws for physical fields. Equations (1.1) and (1.2) in this latter case give rise to the basic transformation laws of Poincaré gauge theory, under the Wigner–Inönü contraction of the de Sitter group to the Poincaré group.

II. THE CONFORMAL GAUGE THEORY

The commutation relations for the generators of SO(4,2) can be displayed in the following form:

$$\begin{aligned}
&[\pi_\alpha, \pi_\beta] = 0,\\
&[\pi_\alpha, S_{\beta\gamma}] = \eta_{\alpha\beta}\pi_\gamma - \eta_{\alpha\gamma}\pi_\beta, \quad [\pi_\alpha, \Delta] = \pi_\alpha,\\
&[\pi_\alpha, \kappa_\beta] = 2(\eta_{\alpha\beta}\Delta - S_{\alpha\beta}),\\
&[S_{\alpha\beta}, S_{\gamma\delta}] = \eta_{\beta\gamma}S_{\alpha\delta} - \eta_{\alpha\gamma}S_{\beta\delta} + \eta_{\alpha\delta}S_{\beta\gamma} - \eta_{\beta\delta}S_{\alpha\gamma},\\
&[S_{\alpha\beta}, \Delta] = 0, \quad [S_{\alpha\beta}, \kappa_\gamma] = \kappa_\alpha\eta_{\beta\gamma} - \kappa_\beta\eta_{\alpha\gamma},\\
&[\Delta, \kappa_\alpha] = \kappa_\alpha, \quad [\kappa_\alpha, \kappa_\beta] = 0,
\end{aligned} \tag{2.1}$$

where $\eta_{\alpha\beta}$ is the Minkowskian metric with signature $(+++-)$.

The connection for SO(4,2) can be written

$$\Gamma_i = e_i{}^\alpha \pi_\alpha + \bar{\Gamma}_i, \tag{2.2}$$

where

$$\bar{\Gamma}_i = \tfrac{1}{2}\Gamma_i{}^{\alpha\beta} S_{\alpha\beta} + \phi_i \Delta + \phi_i{}^\alpha \kappa_\alpha. \tag{2.3}$$

The matrix $(e_i{}^\alpha)$ is assumed to be nonsingular, with inverse $(e_\alpha{}^i)$, which can be regarded as the matrix of components of a tetrad. We may employ these matrices to convert Latin (holonomic) to Greek (anholonomic) indices and vice versa. The Minkowskian metric $\eta_{\alpha\beta}$ will be employed for raising and lowering Latin indices.

The infinitesimal element $\bar{\epsilon}$ of the Lie algebra of H can be written

$$\bar{\epsilon} = \tfrac{1}{2}\epsilon^{\alpha\beta} S_{\alpha\beta} + \zeta\Delta + \zeta^\alpha \kappa_\alpha. \tag{2.4}$$

The transformation law (1.1) then has the explicit forms

$$\delta e_i{}^\alpha = \xi^j \partial_j e_i{}^\alpha + e_j{}^\alpha \partial_i \xi^j - e_i{}^\beta(\epsilon_\beta{}^\alpha + \delta_\beta{}^\alpha \zeta), \tag{2.5}$$

$$\begin{aligned}\delta\bar{\Gamma}_i = {}& \xi^j \partial_j \bar{\Gamma}_i + \bar{\Gamma}_j \partial_i \xi^j + \partial_i \bar{\epsilon} + [\bar{\epsilon}, \bar{\Gamma}_i]\\ &- 2e_i{}^\beta(\zeta_\beta \Delta + \zeta^\alpha S_{\alpha\beta}).\end{aligned} \tag{2.6}$$

Observe that the tetrad undergoes Lorentz rotation and dilation under the action of H. Observe also that, due to the final term in (2.6), $\bar{\Gamma}_i$ is *not* a connection for the group H.

At this stage it is possible to impose a metric on the base space (space-time) in a natural way. We define the space-time metric to be the one with respect to which the tetrad is orthonormal:

$$g_{ij} = e_i{}^\alpha e_j{}^\beta \eta_{\alpha\beta}. \tag{2.7}$$

Under the local action of H, this metric responds according to

$$\delta g_{ij} = -2\zeta g_{ij}. \tag{2.8}$$

Thus, the subgroup of SO(4,2) generated by Δ can be identi-

 0022-2488/86/123051-04$02.50

fied as Weyl's group of scale transformations.

It is also possible to impose a holonomic linear connection on space-time. We introduce the generalized derivative of the tetrad field (see I):

$$D_i e_j{}^\alpha = \partial_i e_j{}^\alpha + e_j{}^\beta \Gamma_{i\beta}{}^\alpha + e_j{}^\alpha \phi_i, \tag{2.9}$$

and then define

$$\Gamma_{ij}{}^k = e_\alpha{}^k D_i e_j{}^\alpha. \tag{2.10}$$

The $\Gamma_{ij}{}^k$ transform under space-time diffeomorphisms like the components of a linear connection. Moreover, it is a metric-compatible connection:

$$\partial_k g_{ij} - \Gamma_{ki}{}^l g_{lj} - \Gamma_{kj}{}^l g_{il} = 0. \tag{2.11}$$

Under the action of H, it has the transformation law

$$\delta \Gamma_{ij}{}^k = 2(\zeta^k g_{ij} - \delta_i{}^k \zeta_j - \delta_j{}^k \zeta_i). \tag{2.12}$$

We now consider the limiting case with vanishing SO(4,2) curvature:

$$G_{ij} = 0. \tag{2.13}$$

The coordinate system and H-gauge can then be chosen so that

$$e_i{}^\alpha = \delta_i{}^\alpha, \quad \overline{\Gamma}_i = 0. \tag{2.14}$$

We then see from (2.7) that the space-time has become Minkowskian. In this reference system, the distinction between Latin and Greek indices is lost and the conditions for the transformations (1.1) to preserve the relations (2.14) are

$$\begin{aligned} &\partial_\gamma \xi^\alpha = \epsilon_\gamma{}^\alpha + \zeta \delta_\gamma{}^\alpha, \\ &\partial_\gamma \epsilon^{\alpha\beta} = 2(\delta_\gamma{}^\beta \zeta^\alpha - \delta_\gamma{}^\alpha \zeta^\beta), \\ &\partial_\gamma \zeta = 2\zeta_\gamma, \quad \partial_\gamma \zeta^\alpha = 0, \end{aligned} \tag{2.15}$$

[cf. Eqs. (8.2) of I]. The integration of these equations is straightforward. We get, successively,

$$\begin{aligned} &\zeta^\alpha = c^\alpha, \quad \zeta = 2c_\alpha x^\alpha + \rho, \\ &\epsilon^{\alpha\beta} = 2(x^\beta c^\alpha - x^\alpha c^\beta) + \omega^{\alpha\beta}, \end{aligned} \tag{2.16}$$

and finally

$$\xi^\alpha = a^\alpha + x_\gamma \omega^{\gamma\alpha} + \rho x^\alpha + 2x^\alpha c\cdot x - c^\alpha x^2, \tag{2.17}$$

where $a^\alpha, \rho, \omega^{\alpha\beta}$, and c^α are constants of integration; x^2 and $c\cdot x$ denote $x^\alpha x^\beta \eta_{\alpha\beta}$ and $c^\alpha x^\beta \eta_{\alpha\beta}$, respectively. We recognize that the diffeomorphisms $x^\alpha \to x^\alpha - \xi^\alpha$ are the infinitesimal conformal mappings on Minkowski space-time.

The transformation law (1.2) for a field ψ becomes

$$\delta\psi = \xi^\alpha \partial_\alpha \psi + (\tfrac{1}{2}\epsilon^{\alpha\beta} S_{\alpha\beta} + \zeta\Delta + \zeta^\alpha \kappa_\alpha)\psi, \tag{2.18}$$

with $\xi^\alpha, \epsilon^{\alpha\beta}, \zeta$, and ζ^α given by (2.16) and (2.17). Thus we have precisely the transformation law of a physical field on Minkowski space-time, under the action of infinitesimal conformal transformations[2]:

$$\begin{aligned} \delta\psi = [&a^\alpha \partial_\alpha + \tfrac{1}{2}\omega^{\alpha\beta}(S_{\alpha\beta} + x_\alpha\, \partial_\beta - x_\beta\, \partial_\alpha) \\ &+ \rho(\Delta + x^\alpha\, \partial_\alpha) + c^\alpha(\kappa_\alpha + 2(x_\alpha \Delta + x^\beta S_{\alpha\beta}) \\ &+ 2(x_\alpha x^\beta - x^2 \delta_\alpha^\beta)\partial_\beta)]\psi. \end{aligned} \tag{2.19}$$

The reverse of the procedure carried out above is to start with the (global) action of the conformal group on Minkowski space and on fields ψ [given by (2.17) and (2.19)] and then "gauge" the group by making the parameters space-time dependent and introducing auxiliary fields. That is, the conformal group can be gauged in a manner analogous to Kibble's[3,4] gauging of the Poincaré group. The details have been presented elsewhere.[5]

III. THE DE SITTER GAUGE THEORY

The commutation relations for the generators of SO(4,1) can be displayed in the form

$$\begin{aligned} &[\pi_\alpha, \pi_\beta] = -\kappa S_{\alpha\beta}, \\ &[\pi_\alpha, S_{\beta\alpha}] = \eta_{\alpha\beta}\pi_\gamma - \eta_{\alpha\gamma}\pi_\beta, \\ &[S_{\alpha\beta}, S_{\gamma\delta}] = \eta_{\beta\gamma} S_{\alpha\delta} - \eta_{\alpha\gamma} S_{\beta\delta} + \eta_{\alpha\delta} S_{\beta\gamma} - \eta_{\beta\delta} S_{\alpha\gamma}. \end{aligned} \tag{3.1}$$

The subgroup H, generated by the $S_{\alpha\beta}$, is just the Lorentz group. The constant κ is inserted so that the Poincaré group can be regarded as a limiting case.

Introduce the connection

$$\Gamma_i = e_i{}^\alpha \pi_\alpha + \tfrac{1}{2}\Gamma_i{}^{\alpha\beta} S_{\alpha\beta}. \tag{3.2}$$

The transformation law (1.1) becomes

$$\delta e_i{}^\alpha = \xi^j\, \partial_j e_i{}^\alpha + e_j{}^\alpha\, \partial_i \xi^j - e_i{}^\beta \epsilon_\beta{}^\alpha, \tag{3.3}$$

$$\begin{aligned} \delta\Gamma_{i\alpha}{}^\beta = {}&\xi^j\, \partial_j \Gamma_{i\alpha}{}^\beta + \Gamma_{j\alpha}{}^\beta\, \partial_i \xi^j + \partial_i \epsilon_\alpha{}^\beta \\ &+ \epsilon_\alpha{}^\gamma \Gamma_{i\gamma}{}^\beta - \Gamma_{i\alpha}{}^\gamma \epsilon_\gamma{}^\beta \end{aligned} \tag{3.4}$$

(in which the Minkowskian metric has been used for raising and lowering Greek indices). Observe that the tetrad is Lorentz rotated by the action of H and that $\Gamma_i{}^{\alpha\beta}$ transforms like a connection for the Lorentz group. We shall employ the symbol D_i to denote the corresponding covariant differentiation. For example,

$$D_i\psi = \partial_i \psi - \tfrac{1}{2}\Gamma_i{}^{\alpha\beta} S_{\alpha\beta}\psi$$

and

$$D_i e_j{}^\alpha = \partial_i e_j{}^\alpha + e_j{}^\beta \Gamma_{i\beta}{}^\alpha.$$

The Lorentz torsion and Lorentz curvature are defined by

$$F_{ij}{}^\alpha = D_i e_j{}^\alpha - D_j e_i{}^\alpha, \tag{3.5}$$

and

$$F_{ij\alpha}{}^\beta = \partial_i \Gamma_{j\alpha}{}^\beta - \partial_j \Gamma_{i\alpha}{}^\beta - \Gamma_{i\alpha}{}^\gamma \Gamma_{j\gamma}{}^\beta + \Gamma_{j\alpha}{}^\gamma \Gamma_{i\gamma}{}^\beta. \tag{3.6}$$

The SO(4,1) curvature is

$$G_{ij} = F_{ij}{}^\alpha \pi_\alpha + \tfrac{1}{2}(F_{ij}{}^{\alpha\beta} + 2\kappa e_i{}^\alpha e_j{}^\beta) S_{\alpha\beta}. \tag{3.7}$$

A holonomic metric and holonomic connection on space-time can be constructed in a natural way from the SO(4,1) connection coefficients. We define

$$g_{ij} = e_i{}^\alpha e_j{}^\beta \eta_{\alpha\beta} \tag{3.8}$$

and

$$\Gamma_{ij}{}^k = e_\alpha{}^k D_i e_j{}^\alpha. \tag{3.9}$$

The connection (3,9) is metric compatible, that is,

$$\partial_i g_{jk} - \Gamma_{ij}{}^l g_{lk} - \Gamma_{ik}{}^l g_{jl} = 0. \tag{3.10}$$

It is not, in general, symmetric:

$$\Gamma_{ij}{}^k - \Gamma_{ij}{}^k = F_{ij}{}^k. \tag{3.11}$$

Thus, the definitions (3.8) and (3.9) impose on the space-time a U(4) structure[6] [in fact, as is apparent from (3.3) and (3.4), the gauged de Sitter group and the gauged Poincaré group are identical].

Now consider the limiting case in which the SO(4,1) curvature (3.7) vanishes. The torsion then vanishes, so the connection (3.9) becomes the Christoffel connection

$$\Gamma_{ij}{}^{k} = \begin{Bmatrix} k \\ ij \end{Bmatrix}. \tag{3.12}$$

The Lorentz curvature (which is now just an anholonomic version of the Riemann tensor constructed from g_{ij}) does not vanish. We have

$$R_{ij}{}^{kl} = F_{ij}{}^{kl} = \kappa(\delta_i{}^l\delta_j{}^k - \delta_j{}^l\delta_i{}^k). \tag{3.13}$$

Thus, the space-time has become a *space of constant curvature*. We can therefore choose the coordinate system to be a system of stereographic coordinates for which

$$g_{ij} = \sigma^2\,\eta_{ij}, \tag{3.14}$$

$$\sigma = 1/[1 + (\kappa x^2/4)], \quad x^2 = \eta_{ij}x^ix^j. \tag{3.15}$$

We can then take the tetrad components to be

$$e_i{}^\alpha = \sigma\delta_i{}^\alpha. \tag{3.16}$$

It is convenient from now on to convert Latin indices to Greek indices, and vice versa, by means of δ_i^α rather than e_i^α. With this understood, the Lorentz connection determined by (3.9) and (3.12) turns out to be

$$\Gamma_i{}^{\alpha\beta} = \kappa\sigma\delta_i^{[\alpha}x^{\beta]}. \tag{3.17}$$

The transformations (3.3) and (3.4) that leave unchanged these particular functional forms for the tetrad and Lorentz connection are those with parameters ξ^i and $\epsilon^{\alpha\beta}$ satisfying

$$\partial^\alpha\xi^\beta - \tfrac{1}{2}\kappa\sigma\xi\cdot x\eta^{\alpha\beta} - \epsilon^{\alpha\beta} = 0, \tag{3.18}$$

and

$$\partial_i\epsilon^{\alpha\beta} + \tfrac{1}{2}\kappa\sigma[\delta_i{}^\alpha(\xi^\beta + \epsilon^{\beta\gamma}x_\gamma) - \delta_i^\beta(\xi^\alpha + \epsilon^{\alpha\gamma}x_\gamma)] = 0. \tag{3.19}$$

Fortunately, we already have partial knowledge about the solution of these equations. The diffeomorphisms that preserve the de Sitter metric are the de Sitter transformations, which, in terms of the stereographic coordinate system, have the infinitesimal form $x^\alpha \to x^\alpha - \xi^\alpha$, where

$$\xi^\alpha = x_\beta\omega^{\beta\alpha} + a^\alpha(1 - (\kappa x^2/4)) + (\kappa/2)x^\alpha a\cdot x; \tag{3.20}$$

$\omega^{\alpha\beta}$ and a^α being the (constant) parameters of the group. Substituting this expression into (3.18) gives

$$\epsilon^{\alpha\beta} = \omega^{\alpha\beta} + (\kappa/2)(a^\alpha x^\beta - a^\beta x^\alpha). \tag{3.21}$$

It is then not difficult to check that Eq. (3.19) is also satisfied.

Equation (1.2) now gives the transformation law for a physical field (belonging to a representation of the Lorentz group) under the action of an infinitesimal de Sitter transformation on a de Sitter space-time:

$$\delta\psi = a^\alpha\left[\left(1 - \frac{\kappa x^2}{4}\right)\partial_\alpha + \frac{\kappa}{2}x_\alpha x^\beta\partial_\beta + \frac{\kappa}{2}S_{\alpha\beta}x^\beta\right]\psi + \tfrac{1}{2}\omega^{\alpha\beta}[x_\alpha\,\partial_\beta - x_\beta\,\partial_\alpha + S_{\alpha\beta}]\psi. \tag{3.22}$$

IV. CONCLUDING REMARKS

Many attempts to construct a gauge theory of a space-time symmetry group encounter difficulties and complications. The reader is referred to the review article of Ivanenko and Sardanashvily[7] and the references cited therein. The difficulties arise from attempting a too close analogy with the pattern established by gauge theories of internal symmetries; if the whole of a space-time group G is "gauged" in the Yang–Mills sense, the gauged "internal translations" destroy the possibility of identifying the translational gauge potentials with a tetrad.[7,8] In our view, in a correct approach to gauging a space-time symmetry G, only the subgroup H is localized in the Yang–Mills sense; the gauged generalization of G in our scheme consists of a local action of H together with *general diffeomorphisms* [or alternatively, general coordinate transformations (GCT)] on space-time M. This viewpoint is already implicit in the de Sitter gauge theory of MacDowell and Mansouri,[9] where invariance of the Lagrangian only under local Lorentz transformations and GCT was imposed. The geometrical background to the MacDowell and Mansouri de Sitter gauge theory corresponds to our scheme [where G is the de Sitter group or its covering group Sp(2,2)].

That the local action of H together with general diffeomorphisms (or GCT) on M does indeed constitute a true gauge theory of a space-time group G is fully justified only when one has shown that the limiting case of "ungauged" transformations does in fact correspond to the correct global action of G on M and on fields in M. The purpose of this work was to demonstrate that this is so for the conformal group, the de Sitter group, and (by Wigner–Inönü contraction of the de Sitter case) the Poincaré group. The "ungauged" limit of Poincaré gauge theory was obtained by Hehl.[6] The gauging of the affine group in accordance with our scheme has been presented elsewhere.[9]

The transformation laws for the points x of M and the matter fields ψ on M, under the global ("ungauged") action of G constitute essentially a *nonlinear realization* of G in the sense of Coleman, Wess, and Zumino[10] or Salam and Strathdee.[11] However, *space-time itself takes the place of the Goldstone fields*, so the usual dynamics of nonlinear realization (Higgs mechanism, spontaneous symmetry breakdown) is not called into play. Thus, our scheme differs radically from that of the Poincaré and de Sitter gauge theories of Tseytlin,[12] in which the whole of G rather than just H acts "internally," but the usual difficulties associated with such a scheme are avoided by realizing the translations nonlinearly. This nonlinear realization of G is associated with spontaneous symmetry breakdown, the broken symmetries being the internal translations.

The relationship between our approach to the gauging of space-time symmetries and that of other approaches becomes clearer when our scheme is expressed in the language of fiber bundles. It is clear from our foregoing remarks that only the subgroup H should act on the fibers, not the whole of G ("no internal translation"). The simplest and most natural translation of our scheme into fiber bundle language consists of expressing the gauge theory of a group G involving space-time and internal symmetries in terms of the group manifold G; specifically, in terms of the principal fiber bundle $G(G/H,H)$ where the coset space G/H is space-time[13] (note that H, not G, is the structural group). This aspect will be dealt with in a subsequent paper.

[1]E. A. Lord and P. Goswami, J. Math. Phys. **27**, 2415 (1986).
[2]G. Mack and A. Salam, Ann. Phys. (NY) **53**, 174 (1969).
[3]T. W. B. Kibble, J. Math. Phys. **2**, 212 (1961).
[4]N. Mukunda, "Gauge approach to classical gravity," in *Proceedings of the Workshop on Gravitation and Relativistic Astrophysics*, Ahmedabad, India, edited by A. R. Prasanna, J. V. Narlikar, and C. V. Vishveshwara (Indian Academy of Science, Bangalore, 1982).
[5]E. A. Lord and P. Goswami, Pramana **25**, 635 (1985).
[6]F. W. Hehl, "Four lectures on Poincaré gauge field theory," in *Proceedings of the 6th course of the International School of Cosmology and Gravitation*, edited by P. G. Bergmann and V. de Sabbata (Plenum, New York, 1978).
[7]D. Ivanenko and G. Sardanashvily, Phys. Rep. **94**, 1 (1983).
[8]T. Dass, Pramana **23**, 433 (1984).
[9]S. W. MacDowell and F. Mansouri, Phys. Rev. Lett. **38**, 739 (1977).
[10]S. Coleman, J. Wess, and B. Zumino, Phys. Rev. D **12**, 1711 (1969).
[11]A. Salam and J. Strathdee, Phys. Rev. **184**, 1750 (1969).
[12]A. A. Tseytlin, Phys. Rev. D **26**, 3327 (1982).
[13]Y. Ne'eman and T. Regge, Riv. Nuovo Cimento **1** (5), 1 (1978).

An exactly soluble relativistic quantum two-fermion problem

A. O. Barut[a)] and N. Ünal[b)]
Sektion Physik der Universität München, 8 München 2, West Germany

(Received 4 April 1986; accepted for publication 18 June 1986)

The two-fermion problem of quantum electrodynamics in which both particles are treated relativistically and full spin degrees of freedom are taken into account is shown to be exactly soluble when potentials up to order α^4 are kept. It is therefore a good starting point for radiative corrections of higher order for the precision tests of QED in bound-state problems. Recoil corrections are included to all orders.

I. INTRODUCTION

There are, to our knowledge, no examples of exactly soluble realistic spin-$\frac{1}{2}$ two-body problems in which both particles are treated relativistically. We present here a case that has been extracted from a fully covariant two-body equation in quantum electrodynamics. It is realistic in the sense that it gives a spectrum for the H or positronium atoms correct up to order α^4 and contains moreover the recoil corrections to all orders. It can therefore be used as a good starting point for radiative corrections in the precision tests of quantum electrodynamics for the remaining terms of the order of α^5 and higher.

We shall also compare this system with the covariant infinite-component wave equation with exactly the same spectrum. In the latter case the composite structure of the system is characterized algebraically by a dynamical group rather than in terms of the parameters of the constituents as a dynamical bound state of two particles.

II. COVARIANT TWO-BODY EQUATION

The starting point is the covariant two-body equation

$$\Big[(\overset{(1)}{\gamma}{}^{\mu}\cdot p_{1\mu}-m_1)\otimes\overset{(2)}{\gamma}\cdot n + \overset{(1)}{\gamma}\cdot n\otimes(\overset{(2)}{\gamma}\cdot p_2-m_2)+V(d)\Big]\Phi=0, \qquad (1)$$

derived directly from the coupled Maxwell–Dirac equations in a nonperturbative way by a variational principle.[1–3]

Here $\overset{(1)}{\gamma}_{\mu}$ and $\overset{(2)}{\gamma}_{\mu}$ are the Dirac algebras for both particles so that Eq. (1) is a (16×16)-spinor equation. Further n_μ is a four-vector normal to the spacelike surface associated with the relative coordinate and $\gamma\cdot n\equiv\gamma^\mu n_\mu$. The relativistic potential $V(d)$ is a function of the covariant relative distance of the two particles $d=\sqrt{(x\cdot n)^2-x^2}$. For the explicit solutions in this paper we shall from now on choose $n^\mu=(1,0,0,0)$, whence $\gamma\cdot n=\gamma^0$ and $d=r$, the magnitude of the relative three-vector $\mathbf{r}$. The spin matrices we write always as the direct products $A\otimes B$, where A refers to particle 1 and B to particle 2.

Equation (1) has many remarkable properties, among them the exact separability of the center of mass and relative coordinates. One then sees that it is actually a *one-time equation*. The dependence on the relative time drops out automatically. The equation for the center of mass is[2]

$$((m_1/M)\boldsymbol{\alpha}_1+(m_2/M)\boldsymbol{\alpha}_2)\cdot\mathbf{P}\phi(R)=(E_0-E)\phi(R), \qquad (2)$$

whereas the relative motion is given by

$$[(\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2)\cdot\mathbf{p}+\beta_1 m_1+\beta_2 m_2+\beta_1\otimes V\otimes\beta_2]\psi(r)=E\psi(r), \qquad (3)$$

where E is the energy in the center of mass frame (total mass of the system) and E_0 the total energy of the moving system so that the difference (E_0-E) in (2) is the relative kinetic energy of the center of mass: $M=m_1+m_2$.

For the coupling of the spinor fields to a vector field A_μ of the form $e\bar\psi\gamma^\mu\psi A_\mu$ and for an effective anomalous magnetic moment coupling of the form $a\bar\psi\sigma_{\mu\nu}\psi F^{\mu\nu}$ the form of the relativistic potential has been derived. The first coupling gives

$$V(r)=(e_1e_2/r)\overset{(1)}{\gamma}{}^{\mu}\otimes\overset{(2)}{\gamma}_{\mu}. \qquad (4)$$

The second potential coming from the Pauli coupling is rather lengthy and since it has been given elsewhere,[2,3] we do not write it here but shall give its radial form later.

In the derivation of Eq. (3) from field theory there are also self-energy terms corresponding to Lamb shift and spontaneous emission. These are of order of $\alpha(Z\alpha)^4$ and higher and will be taken into account separately.

III. RADIAL EQUATIONS

For Eq. (3) we can also separate completely the radial and angular parts.[3] This results in two sets of eight first-order radial wave equations. In each set four of the eight equations are algebraic and the other four are first-order differential equations. Eliminating some of the components of the wave functions we arrive *for the first set* at the following *two coupled second-order equations* (including Pauli terms):

$$\Big\{\frac{V_1V_6}{4V_3}+\Big[\frac{V_6}{V_3}\partial_+\frac{V_3}{V_6}+\frac{2\lambda Mj(j+1)}{Er^4V_3}\Big(\frac{1}{E}+\frac{3}{V_5}\Big)\Big]\partial_- -\frac{2\lambda Mj(j+1)V_6}{EV_3}\partial_+ \times\frac{1}{r^4V_6}\Big(\frac{1}{E}+\frac{3}{V_5}\Big)-\frac{M^2j(j+1)}{E^2r^2V_3}V_2\Big\}(ru_2)$$

$$+M\sqrt{j(j+1)}\Big\{\frac{V_6}{V_3}\partial_+\frac{V_3}{V_6rV_5}$$

---

a) Permanent address: Physics Department, The University of Colorado, Boulder, Colorado 80309.

b) Permanent address: Dicle University, Diyarbakir, Turkey.

 0022-2488/86/123055-06$02.50

$$+\frac{2\lambda Mj(j+1)}{Er^5V_3V_5}\left(\frac{1}{E}+\frac{3}{V_5}\right)$$
$$-\left[\frac{V_2}{rEV_3}+\frac{2\lambda V_6}{MV_3}\partial_+\frac{1}{V_6r^3}\left(\frac{1}{E}+\frac{3}{V_5}\right)\right]\tilde{\partial}_+\Big\}(rv_0)$$
$$=0, \tag{5}$$
$$\Big\{\frac{V_4V_6}{4V_2}+\left[\frac{V_6}{V_2}\tilde{\partial}_- \frac{V_2}{V_6}+\frac{2\lambda Mj(j+1)}{r^4V_2V_5}\left(\frac{1}{E}+\frac{3}{V_5}\right)\right]\tilde{\partial}_+$$
$$-\frac{2\lambda Mj(j+1)V_6}{V_2}\tilde{\partial}_-\frac{1}{V_6V_5r^4}\left(\frac{1}{E}+\frac{3}{V_5}\right)$$
$$-\frac{M^2j(j+1)}{r^2V_2V_5^2}V_3\Big\}(rv_0)-M\sqrt{j(j+1)}$$
$$\times\Big\{-\frac{V_6}{EV_2}\tilde{\partial}_-\frac{V_2}{rV_6}-\frac{2\lambda Mj(j+1)}{Er^5V_2V_5}\left(\frac{1}{E}+\frac{3}{V_5}\right)$$
$$+\left[\frac{V_3}{rV_5V_2}+\frac{2\lambda}{M}\frac{V_6}{V_2}\tilde{\partial}_-\frac{1}{r^3V_6}\left(\frac{1}{E}+\frac{3}{V_5}\right)\right]\partial_-\Big\}(ru_2)$$
$$=0, \tag{6}$$

where the following abbreviations have been used:

$$\begin{aligned}
V_1(r)&\equiv E+\frac{2\alpha}{r}-\frac{M^2}{E}-\frac{4a_1a_2}{r^3}-\frac{\tau^2}{Er^4},\\
V_2(r)&\equiv E+\frac{2\alpha}{r}-\frac{\Delta m^2}{E}-\frac{4j(j+1)}{r^2V_5}+\frac{4a_1a_2}{r^3}-\frac{\lambda^2}{Er^4},\\
V_3(r)&\equiv E+\frac{4\alpha}{r}-\frac{\Delta m^2}{E-2\alpha/r}-\frac{4j(j+1)}{Er^2}-\frac{9\lambda^2}{r^4V_5},\\
V_4(r)&\equiv E-\frac{M^2}{V_5}-\frac{\tau^2}{r^4(E-2\alpha/r)},\\
V_5(r)&\equiv E+\frac{2\alpha}{r}+\frac{8a_1a_2}{r^3},\\
V_6(r)&\equiv V_2V_3-\frac{4\lambda^2j(j+1)}{r^6}\left(\frac{1}{E}+\frac{3}{V_5}\right),
\end{aligned} \tag{7}$$

and

$$\begin{aligned}
\partial_\pm&\equiv\partial_r\pm\frac{\lambda M+\tau\Delta m}{2Er^2},\\
\tilde{\partial}_\pm&\equiv\partial_r\pm\frac{1}{r}\mp\left(\frac{3\lambda M}{2r^2V_5}-\frac{\tau\Delta m}{2r^2(E-2\alpha/r)}\right).
\end{aligned} \tag{8}$$

Further

$$\lambda=e_1a_2+e_2a_1,\quad \tau=e_1a_2-e_2a_1. \tag{9}$$

We are interested in the solutions of Eq. (5). They are rather complicated. However, if we consider some of the small terms (which are, in the electromagnetic problem, of order α^5 and smaller) as perturbations, we have found that these coupled equations are exactly soluble.

In order to motivate the method of solution and to interpret the angular momentum quantum numbers, we begin with a much simpler case, namely the radial equations of two relativistic free particles in the center of mass frame. Even this case is not trivial in this form[3] and provides us actually the tools to solve the case with interactions.

IV. SOLUTIONS OF THE RADIAL EQUATIONS FOR TWO RELATIVISTIC FREE PARTICLES

First we set all the coupling constants equal to zero:

$$\alpha=0,\quad \lambda=0,\quad \tau=0,\quad a_1a_2=0. \tag{10}$$

Then Eqs. (5) and (6) become

$$\Big\{\frac{1}{4}\left(E-\frac{M^2}{E}\right)\left(E-\frac{\Delta m^2}{E}\right)-\frac{j(j+1)}{r^2}+\partial_r^2$$
$$-\frac{j(j+1)/r^3}{(E^2-\Delta m^2)/4-j(j+1)/r^2}\partial_r\Big\}(ru_2)$$
$$-\frac{2M}{Er^2}\sqrt{j(j+1)}\Big\{\frac{E^2-\Delta m^2}{4}$$
$$\times\frac{1}{(E^2-\Delta m^2)/4-j(j+1)/r^2}\Big\}(rv_0)=0, \tag{11}$$
$$\Big\{\frac{1}{4}\left(E-\frac{M^2}{E}\right)\left(\frac{E-\Delta m^2}{E}\right)-\frac{j(j+1)+2}{r^2}+\partial_r^2$$
$$-\frac{j(j+1)/r^3}{(E^2-\Delta m^2)/4-j(j+1)/r^2}\partial_r\Big\}(rv_0)$$
$$-\frac{2M}{Er^2}\sqrt{j(j+1)}\Big\{\frac{E^2-\Delta m^2}{4}$$
$$\times\frac{1}{(E^2-\Delta m^2)/4-j(j+1)/r^2}\Big\}(ru_2)=0. \tag{12}$$

The only difference between Eqs. (11) and (12) is the term $(-(2/r^2)(rv_0))$ in the first part of the second equation. The other terms are completely symmetrical. In the dimensionless units these equations can be written as

$$\Big\{\partial_\rho^2+1-\frac{j(j+1)}{\rho^2}-\frac{2j(j+1)/\rho^3}{\epsilon^2-j(j+1)/\rho^2}\partial_\rho\Big\}f(\rho)$$
$$-\frac{2a/\rho^3}{\epsilon^2-j(j+1)/\rho^2}g(\rho)=0, \tag{13}$$
$$\Big\{\partial_\rho^2+1-\frac{j(j+1)+2}{\rho^2}-\frac{2j(j+1)/\rho^3}{\epsilon^3-j(j+1)/\rho^2}\partial_\rho\Big\}g(\rho)$$
$$-\frac{2a/\rho^3}{\epsilon^2-j(j+1)/\rho^2}f(\rho)=0, \tag{14}$$

with

$$\rho\equiv kr,$$
$$4k^2\equiv(E^2-M^2)(E^2-\Delta m^2)/E^2,$$

where k has the meaning of momentum in the center of mass frame when E is the center of mass energy

$$\epsilon^2\equiv\frac{E^2-\Delta m^2}{4k^2}=\frac{E^2}{E^2-M^2},\quad a\equiv\epsilon^2M/E,$$
$$f(\rho)\equiv\rho u_2(\rho),\quad g(\rho)\equiv\rho v_0(\rho).$$

We note the following.

(i) Except the coupling terms and the

$$\frac{-2j(j+1)/\rho^3}{\epsilon^2-j(j+1)/\rho^2}\partial_\rho$$

terms these are the equations for the spherical Bessel functions $(\rho j_l(\rho))$.

(ii) Although the Bessel differential equations have singularities at $\rho=0$ and $\rho=\infty$, Eqs. (13) and (14) have thus more singularities at $\rho=\pm\epsilon/\sqrt{j(j+1)}$. These additional singularities are artificial, since the original first-order equations have only the two singularities at $\rho=0$ and $\rho=\infty$. The additional singularities have been introduced in the process of going from the first-order differential equations to second-order differential equations. For this reason we search a regular solution of Eqs. (13) and (14) at $\rho = \pm\epsilon\sqrt{j(j+1)}$. Now we try a solution of Eqs. (13) and (14) in the form of a series of spherical Bessel functions (times ρ). These are

$$f(\rho) = \rho\sum_{n=0}^{\infty} A_n j_{n+s}(\rho),$$

$$g(\rho) = \sum_{n=0}^{\infty} B_n j_{n+s}(\rho).$$

The first and second derivations of $f(\rho)$ and $g(\rho)$ are

$$\partial_\rho f(\rho) = \sum_{n=0}^{\infty} A_n\Big[j_{n+s}(\rho) + \frac{\rho}{2(n+s)+1}\big((s+n)j_{n+s-1} - (s+n+1)j_{n+s+1}\big)\Big], \qquad (15)$$

$$\partial_\rho^2 f(\rho) = \rho\sum_{n=0}^{\infty} A_n\Big(-1 + \frac{(n+s)(n+s+1)}{\rho^2}\Big)j_{n+s}, \qquad (16)$$

and similar expressions for $g(\rho)$. Inserting these relations into Eqs. (13) and (14) we obtain

$$\sum_{n=0}^{\infty}\Big\{\Big[-\frac{j(j+1)}{\rho^3}\Big(\frac{(n+s+1)(n+s+2)-j(j+1)}{2(n+s)+1}j_{n+s-1} + \frac{(n+s)(n+s-1)-j(j+1)}{2(n+s)+1}j_{n+s+1}\Big)A_n\Big] + \frac{1}{\rho^2}\Big[\epsilon^2((n+s)(n+s+1)-j(j+1))A_n - 2aB_n\Big]j_{n+s}\Big\} = 0, \qquad (17)$$

$$\sum_{n=0}^{\infty}\Big\{\Big[-\frac{j(j+1)}{\rho^3}\Big(\frac{(n+s+1)(n+s+2)-j(j+1)}{2(n+s)+1}j_{n+s-1} + \frac{(n+s)(n+s-1)-j(j+1)}{2(n+s)+1}j_{n+s+1}\Big)B_n\Big] + \frac{1}{\rho^2}\Big[\epsilon^2((n+s)(n+s+1)-j(j+1)-2)B_n - 2aA_n\Big]j_{n+s}\Big\} = 0. \qquad (18)$$

For $n = 0$, these equations give the following condition:

$$-\frac{j(j+1)}{\rho^3}\,\frac{(s+1)(s+2)-j(j+1)}{2s+1}j_{s-1}\begin{Bmatrix}A_0\\B_0\end{Bmatrix} = 0. \qquad (19)$$

The solutiuon of these individual equations gives

$$s = j-1 \quad\text{or}\quad s = j, \qquad (20)$$

with $A_0 \neq 0$; $B_0 \neq 0$. We choose the positive one: $s = j-1$. In order to get the recursion relations for the coefficients we eliminate the $(1/\rho)$-term in the first part of Eqs. (17) and (18). This can be done by using the following functional relations of spherical Bessel functions

$$\frac{1}{\rho}j_{l-1}(\rho) = \frac{1}{2l-1}[j_{l-2}(\rho) + j_l(\rho)]. \qquad (21)$$

By inserting this into Eqs. (17) and (18) we find the following recursion relations:

$$-j(j+1)\Big[\frac{(n+4)(n+5+2j)}{(2(n+4+j)-1)(2(n+j+4)-3)}A_{n+4} + \frac{1}{2(n+j+2)-1}\Big(\frac{(n+2)(n+3+2j)}{2(n+j+2)-3} + \frac{n(n+2j+1)}{2(n+j+2)+1}\Big)A_{n+2} + \frac{(n-2)(n+2j-1)}{2(n+j+1)(2(n+j)-1)}A_n\Big] + \Big[\epsilon^2((n+j+1)(n+j+2)-j(j+1))A_{n+2} - 2aB_{n+2}\Big] = 0, \qquad (22)$$

$$-j(j+1)\Big[\frac{(n+4)(n+2j+5)}{(2(n+j)+7)(2(n+j)+5)}B_{n+4} + \frac{1}{2(n+j)+3}\Big(\frac{(n+2)(n+2j+3)}{2(n+j)+1} + \frac{n(n+2j+1)}{2(n+j)+5}\Big)B_{n+2} + \frac{(n-2)(n+2j-1)}{(2(n+j)+1)(2(n+j)-1)}B_n\Big] + [-2aA_{n+2} + \epsilon^2((n+j+1)(n+j+2)-j(j+1)-2)B_{n+2}] = 0. \qquad (23)$$

Starting from $n = -2$, with $A_{-2} = B_{-2} = 0$, we then obtain the following relations between A_2, B_2, A_0, and B_0:

$$\frac{j+1}{2j+1}A_2 + \Big(\epsilon^2 - \frac{j+1}{2j+1}\Big)A_0 + \frac{aB_0}{j} = 0, \qquad (24a)$$

$$\Big(\epsilon^2 - \frac{j}{2j+1}\Big)A_2 - \frac{a}{j+1}B_2 + \frac{j}{2j+1}A_0 = 0, \qquad (24b)$$

$$-\frac{a}{j}A_2 + \Big(\epsilon^2 - \frac{j+1}{2j+1}\Big)B_2 + \frac{j+1}{2j+1}B_0 = 0, \qquad (24c)$$

$$\frac{j}{2j+1}B_2 + \frac{a}{j+1}A_0 + \Big(\epsilon^2 - \frac{j}{2j+1}\Big)B_0 = 0. \qquad (24d)$$

The determinant of the coefficients of Eq. (24) is zero. So it has a nontrivial solution given by

$$A_2=\left[\left(1-\epsilon^2\frac{2j+1}{j+1}\right)A_0-\frac{2j+1}{j(j+1)}\,aB_0\right],$$
$$B_2=\left[-\frac{2j+1}{j(j+1)}\,aA_0+\left(1-\epsilon^2\frac{2j+1}{j}\right)B_0\right]. \tag{25}$$

Going back to the recursion relations (22) and (24) with $n=0$ and the solution given by (25) we get

$$A_4=B_4=0.$$

Next for $n=2$ we obtain

$$A_6=B_6=0.$$

This means that

$$A_{2n+4}=B_{2n+4},\quad \text{for all } n>0.$$

Hence the solutions of the coupled differential Eqs. (13) and (14) are

$$f(\rho)=A_0 j_{j-1}(\rho)+A_2 j_{j+1}(\rho), \tag{26}$$
$$g(\rho)=B_0 j_{j-1}(\rho)+B_2 j_{j+1}(\rho), \tag{27}$$

where the relation between A_2, B_2 and A_0, B_0 is given in Eq. (25).

Physically we see that the components (ρu_2) and (ρv_0) of our wave functions in Eqs. (11) and (12) represent states that are superpositions of two angular momenta $l=j+1$ and $l=j-1$.

The spectrum is given by

$$E^2=2k^4+m_1^2+m_2^2\pm 2\sqrt{k^4+k^2(m_1^2+m_2^2)+m_1^2\,m_2^2}.$$

V. INTERACTING PARTICLES

In this section we discuss a second limit of Eqs. (5) and (6). This limit is obtained by expanding the potentials as a power series of α/r and taking the terms up to the fifth power of α. In the power counting l/r is counted as α. This process gives the following set of coupled second-order differential equations:

$$\left[\frac{1}{4}\left(E-\frac{M^2}{E}\right)\left(E-\frac{\Delta m^2}{E}\right)+\frac{\alpha}{r}\left(E-\frac{M^2+\Delta m^2}{2E}\right)-\frac{j(j+1)-\alpha^2}{r^2}+\partial_r^2\right](ru_2)-\frac{2\sqrt{j(j+1)}}{r^2}\,rv_0(r)=0, \tag{28}$$

$$\left[\frac{1}{4}\left(E-\frac{M^2}{E}\right)\left(E-\frac{\Delta m^2}{E}\right)+\frac{\alpha}{r}\left(E-\frac{M^2+\Delta m^2}{2E}\right)-\frac{j(j+1)+2-\alpha^2}{r^2}+\partial_r^2\right](rv_0)-\frac{2\sqrt{j(j+1)}}{r^2}\,ru_2(r)=0. \tag{29}$$

Here again, the only difference between Eqs. (28) and (29) is the term $-(2/r^2)(rv_0)$ in Eq. (29). The remaining terms are symmetrical in both equations. In the dimensionless units these equations are

$$\left[-\frac{1}{4}+\frac{2Z}{\rho}-\frac{l(l+1)}{\rho^2}+\partial_\rho^2\right]f(\rho)-\frac{2\sqrt{j(j+1)}}{\rho^2}\,g(\rho)=0, \tag{30}$$

$$\left[-\frac{1}{4}+\frac{2Z}{\rho}-\frac{l(l+1)+2}{\rho^2}+\partial_\rho^2\right]g(\rho)-\frac{2\sqrt{j(j+1)}}{\rho^2}\,f(\rho)=0, \tag{31}$$

where

$$\rho=2\lambda r, \tag{32a}$$
$$4\lambda^2=(M^2-E^2)(E^2-\Delta m^2)/E^2, \tag{32b}$$
$$2Z=\frac{\alpha}{2\lambda}\left(E-\frac{M^2+\Delta m^2}{2E}\right), \tag{32c}$$
$$l(l+1)=j(j+1)-\alpha^2. \tag{32d}$$

These equations have two singular points, $\rho=0$ and $\rho=\infty$. The point $r=0$ is a regular singularity, while $\rho=\infty$ is an irregular singularity. At $\rho=\infty$ the equations simplify

$$\left(-\frac{1}{4}+\frac{d^2}{d\rho^2}\right)\begin{pmatrix}f(\rho)\\ g(\rho)\end{pmatrix}=0, \tag{33}$$

so that the regular solution at infinity is

$$\begin{pmatrix}f(\rho)\\ g(\rho)\end{pmatrix}\to e^{-(1/2)\rho}. \tag{34}$$

At $\rho\simeq 0$ the equations are

$$\left(\frac{d^2}{d\rho^2}-\frac{l(l+1)}{\rho^2}\right)f(\rho)-\frac{2\sqrt{j(j+1)}}{\rho^2}\,g(\rho)=0, \tag{35}$$

$$\left(\frac{d^2}{d\rho^2}-\frac{l(l+1)+2}{\rho^2}\right)g(\rho)-\frac{2\sqrt{j(j+1)}}{\rho^2}\,f(\rho)=0. \tag{36}$$

We assume a powerlike behavior of the solution at the origin

$$\begin{pmatrix}f(\rho)\\ g(\rho)\end{pmatrix}\to\begin{pmatrix}A_0\\ B_0\end{pmatrix}\rho^s. \tag{37}$$

Insertion of this ansatz into Eqs. (35) and (36) gives the following relation:

$$(s(s-1)-l(l+1))A_0-2\sqrt{j(j+1)}\,B_0=0,$$
$$-2\sqrt{j(j+1)}\,A_0+(s(s-1)-l(l+1)-2)B_0=0. \tag{38}$$

Hence the condition for the existence of a nontrivial solution is

$$s(s-1)=j(j-1)-\alpha^2 \tag{39}$$

or

$$s=\frac{1}{2}+\sqrt{\frac{1}{4}+j(j-1)-\alpha^2}=j+\left(j-\frac{1}{2}\right)\left[\sqrt{1-\frac{\alpha^2}{(j-\frac{1}{2})^2}}-1\right]\simeq j-\frac{\alpha^2}{2(j-\frac{1}{2})}+O(\alpha^4). \tag{40}$$

This s-value is in agreement with the s-value in Eq. (17) for $\alpha=0$ case. In order to find a regular solution for all ρ's we write $f(\rho)$ and $g(\rho)$ in the form

$$f(\rho)=e^{-\rho/2}\rho^s y(\rho), \tag{41}$$
$$g(\rho)=e^{-\rho/2}\rho^s z(\rho). \tag{42}$$

Now instead of searching a power series solution for $y(\rho)$ and $g(\rho)$ we assume a solution that is a series of confluent hypergeometric functions. In the free-particle case we already obtained in the previous section such a two-term series with orbital angular momentum $l=j-1$ and $l=j+1$. Except for the term $-2/r^2 g(\rho)$ and the coupling terms Eqs. (30) and (31) are the same as the Schrödinger equation for the hydrogen atom. The Coulomb problem has the following solutions:

$$R_{n,l}(\rho)=e^{-\rho/2}\rho^{l+1}\,{}_1F_1(-n+l,2l+2;\rho)\,, \tag{43}$$

where F_1 is the confluent hypergeometric function.

For Eqs. (30) and (31) we try again a two-term solution with $l=s$ and $l=s'$ of the form

$$f(\rho)=e^{-\rho/2}[A_0\rho^s\,{}_1F_1(-n+s,2s;\rho) + A_2\rho^{s'+2}\,{}_1F_1(-n+s'+2,2s'+4;\rho)]\,, \tag{44}$$

$$g(\rho)=e^{-\rho/2}[B_0\rho^s\,{}_1F_1(-n+s,2s;\rho) + B_2\rho^{s'+2}\,{}_1F_1(-n+s'+2,2s'+4;\rho)]\,. \tag{45}$$

We shall make use of the following property of the functions R_{nl} that can be proved by using the functional relations of confluent hypergeometric functions:

$$\frac{d^2}{d\rho^2}R_{nl}(\rho) =\frac{d^2}{d\rho^2}\left[e^{-\rho/2}\rho^{l+1}\,{}_1F_1(-n+l+1,2l+2;\rho)\right] =\left(\frac{1}{4}+\frac{n}{\rho}+\frac{l(l+1)}{\rho^2}\right)R_{nl}\,. \tag{46}$$

We insert (44) and (45) into (30) and (31), and, by using (46), we obtain the following relations

$$\left[\frac{2Z-n}{\rho}+\frac{s(s-1)-l(l+1)}{\rho^2}\right]A_0R_{ns} +\left[\frac{2Z-n}{\rho}+\frac{(s'+2)(s'+1)-l(l+1)}{\rho^2}\right]A_2R_{n,s'+2} -\frac{2\sqrt{j(j+1)}}{\rho^2}[B_0R_{n,s}+B_2R_{n,s'+2}]=0\,, \tag{47}$$

$$\left[\frac{2Z-n}{\rho}+\frac{s(s-1)-l(l+1)-2}{\rho^2}\right]B_0R_{n,s} +\left[\frac{2Z-n}{\rho}+\frac{(s'+1)(s'+2)-l(l+1)-2}{\rho^2}\right] \times B_2R_{n,s'+2} -\frac{2\sqrt{j(j+1)}}{\rho^2}[A_0R_{n,s}+A_2R_{n,s'+2}]=0\,. \tag{48}$$

If we choose

$$2Z=n\,, \tag{49}$$

then the $1/\rho$-terms drop out, and we get relations between the coefficient of $R_{n,s}$ and $R_{n,s'}$. The relation between A_0 and B_0 are the same as Eq. (38). The relation between A_2 and B_2 are

$$((s'+1)(s'+2)-l(l+1))A_2-2\sqrt{j(j+1)}B_2=0\,,$$
$$-2\sqrt{j(j+1)}A_2 +((s'+1)(s'+2)-l(l+1)-2)B_2=0\,. \tag{50}$$

For the existence of a nontrivial solution of Eq. (50), s' must satisfy the following condition

$$(s'+1)(s'+2)=(j+1)(j+2)-\alpha^2\,. \tag{51}$$

Hence the relation between A_2 and B_2, and A_0 and B_0 are

$$A_2=\sqrt{j/j+1}B_2\,. \tag{52}$$

Equation (38) gives also the following solution:

$$A_0=-\sqrt{j+1/j}B_0\,. \tag{53}$$

Thus the final solutions of our problem are

$$f(\rho)=\rho u_2(\rho) =e^{-\rho/2}[A_0\sqrt{j+1}\,\rho^{s_-}\,{}_1F_1(-n+s_-;2s_-;\rho) +A_2\sqrt{j}\,\rho^{s_+}\,{}_1F_1(-n+s_+;2s_+;\rho)]\,, \tag{54}$$

$$g(\rho)=\rho\, v_0(\rho) =e^{-\rho/2}[A_0\sqrt{j}\,\rho^{s_-}\,{}_1F_1(-n+s_-;2s_-;\rho) +A_2\sqrt{(j+1)}\,\rho^{s_+}\,{}_1F_1(-n+s_+;2s_+;\rho)]\,, \tag{55}$$

where s_- and s_+ are the obtained from Eqs. (39) and (51), respectively. Finally the quantization condition (49) gives using Eq. (32) the following energy or mass spectrum:

$$E^2=\frac{M^2+\Delta m^2}{2}\pm\frac{M^2-\Delta m^2}{2}\left[1+\frac{\alpha^2}{n^2}\right]^{-1/2} =m_1^2+m_2^2\pm 2m_1m_2\left(1+\frac{\alpha^2}{n^2}\right)^{-1/2}. \tag{56}$$

Here the principal quantum number n is related to the radial quantum number n_r by

$$n=n_r+l_0\,, \tag{57}$$

where l_0 (the nonrelativistic label of the angular momentum) is equal to $l_0=j-1$ or $l_0=j+1$, for the two states we have discussed.

The bound states in E^2 are slightly below the continuum $E^2 > (m_1+m_2)^2$ for the $(+)$ sign in the spectrum (57) and for the $(-)$ sign, slightly above the negative continuum $E^2 < (m_1-m_2)^2$. If we expand Eq. (56) in powers of α and pass to from E^2 to E we obtain

$$E(n,l)=m_1+m_2-\frac{m_1\alpha^2}{2n^2(1+m_1/m_2)} -\frac{m_1\alpha^4}{4n^3(1+m_1/m_2)(l+\frac{1}{2})} +\frac{3}{8}\frac{m_1\alpha^4}{n^4(1+m_1/m_2)}-\frac{1}{8}\frac{(m_1^2/m_2)\alpha^4}{n^4(1+(m_1/m_2))^2} +O(\alpha^6)\,,$$
$$l=-\tfrac{1}{2}+\sqrt{(j+\tfrac{1}{2})^2-\alpha^2} \cong j-\alpha^2/(2j+1)+O(\alpha^4)\,, \tag{58}$$

which shows that the mass spectrum agrees with the usual QED up to order α^4. But the exact expression (56) should be used for recoil correction to all orders in α. Usually the nonrelativistic quantum number $n\equiv n_r+j$ is used. But it is better to keep n_r and j separately for really relativistic systems, e.g., positronium, in which l is not quite an integer. In fact one of the interesting problems of relativistic two-body dynamics is to find exact quantum numbers, besides energy and total angular momentum J, to label the states. The usual nonrelativistic labeling of positronium states, for example, as ${}^1S_0$, ${}^3S_1$,... means only that these states have these corresponding values in the nonrelativistic limit.

VI. THE SECOND SET. TOTAL SPIN $S=0$ AND $S=1$ EQUATIONS

In a similar manner we treat the second set of eight radial linear equations arising from Eq. (3). We eliminate half of the components using the algebraic equations and obtain two coupled second-order equations, namely the counter parts of Eqs. (5) and (6). The exactly soluble part of these equations, up to order α^4, are the following two uncoupled equations:

$$\left[\frac{1}{4}\left(E-\frac{M^2}{E}\right)\left(E-\frac{\Delta m^2}{E}\right)+\frac{\alpha}{r}\left(E-\frac{M^2+\Delta m^2}{2E}\right) - \frac{j(j+1)-\alpha^2}{r^2}+\partial_r^2\right](ru_1)=0, \tag{59a}$$

$$\left[\frac{1}{4}\left(E-\frac{M^2}{E}\right)\left(E-\frac{\Delta m^2}{E}\right)+\frac{\alpha}{2r}\left(2E-\frac{M^2+\Delta m^2}{E}\right) - \frac{j(j+1)-\alpha^2\delta_E}{r^2}+\partial_r^2\right](rv_{00})=0. \tag{59b}$$

The free-particle solution of (59a) and (59b) are simpler than in the first set, Eqs. (26) and (27), namely

$$\rho u_2(\rho)=A\rho j_j(\rho),\qquad \rho v_{00}(\rho)=B\rho j_j(\rho). \tag{60}$$

In Eq. (59b) the factor δ_E is given by

$$\delta_E=\frac{\Delta m^2-M^2}{E^2}+\frac{M^2\Delta m^2}{E^4}. \tag{61}$$

The solutions of (59a) and (59b), because they are uncoupled, can be written down immediately in terms of hydrogenic wave functions

$$\rho u_1(\rho)=Ae^{-\rho/2}\rho^{l+1}\,{}_1F_1(-n+l+1;2l+2;\rho), \tag{62a}$$

$$\rho v_{00}(\rho)=Be^{-\rho/2}\rho^{l_0+1}\,{}_1F_1(-n+l_0+1;2l_0+2;\rho), \tag{62b}$$

and the spectrum has the same general form as in Eq. (56),

$$E^2=\frac{M^2+\Delta m}{2}\pm\frac{M^2-\Delta m^2}{2}\left(1+\frac{\alpha^2}{(n_r+l)^2}\right)^{-1/2}. \tag{63}$$

But the range of the angular momentum l is now given by

$$l(l+1)=j(j+1)-\alpha^2,\quad \text{for Eq. (59a)} \tag{64}$$

and

$$l(l+1)=j(j+1)-\alpha^2\delta_E,\quad \text{for Eq. (59b)}. \tag{65}$$

VII. COMPARISON WITH THE EXACTLY SOLUBLE INFINITE-COMPONENT WAVE EQUATIONS

Infinite-component wave equations appropriate for two-body Coulomb systems are generalizations of the original infinite component Majorana equation.[4] They make use of the dynamical group SO(4,2) rather than the Lorentz group of the Majorana equation and account for the correct degeneracy of states. They have been used to describe the relativistic H-atom and hadrons, and to describe many properties of these composite systems in a relativistic way, such as form factors and transition amplitudes in external fields.[5] It is interesting that our exactly soluble models give precisely the same spectrum as the infinite-composite wave equation for the relativistic Coulomb problem. We thus have in the one hand the group structure of our model, and on the other hand, the infinite-component wave equation acquires an explicit dynamical realization in terms of constituents.

The wave equation is a generalized Dirac equation

$$(J^\mu P_\mu+K)\psi(P)=0, \tag{66}$$

where P_μ is the total momentum of the composite system, and the current and mass operators are given by

$$J_\mu=\alpha_1\Gamma_\mu+\alpha_2P_\mu+\alpha_3P_\mu\Gamma_4,\qquad K=\beta\Gamma_4+\gamma. \tag{67}$$

Here Γ_μ and Γ_4 are the generators of the dynamical group SO(4,2); P_μ the total momentum of the atom. The choice of the constants[6]

$$\alpha_1=1,\quad \alpha_2=\alpha/2m_2,\quad \alpha_3=1/2m_2,\qquad \beta=(m_2^2-m_1^2)/2m_2,\quad \gamma=-\alpha(m_1^2+m_2^2)/2m \tag{68}$$

gives the spectrum

$$M_n^{\pm^2}=m_1^2+m_2^2\pm2m_1m_2(1+\alpha^2/n^2)^{-1/2} \tag{69}$$

which coincides with (56) or (63).

In fact the form of infinite-component wave equation (66) can be inferred directly from our basic equation (1), but the operators J_μ and K have a more complicated form for Eq. (1); the simpler forms given in (67) and (68) correspond to the exactly soluble part of our equation. The infinite-component equation is very useful in treating further the external interactions of our composite atom because it treats the whole atom now as a single relativistic "particle."

The discussion of the perturbations of order α^5 is given elsewhere.[7]

ACKNOWLEDGMENTS

We thank Professor H. Salecker for his hospitality at the University of München.

This work was supported in part by Alexander von Humboldt Foundation, by a NATO Research Grant, by National Science Foundation-International SDC-programs, and by the Scientific and Technical Research Council of Turkey.

[1] A. O. Barut, in *Proceedings of the* XIth *International Colloquium on Group Theoretical Methods in Physics, Lecture Notes in Physics*, Vol. 180 (Springer, Berlin, 1983), p. 332.
[2] A. O. Barut and S. Komy, Fortschr. Phys. **33**, 309 (1985).
[3] A. O. Barut and N. Ünal, Fortschr. Phys. **33**, 319 (1985).
[4] A. O. Barut, *Dynamical Groups and Generalized Symmetries in Quantum Theory* (Univ. Canterbury P., Canterbury, 1972).
[5] A. O. Barut and R. Wilson, Phys. Rev. D **13**, 2629 (1976); Phys. Rev. A **13**, 918 (1976).
[6] A. O. Barut and W. Rasmussen, J. Phys. B **6**, 1695, 1713 (1973).
[7] A. O. Barut and N. Ünal, Physica (to be published).

First-order conserved densities for gas dynamics

John M. Verosky
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

(Received 5 May 1986; accepted for publication 6 August 1986)

For a certain choice of pressure the equations of gas dynamics can be diagonalized, making it easy to calculate some first-order conserved densities of the system. Using the Hamiltonian structure of these equations, third-order symmetries are seen to exist by the Noether correspondence. Such higher-order phenomena are usually associated with a linearization, but in this case, no linearization is obvious.

I. INTRODUCTION

The equations of gas dynamics are

$$u_t + uu_x + (1/\rho)P_x = 0,$$

$$\rho_t + \rho u_x + u\rho_x = 0, \quad s_t + us_x = 0,$$

where u,ρ,s are the velocity, density, and entropy, respectively, of the gas. Here P is the pressure and is a function of ρ and s. When the pressure is chosen to be

$$P = -1/\rho + s,$$

it is shown that the above system has first-order conserved densities, that is, functions

$$T(u,\rho,s,u_x,\rho_x,s_x)$$

such that

$$\int T\,dx$$

is independent of time for suitable solutions. Working in the spirit of the formal variational calculus, T is such a function if there is an X such that

$$D_t T = D_x X,$$

where D_t and D_x are total t and x derivatives, respectively.

A symmetry for gas dynamics would be a system of equations

$$u_r = U, \quad \rho_r = R, \quad s_r = S,$$

where U,R,S depend on x,t,u,ρ,s and derivatives of u,ρ,s. The symmetry is higher order if it depends on derivatives of order 2 or more, and conserved densities with derivatives of order 1 or more are also called higher order. Finding higher-order symmetries and conserved densities has been a recent theme in the study of nonlinear partial differential equations ever since the astounding success of these techniques for dealing with the "integrable" evolution equations, such as the KdV equation.[1–5] Using the theory of noncanonical Hamiltonian structures[6–9] for nonlinear partial differential equations (which also has as its origin the integrable equations), the higher-order conserved density leads to a higher-order symmetry by an application of Noether's theorem.

In an earlier paper,[10] similar work was done for the isentropic ($s =$ const) equations with pressure ρ^γ. There, a conserved density

$$T = \rho_x/(u_x^2 - \gamma\rho^{\gamma-3}\rho_x^2)$$

was found. Note that there are derivatives in the denominator, which vanishes for simple wave solutions (where ρ and u are functionally dependent). In fact, T may be said to measure the "simpleness" of the solution. The conserved density found in this paper has a similar property.

The isentropic equations are linearizable by the hodograph transformation, and the existence of higher-order phenomena may be blamed on this, because the linear equations have recursion operators that produce new symmetries, possibly of higher order, from given ones. In Ref. 10, some recursion operators were found that guarantee the existence of each positive integer order. The integrable systems are also linearizable by the inverse scattering transform and some people believe that higher-order phenomena are always associated with the ability to linearize. The system considered here, however, has no obvious (to this author) linearization.

The plan of the paper is as follows. In Sec. II, it is shown how the equations of gas dynamics, for this particular pressure, may be diagonalized. Only for pressures of the form $F(-\rho^{-1} + g(s))$ can the diagonalization be carried out, but only when F is the identity function can a first-order conserved density be found. Of course, there may be pressures for which the equations may not be diagonalized but which have first-order conserved densities. In this case the calculations become extremely long or impossible and the existence of higher-order phenomena is an open question. In Sec. III conditions are given for the existence of certain first-order conserved densities of diagonal quasilinear evolution equations in three dependent variables. In Sec. IV, the conditions are verified for the system obtained in Sec. II. Finally, in Sec. V, the Hamiltonian structure for gas dynamics in general is discussed and Noether's theorem is applied to obtain a third-order symmetry.

II. DIAGONALIZATION OF GAS DYNAMICS EQUATIONS

Replace t by $-t$ to write the equations as

$$u_t = uu_x + (1/\rho^2)\rho_x + (1/\rho)s_x,$$

$$\rho_t = \rho u_x + u\rho_x, \quad s_t = us_x.$$

We will make a change of dependent variables: let a,b,c be functions of u,ρ,s. In fact if

$$a = u - 1/\rho + s, \quad b = u + 1/\rho - s, \quad c = s,$$

then the system becomes

$$a_t = (b + c)a_x, \quad b_t = (a - c)b_x,$$

$$c_t = [(a + b)/2]c_x.$$

 0022-2488/86/123061-03$02.50

The following occurs:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}_t = \begin{pmatrix} a_u & a_\rho & a_s \\ b_u & b_\rho & b_s \\ c_u & c_\rho & c_s \end{pmatrix} \begin{pmatrix} u \\ \rho \\ s \end{pmatrix}_t = \begin{pmatrix} a_u & a_\rho & a_s \\ b_u & b_\rho & b_s \\ c_u & c_\rho & c_s \end{pmatrix} \begin{pmatrix} u & \rho^{-3} & \rho^{-1} \\ \rho & u & 0 \\ 0 & 0 & u \end{pmatrix} \begin{pmatrix} u \\ \rho \\ s \end{pmatrix}_x .$$

But it is also true that

$$\begin{pmatrix} u \\ \rho \\ s \end{pmatrix}_x = \begin{pmatrix} u_a & u_b & u_c \\ \rho_a & \rho_b & \rho_c \\ s_a & s_b & s_c \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}_x = \begin{pmatrix} a_u & a_\rho & a_s \\ b_u & b_\rho & b_s \\ c_u & c_\rho & c_s \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}_x .$$

Thus, a,b,c evolve according to

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}_t = M \begin{pmatrix} a \\ b \\ c \end{pmatrix}_x ,$$

where M is

$$\begin{aligned}
&\begin{pmatrix} a_u & a_\rho & a_s \\ b_u & b_\rho & b_s \\ c_u & c_\rho & c_s \end{pmatrix} \begin{pmatrix} u & \rho^{-3} & \rho^{-1} \\ \rho & u & 0 \\ 0 & 0 & u \end{pmatrix} \begin{pmatrix} a_u & a_\rho & a_s \\ b_u & b_\rho & b_s \\ c_u & c_\rho & c_s \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 1 & \rho^{-2} & 1 \\ 1 & -\rho^{-2} & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u & \rho^{-3} & \rho^{-1} \\ \rho & u & 0 \\ 0 & 0 & u \end{pmatrix} \begin{pmatrix} 1 & \rho^{-2} & 1 \\ 1 & -\rho^{-2} & -1 \\ 0 & 0 & -1 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 1 & \rho^{-2} & 1 \\ 1 & -\rho^{-2} & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u & \rho^{-3} & \rho^{-1} \\ \rho & u & 0 \\ 0 & 0 & u \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \rho^2/2 & -\rho^2/2 & -\rho^2 \\ 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} u+1/\rho & 0 & 0 \\ 0 & u-1/\rho & 0 \\ 0 & 0 & u \end{pmatrix} = \begin{pmatrix} b+c & 0 & 0 \\ 0 & a-c & 0 \\ 0 & 0 & (a+b)/2 \end{pmatrix} .
\end{aligned}$$

III. CONSERVED DENSITIES FOR A DIAGONAL SYSTEM

A system of the form

$$a_t = A a_x, \quad b_t = B b_x, \quad c_t = C c_x$$

(where A,B,C are functions of a,b,c) has a conserved density of the form

$$T = \frac{f}{a_x} + \frac{g}{b_x} + \frac{h}{c_x}$$

(where f,g,h are functions of a,b,c) with flux

$$X = \frac{Af}{a_x} + \frac{Bg}{b_x} + \frac{Cc}{c_x}$$

when A,B,C,f,g,h satisfy certain first-order partial differential equations. These are found by expanding $D_t T = D_x X$ and setting the coefficients of

$$\frac{a_x}{b_x}, \frac{a_x}{c_x}, \frac{b_x}{a_x}, \frac{b_x}{c_x}, \frac{c_x}{a_x}, \frac{c_x}{b_x}$$

to zero. There is also one remaining equation. Thus there are seven equations

$$\frac{f_b}{f} = \frac{2A_b}{B-A}, \quad \frac{f_c}{f} = \frac{2A_c}{C-A},$$

$$\frac{g_a}{g} = \frac{2B_a}{A-B}, \quad \frac{g_c}{g} = \frac{2B_c}{C-B},$$

$$\frac{h_a}{h} = \frac{2C_a}{A-C}, \quad \frac{h_b}{h} = \frac{2C_b}{B-C},$$

and

$$fA_a + gB_b + hC_c = 0.$$

It should be noted that obtaining these equations is relatively easy because the system is diagonal. Solving $D_t T = D_x X$ explicitly for the first-order conserved densities of an arbitrary quasilinear first-order system is usually impossible.

IV. THE FIRST-ORDER CONSERVED DENSITIES

Consider the choices

$$A = b + c, \quad B = a - c, \quad C = (a+b)/2.$$

Note that the seventh equation disappears since $A_a = B_b = C_c = 0$. The other six are readily solved, giving

$$f = \alpha(a)(a-b-2c)^{-2},$$

$$g = \beta(b)(a-b-2c)^{-2},$$

$$h = \gamma(c)(a-b-2c)^{-2},$$

where α,β,γ are arbitrary functions of one variable. Note that since

$$a - b - 2c = -2/\rho,$$

f,g,h can also be expressed as

$f=\alpha(a)\rho^2, \quad g=\beta(b)\rho^2, \quad h=\gamma(c)\rho^2.$

Hence the first-order conserved densities are of the form

$$T=\rho^2\left(\frac{\alpha(a)}{a_x}+\frac{\beta(b)}{b_x}+\frac{\gamma(c)}{c_x}\right)=\frac{\text{numerator}}{a_x b_x c_x}.$$

The denominator of T vanishes only when two of a,b,c are functionally dependent. For example, if b and c were dependent then

$$0=b_t c_x - b_x c_t = Bb_x c_x - b_x C c_x = (B-C)b_x c_x$$

would imply that $a_x b_x c_x = a_x \cdot 0 = 0$. Thus, these conserved densities measure the degree of dependence among a,b,c.

It is easy to write T completely in terms of u,ρ,s using

$$a_x = u_x + \rho^{-2}\rho_x + s_x,$$
$$b_x = u_x - \rho^{-2}\rho_x - s_x, \quad c_x = s_x.$$

Two special cases are interesting. For $\alpha = 1, \beta = -1$, and $\gamma = 0$, T is

$$\rho^2\left(\frac{b_x - a_x}{a_x b_x}\right) = \rho^2 \frac{-2\rho^{-2}\rho_x}{u_x^2 - (\rho^{-2}\rho_x + s_x)^2}$$
$$= -2\left(\frac{\rho_x}{u_x^2 - (\rho^{-2}\rho_x + s_x)^2}\right).$$

The -2 is irrelevant. Except for the minus sign in the denominator, this agrees with the conserved density in Ref. 10 with $\gamma = -1$, in the pressure ρ^γ, when entropy is constant. In this paper, pressure is $-\rho^{-1}+s$, and the difference of signs in the pressures leads to different signs in the denominators of the conserved densities. If $\alpha = 1 = \beta$ and $\gamma = 0$, then T becomes

$$\rho^2\left(\frac{a_x + b_x}{a_x b_x}\right) = 2\rho^2\left(\frac{u_x}{u_x^2 - (\rho^{-2}\rho_x + s_x)^2}\right),$$

which has no analog in Ref. 10. However, $\gamma = -1$ is actually a singular value for γ and the isentropic equations have more higher-order symmetries and conserved densities for $\gamma = -1$ than for general γ.

V. HAMILTONIAN STRUCTURE

It can be verified using Olver's methods[11] that

$$\begin{pmatrix} u \\ \rho \\ s \end{pmatrix}_t = \begin{pmatrix} 0 & D_x & -s_x/\rho \\ D_x & 0 & 0 \\ s_x/\rho & 0 & 0 \end{pmatrix} \begin{pmatrix} \partial_u \\ \partial_\rho \\ \partial_s \end{pmatrix} \left(\tfrac{1}{2}\rho u^2 + \epsilon(\rho,s)\right)$$

provides the gas dynamics equations with a nonlinear Hamiltonian structure. The internal energy is $\epsilon(\rho,s)$ and the kinetic energy is $\frac{1}{2}\rho u^2$, hence the Hamiltonian function is the total energy. Expanding the above gives

$$u_t = uu_x + \epsilon_{\rho\rho}\rho_x + (\epsilon_{\rho s} - (1/\rho)\epsilon_s)s_x,$$
$$\rho_t = \rho u_x + u\rho_x, \quad s_t = us_x.$$

This shows the one-to-one correspondence between ϵ and P given by

$$\epsilon_{\rho\rho} = (1/\rho)P_\rho, \quad \rho\epsilon_{\rho s} - \epsilon_s = P_s.$$

These equations are compatible for any choices of ϵ or P since

$$(\rho\epsilon_{\rho\rho})_s = (\rho\epsilon_{\rho s} - \epsilon_s)_\rho,$$

for all ϵ. For the pressure used in this paper, the corresponding ϵ is $1/2\rho - s$. This is a potential energy term possibly arising from electrical forces among the gas particles.

Noether's theorem says that

$$\begin{pmatrix} u \\ \rho \\ s \end{pmatrix}_r = \begin{pmatrix} 0 & D & -s_x/\rho \\ D & 0 & 0 \\ s_x/\rho & 0 & 0 \end{pmatrix} \begin{pmatrix} E_u \\ E_\rho \\ E_s \end{pmatrix} T$$

is a symmetry, whenever T is a conserved density. Here, the E's are Euler operators or variational derivatives when T is first order or more. The symmetry can be expressed in the a,b,c coordinates by making a change of variables:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}_r = \begin{pmatrix} 0 & D & -s_x/\rho \\ D & 0 & 0 \\ s_x/\rho & 0 & 0 \end{pmatrix} \begin{pmatrix} E_a \\ E_b \\ E_c \end{pmatrix} T,$$

where J is the Jacobian matrix

$$\begin{pmatrix} a_u & a_\rho & a_s \\ b_u & b_\rho & b_s \\ c_u & c_\rho & c_s \end{pmatrix}.$$

Completely expanding the above is an absurdly long calculation but it is simple to obtain the highest- (third-) order terms:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}_r = \begin{pmatrix} \alpha(a) & a_{xxx}/a_x^3 \\ \beta(b) & b_{xxx}/b_x^3 \\ & 0 \end{pmatrix} + \text{lower-order terms}.$$

Thus the gas dynamic equations have plenty of third-order symmetries.

ACKNOWLEDGMENT

I would like to thank Peter Olver for several helpful comments.

[1] N. Kh. Ibragimov and A. B. Shabat, Funct. Anal. Appl. **14**, 19 (1980).
[2] N. Kh. Ibragimov and A. B. Shabat, Funct. Anal. Appl. **14**, 313 (1980).
[3] R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).
[4] P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[5] G. Wilson, Math. Proc. Cambridge Philos. Soc. **86**, 131 (1979).
[6] C. S. Gardner, J. Math. Phys. **12**, 1548 (1971).
[7] I. M. Gelfand and I. Ya. Dorfman, Funct. Anal. Appl. **13** 248 (1979).
[8] P. J. Olver, Math. Proc. Cambridge Philos. Soc. **88**, 71 (1980).
[9] F. Magri, J. Math. Phys. **19**, 1156 (1978).
[10] J. Verosky, J. Math. Phys. **25**, 884 (1984).
[11] P. J. Olver, *Applications of Lie Groups to Differential Equations, Graduate Texts in Mathematics*, Vol. 107 (Springer, Berlin, 1986).

Mathematical aspects of quantum fluids. III. Interior Clebsch representations and transformations of symplectic two-cocycles for ^{4}He

B. A. Kupershmidt
The University of Tennessee Space Institute, Tullahoma, Tennessee 37388 and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 1 May 1986; accepted for publication 20 August 1986)

The symplectic two-cocycle on the semidirect product Lie algebra $\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$ is shown to be canonically related to the dual spaces of the Lie algebras (a) $\mathfrak{g}\ltimes(W\oplus(\mathfrak{g}\ltimes V))$ and (b) $\mathfrak{g}\ltimes(W\oplus(\mathfrak{g}\ltimes V^*))$. This fact (a) explains the second Poisson bracket for irrotational ^{4}He and (b) leads to a derivation of a new nonlinear Poisson bracket for rotating ^{4}He.

I. INTRODUCTION

The dynamics of the majority of conservative continuous systems is governed by the elementary Hamiltonian formalism associated with semidirect product Lie algebras invariably of the form $D_n\ltimes V$, where D_n is the Lie algebra of vector fields on $\mathbb{R}^n$ (n is the dimension of the physical space), and V is a subspace in the space of tensor fields on $\mathbb{R}^n$. The only known exceptions to this experimental principle are quantum fluids in general,[1] and four types of superfluid helium in particular.[2,3] The new features appearing in the Hamiltonian description of these exceptional systems are (A) a symplectic form on V, which is a two-cocycle on $D_n\ltimes V$, in the cases of irrotational ^{4}He and rotating ^{4}He, and (B) instead of a subspace V in $D_n\ltimes V$, one has a non-Abelian Lie algebra L in $D_n\ltimes L$, in the cases of spinless ^{3}He–A, ^{3}He–A with spin, and general quantum fluids.

A systematic Lie algebraic analysis of various mathematical phenomena involved in the Hamiltonian description of quantum fluids is the purpose of the series of papers of which this one is the third. In the first two papers I examined symplectic two-cocycles on semidirect product Lie algebras,[4] and described generalized Clebsch representations for them.[5] Here I study the last unexplained observation in the description of irrotational ^{4}He [referred to in the feature (A) above]: the transformation from the symplectic two-cocycle description in the space with the condensate phase variable, into the cocycleless description in the space with the superfluid momentum density variables.

To be more specific, let us start with the *first* Poisson bracket formula for irrotational ^{4}He [formula (5) in Ref. 3]:

$$\{H,F\}\sim\Big\{\frac{\delta F}{\delta M_k}\Big[(M_l\partial_k+\partial_l M_k)\Big(\frac{\delta H}{\delta M_l}\Big)+\rho\partial_k\Big(\frac{\delta H}{\delta\rho}\Big)-\alpha_{,k}\frac{\delta H}{\delta\alpha}+\sigma\partial_k\Big(\frac{\delta H}{\delta\sigma}\Big)\Big]+\Big[\frac{\delta F}{\delta\rho}\partial_l\rho+\frac{\delta F}{\delta\alpha}\alpha_{,l}+\frac{\delta F}{\delta\sigma}\partial_l\sigma\Big]\Big(\frac{\delta H}{\delta M_l}\Big)\Big\} \tag{1.1a}$$

$$+\Big(\frac{\delta F}{\delta\alpha}\frac{\delta H}{\delta\rho}-\frac{\delta F}{\delta\rho}\frac{\delta H}{\delta\alpha}\Big). \tag{1.1b}$$

The notation used here is $\partial_l=\partial/\partial x_l$, where $(x_1,\dots,x_n)$ are coordinates in $\mathbb{R}^n$ ($n\leqslant 3$ as a rule); $(\cdot)_{,l}=\partial(\cdot)/\partial x_l, 1\leqslant k,l\leqslant n$, and the sum is taken over repeated indices; $\mathbf{M}=(M_1,\dots,M_n)$ is the total momentum density of the normal flow; ρ is the mass density; σ is the entropy density; α is the condensate phase that defines the curl-free superfluid velocity $\mathbf{v}^s$ as $\mathbf{v}^s=\nabla\alpha$; $\delta H/\delta(\cdot)$ denotes the variational derivative of H with respect to $(\cdot)$; $\sim$ means equality modulo total derivatives ("divergences").

The part (1.1a) of the Poisson bracket (1.1) is the natural bracket associated to the dual space of the semidirect product Lie algebra

$$\mathfrak{g}(^4\mathrm{He}_{nr})=D_n\ltimes(\Lambda^0\oplus\Lambda^n\oplus\Lambda^0), \tag{1.2}$$

with the commutator

$$[(X;f;\beta;a),\ (\bar X;\bar f;\bar\beta;\bar a)]=([X,\bar X];X(\bar f)-\bar X(f);X(\bar\beta)-\bar X(\beta);X(\bar a)-\bar X(a)), \tag{1.3}$$

where: $\Lambda^k=\Lambda^k(\mathbb{R}^n)$ is the $C^\infty(\mathbb{R}^n)$-module of differential k-forms on $\mathbb{R}^n$; $X,\bar X\in D_n$; $f,a,\bar f,\bar a,\in\Lambda^0$; $\beta,\bar\beta\in\Lambda^n$; the (Lie derivative) action of D_n on Λ^k is denoted $X(\cdot)$ for $X\in D_n$ and $(\cdot)\in\Lambda^k$; the dual coordinates on $(\mathfrak{g}(^4\mathrm{He}_{nr}))^*$ are M_k to $\partial_k\in D_n$, ρ to $1\in\Lambda^0$, α to $dx_1\wedge\cdots\wedge dx_n\in\Lambda^n$, σ to $1\in\Lambda^0$.

The part (1.1b) of the Poisson bracket (1.1) corresponds to the following two-cocycle on the Lie algebra $\mathfrak{g}(^4\mathrm{He}_{nr})$ (1.2):

$$\omega((X;f;\beta;a),(\bar X;\bar f;\bar\beta;\bar a))=-f\bar\beta+\beta\bar f. \tag{1.4}$$

The *second* Poisson bracket formula for irrotational ^{4}He [equivalent to formula (9) in Ref. 3] is

$$\{H,F\}\sim\frac{\delta F}{\delta M_K}\Big[(M_l\partial_k+\partial_l M_k)\Big(\frac{\delta H}{\delta M_l}\Big)+\rho\partial_k\Big(\frac{\delta H}{\delta\rho}\Big)+(P_l\partial_k+\partial_l P_k)\Big(\frac{\delta H}{\delta P_l}\Big)+\sigma\partial_k\Big(\frac{\delta H}{\delta\sigma}\Big)\Big]+\Big[\frac{\delta F}{\delta\rho}\partial_l\rho+\frac{\delta F}{\delta P_k}(P_l\partial_k+\partial_l P_k)+\frac{\delta F}{\delta\sigma}\partial_l\sigma\Big]\times\Big(\frac{\delta H}{\delta M_l}\Big)+\frac{\delta F}{\delta P_k}\Big[(P_l\partial_k+\partial_l P_k)\Big(\frac{\delta H}{\delta P_l}\Big)+\rho\partial_k\Big(\frac{\delta H}{\delta\rho}\Big)\Big]+\frac{\delta F}{\delta\rho}\partial_l\rho\Big(\frac{\delta H}{\delta P_l}\Big), \tag{1.5}$$

with $\mathbf{P}=(P_1,\dots,P_n)$ being superfluid momentum density:

$$P_k=\rho\alpha_{,k},\quad 1\leqslant k\leqslant n. \tag{1.6}$$

The Poisson bracket (1.5) is the natural bracket associated to the dual space of the semidirect product Lie algebra

$$\mathfrak{g}(^4\mathrm{He}_{nr})' = D_n \ltimes (\Lambda^0 \oplus (D_n \ltimes \Lambda^0)), \tag{1.7}$$

with the commutator

$$[(X;a;Y;f),(\bar{X};\bar{a};\bar{Y};\bar{f})]$$
$$= ([X,\bar{X}];X(\bar{a}) - \bar{X}(a);$$
$$[X,\bar{Y}] - [\bar{X},Y] + [Y,\bar{Y}];$$
$$(X+Y)(\bar{f}) - (\bar{X}+\bar{Y})(f)), \tag{1.8}$$

where $X,\bar{X}\in D_n$; $a,\bar{a}\in\Lambda^0$; $Y,\bar{Y}\in D_n$; $f,\bar{f}\in\Lambda^0$; and dual coordinates on $[\mathfrak{g}(^4\mathrm{He}_{nr})']^*$ are M_k to $\partial_k\in D_n$, σ to $1\in\Lambda^0$, P_k to $\partial_k\in D_n$, ρ to $1\in\Lambda^0$.

It is easy to check out that the first Poisson bracket (1.1) and the second Poisson bracket (1.5), describing the same irrotational ^{4}He, are compatible with respect to the relation (1.6). I can now formulate precisely the problem addressed in this paper: What is the nature of the map (1.6) and of the Lie algebra (1.7), and why does this map produce this Lie algebra out of the Lie algebra (1.2) together with the symplectic two-cocycle (1.4) on it? The answer, given by Theorem 3.3, asserts, roughly speaking, that if $\mathfrak{g}$ is a Lie algebra acting on spaces W and V, then the natural Hamiltonian map between the symplectic space $V^*\oplus V$ and the dual to the Lie algebra $\mathfrak{g}\ltimes V$ can be extended into a Hamiltonian map between the dual to the Lie algebra $\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$ together with the symplectic two-cocycle on it, and the dual to the Lie algebra $\mathfrak{g}\ltimes(W\oplus(\mathfrak{g}\ltimes V))$, with $\mathfrak{g}$ acting on itself in the adjoint representation. This result, incidentally, provides the first general class of Clebsch representations (= Hamiltonian maps) for semidirect products $\mathfrak{g}\ltimes L$ with *non-Abelian* L.

The paper is organized as follows: To make the presentation reasonably self-contained, for the reader's convenience I summarize in the next section the basic ingredients of the modern Hamiltonian formalism. (Details can be found in Ref. 6, Chap. VIII.) In Sec. III, the main result of this paper is proved (Theorem 3.3), in the spirit of paper II in this series.[5] Section IV is devoted to applications. First, formulas (1.6) and (1.7) for irrotational ^{4}He are shown to be a particular instance of Theorem 3.3. Then we analyze the case of rotating ^{4}He. This analysis shows that one needs a complementary version of Theorem 3.3, and such a version is then established (Theorem 4.1). Applying Theorem 4.1 to the case of rotating ^{4}He, we find for this case a new nonlinear Poisson bracket [(4.27)] in the space of the physical variables.

II. HAMILTONIAN FORMALISM

Let K be a commutative algebra. Let $\partial_1,\ldots,\partial_n: K\to K$ be n commuting derivations. Let G be a discrete group acting by automorphisms on K, and suppose that the actions of G and ∂'s commute. Such K is called a differential-difference ring. Let I be a countable set. Set $C = K[q_i^{(g|\nu)}]$, $i\in I$, $g\in G$, $\nu\in \mathbf{Z}^n_+$, and extend G and ∂'s to act on C by the rule

$$\hat{h}(q_i^{(g|\nu)}) = q_i^{(hg|\nu)},$$
$$\partial^\mu(q_i^{(g|\nu)}) = q_i^{(g|\mu+\nu)}, \quad h\in G, \quad \mu\in\mathbf{Z}^n_+, \tag{2.1}$$

where

$$(\pm\partial)^\mu = (\pm\partial_1)^{\mu_1}\cdots(\pm\partial_n)^{\mu_n}$$

for $\mu = (\mu_1,\ldots,\mu_n)$, and $\hat{h}(\cdot)$ is the image of $(\cdot)$ under the automorphism h, $h\in G$.

Let N be a natural number or ∞, T a differential-difference ring. Then T^N consists of column vectors with only finite number of nonzero components. An operator E: $T^N\to T^M$ is a map of the form

$$(E(u))_j = \sum E^j_{i,g,\mu}\hat{g}\partial^\mu(u_i), \quad E^j_{\ldots}\in T, \quad u\in T^N, \tag{2.2}$$

finite sums; a bilinear operator $T^{N_1}\times T^{N_2}\to T^{N_3}$ is defined analogously. An algebra structure on T^N is a bilinear operator $T^N\times T^N\to T^N$. The associative ring of operators $T^N\to T^N$, and the corresponding Lie algebra, are both denoted $\mathrm{Diff}(T^N)$.

Trivial elements in T are defined as elements from

$$\mathrm{Im}\,\mathscr{D} = \sum_{s=1}^n \mathrm{Im}\,\partial_s + \sum_{g\in G}\mathrm{Im}(\hat{g}-\hat{e}),$$

where e is the unit element of G; we write $a\sim b$ if $(a-b)$ is trivial.

A bilinear form on T^N is an operator ω: $T^N\times T^N\to T$. To each bilinear form ω one uniquely associates an operator b_ω: $T^N\to T^N$, acting by the rule

$$\omega(X,Y)\sim X^t b_\omega(Y), \tag{2.3}$$

so that if

$$\omega(X,Y) = \sum \omega_{i,g,\mu|j,h,\nu}\hat{g}\partial^\mu(X_i)\cdot\hat{h}\partial^\nu(Y_j), \quad \omega_{\ldots}\in T,$$

then

$$(b_\omega)_{ij} = \hat{g}^{-1}(-\partial)^\mu\omega_{i,g,\mu|j,h,\nu}\hat{h}\,\partial^\nu. \tag{2.4}$$

The form ω is called symmetric (resp. skew symmetric), if $\omega(X,Y)\sim\omega(Y,X)$ [resp. $\omega(X,Y)\sim-\omega(Y,X)$]. The form ω is symmetric (resp. skew symmetric) if and only if the corresponding operator b_ω is symmetric: $(b_\omega)^\dagger = b_\omega$ [resp. b_ω is skew symmetric: $(b_\omega)^\dagger = -b_\omega$]. Recall that for an operator E: $T^N\to T^M$, the adjoint operator $E^\dagger$:$T^M\to T^N$ is uniquely defined by the equation

$$v^tE(u)\sim[E^\dagger(v)]^t u, \quad u\in T^N, \quad v\in T^M, \tag{2.5}$$

so that

$$(E^\dagger)_{ij} = (E_{ji})^\dagger, \tag{2.6}$$

and

$$(a\hat{g}\,\partial^\nu)^\dagger = (-\partial)^\nu\,\hat{g}^{-1}a, \quad a\in T. \tag{2.7}$$

A Lie algebra structure on K^N is an operator $K^N\times K^N\to K^N$,[,]:$X\times Y\mapsto[X,Y]$, satisfying the following conditions:

$$[X,Y] = -[Y,X] \quad \text{(skew symmetry)}; \tag{2.8}$$
$$[X,[Y,Z]] + \text{c.p.} = 0 \quad \text{(Jacobi identity)}, \tag{2.9}$$

where "c.p." stands for "cyclic permutation";

the properties (2.8) and (2.9) remain true under any (differential-difference) extension $K'\supset K$. (2.10)

A skew-symmetric form ω on a Lie algebra $\mathfrak{g}=K^N$ is called a (generalized) two-cocycle on $\mathfrak{g}$ if

$$\omega(X,[Y,Z])+\text{c.p.}\sim 0,\quad \forall X,Y,Z\in\mathfrak{g}. \tag{2.11}$$

A derivation X of C over K is called evolutionary if it commutes with the actions of G and ∂'s, so that

$$X=\sum \hat{g}\partial^{\nu}(X_i)\frac{\partial}{\partial q_i^{(g|\nu)}},\quad X_i:=X(q_i),\quad q_i:=q_i^{(e|0)}. \tag{2.12}$$

The set of all evolution derivations is a Lie algebra denoted $D^{\text{ev}}(C)$.

Set $N=|I|$. The Euler–Lagrange map $\delta=\delta/\delta\bar{q}$: $C\to C^N$, defined by the formula

$$\left(\frac{\delta H}{\delta\bar{q}}\right)_i=\frac{\delta H}{\delta q_i}=\sum\hat{g}^{-1}(-\partial)^{\nu}\left(\frac{\partial H}{\partial q_i^{(g|\nu)}}\right), \tag{2.13}$$

annihilates $\operatorname{Im}\mathscr{D}$ in C: $\delta(\operatorname{Im}\mathscr{D})=0$, while $\delta H/\delta q_i$ is called the variational derivative of H with respect to q_i. For $X\in D^{\text{ev}}(C)$, $H\in C$,

$$X(H)\sim\bar{X}^t\frac{\delta H}{\delta\bar{q}} \tag{2.14}$$

("formula for the first variation"), where "t" stands for "transpose," and

$$(\bar{X})_i=X_i. \tag{2.15}$$

A map $\Gamma: C\to D^{\text{ev}}(C)$, $H\mapsto X_H$, is called Hamiltonian if there exists an operator B: $C^N\to C^N$ such that

$$\overline{X_H}=B\left(\frac{\delta H}{\delta\bar{q}}\right); \tag{2.16}$$

$$\{H,F\}\sim-\{F,H\}\quad\text{(skew symmetry)}, \tag{2.17}$$

where the Poisson bracket $\{H,F\}$ is defined as $X_H(F)$;

$$X_{\{H,F\}}=[X_H,X_F]; \tag{2.18}$$

Properties (2.17) and (2.18) remain true for arbitrary extension $K'\supset K$.

The property (2.18) is equivalent to

$$\{H,\{F,S\}\}+\text{c.p.}\sim 0, \tag{2.18$'$}$$

for any $H,F,S\in C'=K'[q_i^{(g|\nu)}]$. The property (2.17) is equivalent to B being skew symmetric: $B^{\dagger}=-B$, while (2.18) can be reduced to a set of quadratic equations on the matrix elements of B.

Let $C_1=K[p_j^{(g|\nu)}]$, $j\in J$, $g\in G$, $\nu\in Z_+^n$. A (differential-difference) homomorphism Φ: $C\to C_1$ is a homomorphism over K that commutes with the actions of G and ∂'s:

$$\Phi(q_i^{(g|\nu)})=\hat{g}\,\partial^{\nu}(\Phi_i),\quad \Phi_i:=\Phi(q_i). \tag{2.19}$$

If Γ_1: $C_1\to D^{\text{ev}}(C_1)$, $F\mapsto X_F$ is a Hamiltonian structure in the ring C_1, then the map Φ is called Hamiltonian (also: "canonical") if, for any $H\in C$, the evolution derivations X_H in C and $X_{\Phi(H)}$ in C_1 are Φ compatible: $\Phi X_H=X_{\Phi(H)}\Phi$. If B_1 is a Hamiltonian matrix in C_1 such that $\bar{X}_F=X_F(\bar{p})=B_1(\delta F/\delta\bar{p})$, then Φ is Hamiltonian if and only if

$$\Phi(B)=D(\bar{\Phi})B_1D(\bar{\Phi})^{\dagger}, \tag{2.20}$$

where

$$(\bar{\Phi})_i=\Phi_i, \tag{2.21}$$

and $D(\bar{\Phi})$ is the Fréchet derivative of $\bar{\Phi}$:

$$[D(\bar{\Phi})]_{ij}=D_j(\Phi_i)=\sum\frac{\partial\Phi_i}{\partial q_j^{(g|\nu)}}\hat{g}\,\partial^{\nu}. \tag{2.22}$$

An operator $a\hat{g}\partial^{\nu}$: $C\to C$ is called q independent (resp. q linear) if $a\in K$ (resp. if $a=\sum a_{i,h,\mu}q_i^{(h|\mu)}$, $a...\in K$). An operator is affine if it is a sum of a q-independent and a q-linear operator. The same terminology applies to sums of operators, and to matrix operators.

Let $B=B^l+b$ be an affine operator: $C^N\to C^N$, with B^l being q linear and b being q independent. We make K^N into a (differential-difference) algebra setting

$$\bar{q}^t[X,Y]\sim X^tB^l(Y),\quad (\bar{q})_i:=q_i,\quad X,Y\in K^N. \tag{2.23}$$

Conversely, given an algebra structure on K^N, (2.23) defines a q-linear operator B^l. The relation between affine Hamiltonian operators and two-cocycles on Lie algebras is one-to-one: given a Lie algebra $\mathfrak{g}$ and a two-cocycle ω on it, we set $B=B^l+b_\omega$, with B^l defined by (2.23). Conversely, given an affine Hamiltonian matrix $B=B^l+b$, the same formula (2.23) defines a Lie algebra structure on K^N while (2.3) defines a two-cocycle ω via $b_\omega=b$.

III. INTERIOR CLEBSCH REPRESENTATIONS

Let $\mathfrak{g}=K^N$ be a Lie algebra, $W=K^{N_1}$, $V=K^{N_2}$. Let ${}^1\rho$: $\mathfrak{g}\to\operatorname{Diff}(W)$ and ${}^2\rho$: $\mathfrak{g}\to\operatorname{Diff}(V)$ be two representations of $\mathfrak{g}$. Representation ${}^3\rho$: ${}^3\rho(X)=-{}^2\rho(X)^{\dagger}$, $X\in\mathfrak{g}$, of $\mathfrak{g}$ on $V^*:=K^{N_2}$ is called the dual representation of $\mathfrak{g}$ on V^* (Ref. 4, Proposition 3.3). Let $\bar{\mathfrak{g}}_1=\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$ be the semidirect product of $\mathfrak{g}$ with $W\oplus V^*\oplus V$, and let B_1^l be the natural Hamiltonian matrix associated by (2.23) with the Lie algebra $\bar{\mathfrak{g}}_1$:

$$\sum q_k[X,Y]_k+\sum c_i(X\cdot v_1-Y\cdot u_1)_i$$

$$+\sum\lambda_j(X\cdot v_2-Y\cdot u_2)_j \tag{3.1a}$$

$$+\sum\gamma_j(X\cdot v_3-Y\cdot u_3)_j \tag{3.1b}$$

$$\sim\begin{pmatrix}X\\u_1\\u_3\\u_2\end{pmatrix}^t B_1^l\begin{pmatrix}Y\\v_1\\v_3\\v_2\end{pmatrix},$$

in the ring

$$C_1=K[q_k^{(g|\nu)},c_i^{(g|\nu)},\gamma_j^{(g|\nu)},\lambda_j^{(g|\nu)}],$$

$$1\leqslant k\leqslant N,\quad 1\leqslant i\leqslant N_1,\quad 1\leqslant j\leqslant N_2, \tag{3.2}$$

which plays the role of the functions on "the dual space to $\bar{\mathfrak{g}}_1$," where $X,Y\in\mathfrak{g}$, $u_1,v_1\in W$, $u_2,v_2\in V$, $u_3,v_3\in V^*$, and $[X\cdot(\ )]_i:={}^i\rho(X)(\ )$.

Let ω be the symplectic two-cocycle on $\bar{\mathfrak{g}}_1$, with

$$b_\omega=\left(\begin{array}{cc|cc}0&0&0&0\\0&0&0&0\\\hline 0&0&0&1\\0&0&-1&0\end{array}\right), \tag{3.3}$$

and let the corresponding affine Hamiltonian structure B_1 in C_1 be defined as

$$B_1 = B_1^l + b_\omega. \tag{3.4}$$

Let $\bar{\mathfrak{g}} = \mathfrak{g} \ltimes (W \oplus (\mathfrak{g} \ltimes V))$ be the semidirect product of $\mathfrak{g}$ with $W \oplus (\mathfrak{g} \ltimes V)$, and let B^l be the associated Hamiltonian structure in the ring

$$C = K\left[q_k^{(g|\nu)},\ c_i^{(g|\nu)},\ p_k^{(g|\nu)},\ \lambda_j^{(g|\nu)}\right],$$
$$1 < k < N, \quad 1 < i < N_1, \quad 1 < j < N_2: \tag{3.5}$$

$$\sum q_k[X,Y]_k + \sum c_i(X\cdot v_1 - Y\cdot u_1)_i$$
$$+ \sum \lambda_j (X\cdot v_2 - Y\cdot u_2)_j \tag{3.6a}$$
$$+ \sum p_k([X,y]_k - [Y,x]_k + [x,y]_k)$$
$$+ \sum \lambda_j (x\cdot v_2 - y\cdot u_2)_j \tag{3.6b}$$
$$\sim \begin{pmatrix} X \\ u_1 \\ x \\ u_2 \end{pmatrix}^t B^l \begin{pmatrix} Y \\ v_1 \\ y \\ v_2 \end{pmatrix},$$

where $x,y \in \mathfrak{g}$, and the rest of the notation in (3.5) and (3.6) is the same as in (3.2) and (3.1).

To prepare the grounds for the desired Clebsch map $\Phi: C \to C_1$, we need the following result.

Lemma 3.1: Let $\rho: \mathfrak{g} \to \mathrm{Diff}(T)$ be a representation, with $T = K^M$, and with

$$\rho(X)_{\alpha\beta} = \sum \rho_{\alpha\beta}^{k;g,\nu|h,\sigma} X_k^{(h|\sigma)} \hat{g}\, \partial^\nu, \quad \rho^{k\ldots}_{\ldots} \in K. \tag{3.7}$$

Define the map $\nabla: T \times T \to \mathfrak{g}$, $u \times v \mapsto u \nabla v$, by

$$(u\nabla v)_k = \sum \hat{h}^{-1}(-\partial)^\sigma\left(\rho_{\alpha\beta}^{k;g,\nu|h,\sigma} v_\alpha u_\beta^{(g|\nu)}\right). \tag{3.8}$$

Then

$$v^t \rho(X)(u) \sim X^t(u\nabla v), \quad u,v \in T, \quad X \in \mathfrak{g}. \tag{3.9}$$

Proof: We have, by (3.7),

$$v^t\rho(X)(u) = \sum v_\alpha\, \rho_{\alpha\beta}^{k;g,\nu|h,\sigma} X_k^{(h|\sigma)} u_\beta^{(g|\nu)}$$
$$\sim \sum X_k \hat{h}^{-1}(-\partial)^\sigma\left(\rho_{\alpha\beta}^{k;g,\nu|h,\sigma} v_\alpha u_\beta^{(g|\nu)}\right). \quad \blacksquare$$

Remark 3.2: The property (3.9) uniquely defines the product $u\nabla v$.

Denote by $\gamma, \lambda \in C_1^{N_2}$ two vectors with components γ_j and λ_j, respectively. Set $T = V \underset{K}{\otimes} C_1$, $\rho = {}^2\rho$ in Lemma 3.1 and define

$$\Psi_k = (\gamma \nabla \lambda)_k. \tag{3.10}$$

Theorem 3.3: Define a homomorphism $\Phi: C \to C_1$ by the formulas:

$$\Phi(q_k) = q_k, \quad \Phi(c_i) = c_i, \quad \Phi(p_k) = \Psi_k,$$
$$\Phi(\lambda_j) = \lambda_j, \quad 1 < k < N, \quad 1 < i < N_1, \quad 1 < j < N_2. \tag{3.11}$$

Then the map Φ is Hamiltonian between the Hamiltonian structures $B = B^l$ in C and $B_1 = B_1^l + b_\omega$ in C_1.

Proof: We have to check out the equality (2.20). From (3.11), we find the Fréchet derivative $D(\bar{\Phi})$ to be

$$D(\bar{\Phi}) = \begin{array}{c} \\ \Phi(q_k) \\ \Phi(c_i) \\ \Phi(p_k) \\ \Phi(\lambda_j) \end{array} \begin{array}{c} \begin{array}{cccc} q_s & c_r & \gamma_l & \lambda_l \end{array} \\ \begin{pmatrix} \delta_s^k & 0 & 0 & 0 \\ 0 & \delta_r^i & 0 & 0 \\ 0 & 0 & D\Psi_k/D\gamma_l & D\Psi_k/D\lambda_l \\ 0 & 0 & 0 & \delta_l^j \end{pmatrix} \end{array}. \tag{3.12}$$

Using (3.3) and (3.4), we write the matrix B_1 in the form

$$\begin{array}{c} \\ q_s \\ c_r \\ \gamma_l \\ \lambda_l \end{array} \begin{array}{c} \begin{array}{cccc} q_k & c_i & \gamma_j & \lambda_j \end{array} \\ \begin{pmatrix} b_{q_sq_k} & b_{q_sc_i} & b_{q_s\gamma_j} & b_{q_s\lambda_j} \\ b_{c_rq_k} & 0 & 0 & 0 \\ b_{\gamma_lq_k} & 0 & 0 & \delta_j^l \\ b_{\lambda_lq_k} & 0 & -\delta_j^l & 0 \end{pmatrix} \end{array}, \tag{3.13}$$

where $b_{\ldots}$ are the corresponding matrix elements of the matrix B_1^l in (3.1). Hence, for the matrix in the right-hand side of (2.20) we obtain

$$\begin{array}{c} \\ \Phi(q_k) \\ \\ \Phi(c_i) \\ \\ \Phi(p_k) \\ \\ \Phi(\lambda_j) \end{array} \begin{array}{c} \begin{array}{cccc} \Phi(q_l) & \Phi(c_r) & \Phi(p_l) & \Phi(\lambda_s) \end{array} \\ \begin{pmatrix} b_{q_kq_l} & b_{q_kc_r} & \sum_a b_{q_k\gamma_a}\left(\frac{D\Psi_l}{D\gamma_a}\right)^\dagger + \sum_a b_{q_k\lambda_a}\left(\frac{D\Psi_l}{D\lambda_a}\right)^\dagger & b_{q_k\lambda_s} \\ b_{c_iq_l} & 0 & 0 & 0 \\ \sum_a \frac{D\Psi_k}{D\gamma_a} b_{\gamma_aq_l} + \sum_a \frac{D\Psi_k}{D\lambda_a} b_{\lambda_aq_l} & 0 & \sum_a \frac{D\Psi_k}{D\gamma_a}\left(\frac{D\Psi_l}{D\lambda_a}\right)^\dagger - \sum_a \frac{D\Psi_k}{D\lambda_a}\left(\frac{D\Psi_l}{D\gamma_a}\right)^\dagger & \frac{D\Psi_k}{D\gamma_s} \\ b_{\lambda_jq_l} & 0 & -\left(\frac{D\Psi_l}{D\gamma_j}\right)^\dagger & 0 \end{pmatrix} \end{array}. \tag{3.14}$$

To show that the matrix (3.14) equals the matrix $\Phi(B)$ in the left-hand side of (2.20), we apply each of these matrices to the vector $(Y,v_1,y,v_2)^t$, multiply the result from the left by the row vector (X,u_1,x,u_2), and then show that the resulting expressions differ at most by an element from Im $\mathscr{D}$. Since Φ in (3.11) acts identically on the q's, c's, and λ's, and the left-hand sides of the row (3.1a) and the row (3.6a) are the same, it remains to verify the following relation:

$$\sum\Phi(p_k)([X,y]_k-[Y,x]_k+[x,y]_k)$$

$$+\sum\lambda_j(x\cdot v_2-y\cdot u_2)_j \tag{3.15l}$$

$$\sim\sum X_k\left[b_{q_k\gamma_a}\left(\frac{D\Psi_l}{D\gamma_a}\right)^\dagger+b_{q_k\lambda_a}\left(\frac{D\Psi_l}{D\lambda_a}\right)^\dagger\right](y_l) \tag{3.15a}$$

$$+\sum x_k\left(\frac{D\Psi_k}{D\gamma_a}b_{\gamma_a q_l}+\frac{D\Psi_k}{D\lambda_a}b_{\lambda_a q_l}\right)(Y_l) \tag{3.15b}$$

$$+\sum x_k\left[\frac{D\Psi_k}{D\gamma_a}\left(\frac{D\Psi_l}{D\lambda_a}\right)^\dagger-\frac{D\Psi_k}{D\lambda_a}\left(\frac{D\Psi_l}{D\gamma_a}\right)^\dagger\right](y_l) \tag{3.15c}$$

$$+\sum\left[x_k\frac{D\Psi_k}{D\gamma_s}(v_{2s})-u_{2j}\left(\frac{D\Psi_l}{D\gamma_j}\right)^\dagger(y_l)\right]. \tag{3.15d}$$

To check (3.15), we use the following identities:

$$\frac{D\Psi_k}{D\gamma_j}=\sum\hat{h}^{-1}(-\partial)^\sigma\,{}^2\rho_{sj}^{k;\,g,\nu|h,\sigma}\lambda_s\,\hat{g}\,\partial^\sigma, \tag{3.16}$$

$$\frac{D\Psi_k}{D\lambda_j}=\sum\hat{h}^{-1}(-\partial)^\sigma\,{}^2\rho_{js}^{k;\,g,\nu|h,\sigma}\gamma_s^{(g|\nu)}, \tag{3.17}$$

$$(X\cdot\gamma)_j=\sum\left(\frac{D\Psi_k}{D\lambda_j}\right)^\dagger(X_k),\quad\gamma\in V, \tag{3.18}$$

$$\lambda^t(X\cdot v)\sim\sum v_j\left(\frac{D\Psi_k}{D\gamma_j}\right)^\dagger(X_k),\quad v\in V, \tag{3.19}$$

with notation from Lemma 3.1. Since $\Psi_k=(\gamma\nabla\lambda)_k$, formulas (3.16) and (3.17) follow at once from (3.8). Now, from (3.7) we get

$$(X\cdot\gamma)_j=\sum{}^2\rho_{js}^{k;\,g,\nu|h,\sigma}X_k^{(h|\sigma)}\gamma_s^{(g|\nu)}$$

$$=\sum{}^2\rho_{js}^{k;\,g,\nu|h,\sigma}\gamma_s^{(g|\nu)}\hat{h}\,\partial^\sigma(X_k)$$

$$=\sum[\hat{h}^{-1}(-\partial)^\sigma\,{}^2\rho_{js}^{k;\,g,\nu|h,\sigma}\gamma_s^{(g|\nu)}]^\dagger(X_k)$$

$$=\sum\left(\frac{D\Psi_k}{D\lambda_j}\right)^\dagger(X_k)\quad[\text{by }(3.17)],$$

which is (3.18). At last,

$$\lambda^t(X\cdot v)=\lambda_s\,{}^2\rho_{sj}^{k;\,g,\nu|h,\sigma}X_k^{(h|\sigma)}v_j^{(g|\nu)}\quad[\text{by }(3.7)]$$

$$\sim\sum v_j\,\hat{g}^{-1}(-\partial)^\nu\,{}^2\rho_{sj}^{k;\,g,\nu|h,\sigma}\lambda_s\hat{h}\partial^\sigma(X_k)$$

$$\sim\sum v_j\left(\frac{D\Psi_k}{D\gamma_j}\right)^\dagger(X_k)\quad[\text{by }(3.16)],$$

which is (3.19).

We start with (3.15d):

$$\sum x_k\frac{D\Psi_k}{D\gamma_s}(v_{2s})-\sum u_{2j}\left(\frac{D\Psi_l}{D\gamma_j}\right)^\dagger(y_l)$$

$$\sim\sum v_{2s}\left(\frac{D\Psi_k}{D\gamma_s}\right)^\dagger(x_k)-\sum u_{2j}\left(\frac{D\Psi_l}{D\gamma_j}\right)^\dagger(y_l)$$

$$\sim\lambda^t(x\cdot v_2-y\cdot u_2)\quad[\text{by }(3.19)],$$

which is the fourth term in (3.15l).

Next, for the third term in (3.15l) we obtain

$$\sum\Phi(p_k)[x,y]_k=\sum\Psi_k[x,y]_k$$

$$=\sum(\gamma\nabla\lambda)_k[x,y]_k\sim\lambda^t([x,y]\cdot\gamma)\quad[\text{by }(3.9)]$$

$$=\lambda^t[x\cdot(y\cdot\gamma)-y\cdot(x\cdot\gamma)]\quad(\text{since }{}^2\rho\text{ is a representation})$$

$$\sim\sum(y\cdot\gamma)_j\left(\frac{D\Psi_k}{D\gamma_j}\right)^\dagger(x_k)-\sum(x\cdot\gamma)_j\left(\frac{D\Psi_l}{D\gamma_j}\right)^\dagger(y_l)\quad[\text{by }(3.19)]$$

$$=\sum\left(\frac{D\Psi_l}{D\lambda_j}\right)^\dagger(y_l)\cdot\left(\frac{D\Psi_k}{D\gamma_j}\right)^\dagger(x_k)-\sum\left(\frac{D\Psi_k}{D\lambda_j}\right)^\dagger(x_k)\cdot\left(\frac{D\Psi_l}{D\gamma_j}\right)(y_l)\quad[\text{by }(3.18)]$$

$$\sim\sum x_k\left[\frac{D\Psi_k}{D\gamma_j}\left(\frac{D\Psi_l}{D\lambda_j}\right)^\dagger(y_l)-\frac{D\Psi_k}{D\lambda_j}\left(\frac{D\Psi_l}{D\gamma_j}\right)^\dagger(y_l)\right],$$

which is (3.15c).

Further, for (3.15b), we get

$$\sum x_k \frac{D\Psi_k}{D\gamma_a} b_{\gamma_a q_l}(Y_l) + \sum x_k \frac{D\Psi_k}{D\lambda_a} b_{\lambda_a q_l}(Y_l)$$
$$\sim \sum \left(\frac{D\Psi_k}{D\gamma_a}\right)^\dagger (x_k)\cdot b_{\gamma_a q_l}(Y_l) + \sum \left(\frac{D\Psi_k}{D\lambda_a}\right)^\dagger (x_k)\cdot b_{\lambda_a q_l}(Y_l)$$
$$\sim -\sum Y_l \left[b_{q_l\gamma_a}\left(\frac{D\Psi_k}{D\gamma_a}\right)^\dagger + b_{q_l\lambda_a}\left(\frac{D\Psi_k}{D\lambda_a}\right)^\dagger \right](x_k) \quad [\text{since } (b_{\gamma_a q_l})^\dagger = -b_{q_l\gamma_a},\ \ (b_{\lambda_a q_l})^\dagger = -b_{q_l\lambda_a}],$$

which equals minus the expression (3.15a) with (X,y) changed into (Y,x): and this is exactly the relation between the first and the second term in (3.15l). Thus, it remains only to compare (3.15a) with the first term in (3.15l). Denote by $b_\gamma(X)$ and $b_\lambda(X)$ the following vectors:

$$[b_\gamma(X)]_a = \sum b_{\gamma_a q_k}(X_k), \quad [b_\lambda(X)]_a = \sum b_{\lambda_a q_k}(X_k). \tag{3.20}$$

Then, we transform (3.15a) as

$$\sum X_k \left[b_{q_k\gamma_a}\left(\frac{D\Psi_l}{D\gamma_a}\right)^\dagger + b_{q_k\lambda_a}\left(\frac{D\Psi_l}{D\lambda_a}\right)^\dagger \right](y_l)$$
$$\sim -\sum b_{\gamma_a q_k}(X_k)\cdot\left(\frac{D\Psi_l}{D\gamma_a}\right)^\dagger (y_l) - \sum b_{\lambda_a q_k}(X_k)\cdot\left(\frac{D\Psi_l}{D\lambda_a}\right)^\dagger (y_l)$$
$$\sim -\lambda^t(y\cdot b_\gamma(X)) - b_\lambda(X)^t(y\cdot\gamma) \quad [\text{by (3.19) and (3.18)}]$$
$$\sim (y\cdot\lambda)^t b_\gamma(X) - (y\cdot\gamma)^t b_\lambda(X) \quad [\text{since the dual representation } {}^3\rho(X) = -{}^2\rho(X)^\dagger,\ \ \forall X\in\mathfrak{g}]$$
$$\sim -\gamma^t(X\cdot(y\cdot\lambda)) + \lambda^t(X\cdot(y\cdot\gamma)) \quad [\text{by (3.1b) and (3.1a)}]$$
$$\sim -\gamma^t(X\cdot(y\cdot\lambda)) + \gamma^t(y\cdot(X\cdot\lambda)) \quad [\text{by } {}^3\rho(Y) = -{}^2\rho(Y)^\dagger]$$
$$= -\gamma^t([X,y]\cdot\lambda) \quad (\text{since } {}^2\rho \text{ is a representation}) \sim \lambda^t([X,y]\cdot\gamma)$$
$$\sim [X,y]^t(\gamma\nabla\lambda) \quad [\text{by (3.9)}] = [X,y]^t\Psi = \sum \Phi(p_k)[X,y]_k,$$

and this is precisely the first term of (3.15l). ∎

Remark 3.4: The reader may have wondered if the same Φ will provide a more general, non-Abelian, Clebsch map between Lie algebras $\mathfrak{g}\ltimes(W\oplus(\mathfrak{h}\ltimes V))$ and $\mathfrak{g}\ltimes(W\oplus V^*\oplus V)$, with a different Lie algebra $\mathfrak{h}$ acting on V in addition to $\mathfrak{g}$. The answer appears to be negative, as may be seen from the following example. Take $G=\{e\}$, $n=1$, $\partial=\partial_1$, $\mathfrak{g}=D_1$. For each $\lambda\in k=\operatorname{Ker}\partial|_K$, denote V_λ the space K^1 with the following action of D_1:

$$X: f\mapsto X\partial(f) + \lambda f\partial(X) =: L_X^\lambda(f). \tag{3.21}$$

Since

$$gL_X^\lambda(f) = g[X\partial(f) + \lambda f\partial(X)]$$
$$\sim f[-\partial(Xg) + \lambda g\,\partial(X)]$$
$$= -f[X\partial(g) + (1-\lambda)g\,\partial(X)],$$

we see that

$$(V_\lambda)^* \approx V_{1-\lambda}. \tag{3.22}$$

Let us take $W=\{0\}$, $V=V_\lambda$, and consider V_λ also as V_μ for some $\mu\in k$, with respect to the action of another copy of $\mathfrak{g}=D_1$ (every V_λ is a free one-dimensional K-module). Applying the map (3.11) with

$$\Psi = \mu\lambda\,\partial(\gamma) - (1-\mu)\gamma\,\partial(\lambda), \tag{3.23}$$

we can easily see that this map is Hamiltonian iff $\mu=\lambda$.

Remark 3.5: Combining our Hamiltonian map

$$\Phi:\ C((\mathfrak{g}\ltimes(W\oplus(\mathfrak{g}\ltimes V)))^*\to C_1((\mathfrak{g}\ltimes(W\oplus V^*\oplus V))^*)$$

with the Hamiltonian map

$$C_1((\mathfrak{g}\ltimes(W\oplus V^*\oplus V))^*)\to C_2(W\oplus W^*\oplus V^*\oplus V)$$

given by Theorem 3.2 in Ref. 5, we find a genuine canonical representation for the dual space to the Lie algebra $\mathfrak{g}\ltimes(W\oplus(\mathfrak{g}\ltimes V))$.

Remark 3.6: If $N=\dim\mathfrak{g}<\dim V^*=N_2$ and if our map Φ is injective, we can think of the Hamiltonian structure in C as being the restriction from the Hamiltonian structure in C_1. (In the finite-dimensional case, the dual process, in the language of manifolds, is called "reduction.") In practice, however, the inequality $N<N_2$ is rarely available, as we shall see in the next two sections treating irrotational ^{4}He and rotating ^{4}He.

IV. APPLICATIONS TO ^{4}He

We start with the case of irrotational ^{4}He first. To derive formulas (1.5)–(1.7) for the second Poisson bracket out of the data (1.1)–(1.4) describing the first Poisson bracket, we set

$$\mathfrak{g}=D_n,\quad W=\Lambda^0,\quad V=\Lambda^0,\quad V^*=\Lambda^n. \tag{4.1}$$

The action of D_n on $V=\Lambda^0$ is given by

$$\rho(X)(u)=\sum X_k u_{,k}, \quad X=\sum X_k \partial_k \in D_n.$$

Hence, by (3.9)

$$v^t\rho(X)(u)=\sum vX_k u_{,k}=\sum X_k(vu_{,k}),$$

which means

$$(u\nabla v)_k = vu_{,k}. \tag{4.2}$$

In particular, by (3.10),

$$\Psi_k=(\gamma\nabla\lambda)_k=\lambda\gamma_{,k}, \tag{4.3}$$

and the map (3.11) becomes

$$\Phi(M_k)=M_k, \quad \Phi(\sigma)=\sigma, \quad \Phi(P_k)=\rho\alpha_{,k}, \quad \Phi(\rho)=\rho, \tag{4.4}$$

if we adjust our notation to

$$q_k=M_k, \quad c=c_1=\sigma, \quad P_k=P_k, \quad \gamma=\alpha, \quad \lambda=\rho. \tag{4.5}$$

Formula (4.4) is exactly (dual to) the map (1.6). By Theorem 3.3, the Hamiltonian map Φ relates the Hamiltonian structure on the dual space to the Lie algebra $\mathfrak{g}\ltimes(W\oplus(\mathfrak{g}\ltimes V))=D_n\ltimes(\Lambda^0\oplus(D_n\ltimes\Lambda^0))$ [which is the Lie algebra $\mathfrak{g}({}^4\mathrm{He}_{nr})'$ in (1.7)] to the Hamiltonian structure on the dual to the Lie algebra $\mathfrak{g}\ltimes(W\oplus V^*\oplus V)=D_n\ltimes(\Lambda^0\oplus\Lambda^n\oplus\Lambda^0)$ [which is the Lie algebra $\mathfrak{g}({}^4\mathrm{He}_{nr})$ in (1.2)] accompanied by the symplectic two-cocycle on $V^*\oplus V=\Lambda^n\oplus\Lambda^0$ [which coincides with (1.4)]. The reader can see how easily all the complexities apparent at the first encounter with the second Poisson structure of irrotational ^{4}He, disappear at once when viewed from the general point of view provided by Theorem 3.3.

The case of rotating ^{4}He is more instructive. The Poisson bracket in this case [formula (14) in Ref. 3] is

$$\{H,F\}\sim\left\{\frac{\delta F}{\delta M_k}\left[(M_l\partial_k+\partial_l M_k)\left(\frac{\delta H}{\delta M_l}\right)+(\partial_l a_k-a_{l,k})\left(\frac{\delta H}{\delta a_l}\right)+\rho\,\partial_k\left(\frac{\delta H}{\delta\rho}\right)-\alpha_{,k}\frac{\delta H}{\delta\alpha}\right]\right.$$

$$+\left[\frac{\delta F}{\delta a_k}(a_{k,l}+a_l\partial_k)+\frac{\delta F}{\delta\rho}\partial_l\rho+\frac{\delta F}{\delta\alpha}\alpha_{,l}\right]\left(\frac{\delta H}{\delta M_l}\right) \tag{4.6a1}$$

$$+\left(\frac{\delta F}{\delta\rho}\frac{\delta H}{\delta\alpha}-\frac{\delta F}{\delta\alpha}\frac{\delta H}{\delta\rho}\right) \tag{4.6a2}$$

$$+\frac{\delta F}{\delta\pi_k}\left[(\pi_l\partial_k+\partial_l\pi_k)\left(\frac{\delta H}{\delta\pi_l}\right)+\sigma\,\partial_k\left(\frac{\delta H}{\delta\sigma}\right)\right]+\frac{\delta F}{\delta\sigma}\partial_l\sigma\left(\frac{\delta H}{\delta\pi_l}\right), \tag{4.6b}$$

where new notations, in addition to the nonrotating ^{4}He case, are π is the relative normal momentum density; and $\mathbf{a}$ is the vorticial part of the superfluid velocity $\mathbf{v}^s$, which, for rotating ^{4}He, is not curl-free anymore, i.e., $\mathbf{v}^s=\mathbf{a}-\nabla\alpha$.

The bracket (4.6) splits off in two separate brackets: (4.6a) and (4.6b). The bracket (4.6b) is the natural bracket on the semidirect product Lie algebra

$$\bar{\mathfrak{g}}_2=D_n\ltimes\Lambda^0, \tag{4.7}$$

and we shall not need this part for some time. The bracket (4.6a) is of the form $B_1=B_1^l(\bar{\mathfrak{g}}_1)+b$, where $B_1^l(\bar{\mathfrak{g}}_1)$, given by (4.6a1), is naturally associated with the semidirect product Lie algebra

$$\bar{\mathfrak{g}}_1=D_n\ltimes(\Lambda^{n-1}\oplus\Lambda^0\oplus\Lambda^n), \tag{4.8}$$

while b, given by (4.6a2), is the symplectic two-cocycle on the $\Lambda^0\oplus\Lambda^n$- part of $\bar{\mathfrak{g}}_1$. [Notice the transposition of Λ^n and Λ^0 in contrast to the nonrotating ^{4}He case, which is responsible for formulas (1.1b) and (4.6a2) having *opposite* signs.]

From Theorem 3.3 we conclude that there exists a second Poisson bracket description of rotating ^{4}He. Let us find it. From (4.8) we see that

$$\mathfrak{g}=D_n, \quad W=\Lambda^{n-1}, \quad V=\Lambda^n, \quad V^*=\Lambda^0, \tag{4.9}$$

and that the sign of the two-cocycle (4.6 a2) is correct, i.e., it is the same as in (3.3). The action of $\mathfrak{g}=D_n$ on $V=\Lambda^n$ is

$$\left(\sum X_k\partial_k\right)(\omega\,d^nx)=\sum(X_k\omega)_{,k}\,d^nx,$$

$$d^nx:=dx_1\wedge\cdots\wedge dx_n. \tag{4.10}$$

Hence, by (3.9),

$$v^t\rho(X)(u)=\sum v(X_k u)_{,k}\sim-\sum X_k(uv_{,k}). \tag{4.11}$$

Thus

$$(u\nabla v)_k=-uv_{,k}, \tag{4.12}$$

so that

$$\Psi_k=(\gamma\nabla\lambda)_k=-\gamma\lambda_{,k}. \tag{4.13}$$

Therefore, the map (3.11) becomes

$$\Phi(M_k)=M_k, \quad \Phi(a_i)=a_i,$$

$$\Phi(p_k)=-\rho\alpha_{,k}, \quad \Phi(\alpha)=\alpha, \tag{4.14}$$

with the identification

$$q_k=M_k, \quad c_i=a_i, \quad \gamma=\rho, \quad \lambda=\alpha. \tag{4.15}$$

However, (4.14) is not what we would like to have, since we want to *exchange* α for p *keeping* the density ρ *intact*, and not the other way around. The reason for this misfortune is clear: our Theorem 3.3 exchanges V^* for $\mathfrak{g}$ while we would like instead to exchange V for $\mathfrak{g}$ or, equivalently, to be able to change the sign in the symplectic two-cocycle (3.3). The remedy, then, is equally clear.

Theorem 4.1: Let

$$C' = K\,[q_k^{(g|v)}, c_i^{(g|v)}, p_k^{(g|v)}, \gamma_j^{(g|v)}] \tag{4.16}$$

serve the Lie algebra

$$\bar{\mathfrak{g}}' = \mathfrak{g} \ltimes (W \oplus (\mathfrak{g} \ltimes V^*)), \tag{4.17}$$

and let the Hamiltonian matrix $B' = B^{l'}$ in C' be associated by (2.23) with $\bar{\mathfrak{g}}'$:

$$\sum q_k [X,Y]_k + \sum c_i (X \cdot v_1 - Y \cdot u_1)_i$$

$$+ \sum \gamma_j (X \cdot v_3 - Y \cdot u_3)_j \tag{4.18a}$$

$$+ \sum p_k ([X,y]_k - [Y,x]_k + [x,y]_k) $$

$$+ \sum \gamma_j (x \cdot v_3 - y \cdot u_3)_j \tag{4.18b}$$

$$\sim \begin{pmatrix} X \\ u_1 \\ x \\ u_3 \end{pmatrix}^t B' \begin{pmatrix} Y \\ v_1 \\ y \\ v_3 \end{pmatrix},$$

with $X,Y,x,y \in \mathfrak{g}$, $u_1, v_1 \in W$, $u_3, v_3 \in V^*$. In the notation of Theorem 3.3, define the map Φ: $C' \to C_1$ by

$$\Phi(q_k) = q_k, \quad \Phi(c_i) = c_i, \quad \Phi(p_k) = \Psi_k,$$
$$\Phi(\gamma_j) = \gamma_j, \quad 1 \leqslant u \leqslant N, \quad 1 \leqslant i \leqslant N_1, \quad 1 \leqslant j \leqslant N_2. \tag{4.19}$$

Then the map Φ is Hamiltonian.

Proof: We follow the Proof of Theorem 3.1. To check (2.20), we first find

$$D(\bar{\Phi}) = \begin{array}{c} \\ \Phi(q_k) \\ \Phi(c_i) \\ \Phi(p_k) \\ \Phi(\gamma_j) \end{array} \begin{array}{c} \begin{array}{cccc} q_s & c_r & \gamma_l & \lambda_l \end{array} \\ \begin{pmatrix} \delta_k^s & 0 & 0 & 0 \\ 0 & \delta_i^r & 0 & 0 \\ 0 & 0 & \dfrac{D\Psi_k}{D\gamma_l} & \dfrac{D\Psi_k}{D\lambda_l} \\ 0 & 0 & \delta_j^l & 0 \end{pmatrix} \end{array}. \tag{4.20}$$

Using B_1 written in the form (3.13), for the right-hand side of (2.20) we obtain

$$\begin{array}{c} \\ \Phi(q_k) \\ \\ \Phi(c_i) \\ \\ \Phi(p_k) \\ \\ \Phi(\gamma_j) \end{array} \begin{array}{c} \begin{array}{cccc} \Phi(q_l) & \Phi(c_r) & \Phi(p_l) & \Phi(\gamma_s) \end{array} \\ \begin{pmatrix} b_{q_k q_l} & b_{q_k c_r} & \sum b_{q_k \gamma_a} \left(\dfrac{D\Psi_l}{D\gamma_a}\right)^\dagger + \sum b_{q_k \lambda_a} \left(\dfrac{D\Psi_l}{D\lambda_a}\right)^\dagger & b_{q_k \gamma_s} \\ b_{c_i q_l} & 0 & 0 & 0 \\ \sum \dfrac{D\Psi_k}{D\gamma_a} b_{\gamma_a q_l} + \sum \dfrac{D\Psi_k}{D\lambda_a} b_{\lambda_a q_l} & 0 & \sum \dfrac{D\Psi_k}{D\gamma_a} \left(\dfrac{D\Psi_l}{D\lambda_a}\right)^\dagger - \sum \dfrac{D\Psi_k}{D\lambda_a} \left(\dfrac{D\Psi_l}{D\gamma_a}\right)^\dagger & -\dfrac{D\Psi_k}{D\lambda_s} \\ b_{\lambda_j q_l} & 0 & \left(\dfrac{D\Psi_l}{D\lambda_j}\right)^\dagger & 0 \end{pmatrix} \end{array}. \tag{4.21}$$

Again, applying $\Phi(B')$ from the left-hand side of (2.20) and the matrix (4.21) from the right-hand side of (2.20) to the vector $(Y, v_1, y, v_3)^t$, then multiplying the result from the left by the row vector (X, u_1, x, u_3), using (4.18) and the already proved formula (3.15), we are left with the following relation to check out:

$$\sum \gamma_j (x \cdot v_3 - y \cdot u_3)_j$$

$$\sim \sum \left[-x_k \frac{D\Psi_k}{D\lambda_s}(v_{3s}) + u_{3j} \left(\frac{D\Psi_l}{D\lambda_j}\right)(y_l) \right]. \tag{4.22}$$

This relation can be verified as follows:

$$\sum \gamma_j (x \cdot v_3 - y \cdot u_3)_j$$

$$= \gamma^t (x \cdot v_3 - y \cdot u_3) \sim -(x \cdot \gamma)^t v_3 + (y \cdot \gamma)^t u_3,$$

[taking the adjoint representation to ${}^3\rho$]

$$= -\sum v_{3j} \left(\frac{D\Psi_k}{D\lambda_j}\right)^\dagger (x_k)$$

$$+ \sum u_{3j} \left(\frac{D\Psi_k}{D\lambda_j}\right)^\dagger (y_k) \quad \text{[by (3.18)]}$$

$$\sim -\sum x_k \left(\frac{D\Psi_k}{D\lambda_j}\right)(v_{3j}) + \sum u_{3j} \left(\frac{D\Psi_k}{D\lambda_j}\right)^\dagger (y_k),$$

which is the right-hand side of (4.22). ■

Now we are in a position to obtain the correct form of the second Poisson structure for rotating ^{4}He. Applying the map Φ from (4.19), we get, using computations (4.10)–(4.13),

$$\Phi(M_k) = M_k, \quad \Phi(a_i) = a_i,$$
$$\Phi(p_k) = -\rho\alpha_{,k}, \quad \Phi(\rho) = \rho, \tag{4.23}$$

with the identification (4.15). Hence, we obtain the natural Hamiltonian structure associated, by Theorem 4.1, with the Lie algebra

$$[D_n \circledS (\Lambda^{n-1} \oplus (D_n \circledS \Lambda^0))] \oplus [D_n \circledS \Lambda^0], \tag{4.24}$$

the second summand in the square brackets being the unchanged Lie algebra $\bar{\mathfrak{g}}_2$ from (4.7). The Poisson bracket associated with (4.24) is

$$\{H,F\} \sim \frac{\delta F}{\delta M_k}\left[(M_l\partial_k + \partial_l M_k)\left(\frac{\delta H}{\delta M_l}\right) + (\partial_l a_k - a_{l,k})\left(\frac{\delta H}{\delta a_l}\right) + (p_l\partial_k + \partial_l p_k)\left(\frac{\delta H}{\delta p_l}\right)\right.$$
$$\left. + \rho\partial_k\left(\frac{\delta H}{\delta\rho}\right)\right] + \left[\frac{\delta F}{\delta a_k}(a_{k,l} + a_l\partial_k) + \frac{\delta F}{\delta p_k}(p_l\partial_k + \partial_l p_k) + \frac{\delta F}{\delta\rho}\partial_l\rho\right]\left(\frac{\delta H}{\delta M_l}\right)$$
$$+ \frac{\delta F}{\delta p_k}\left[(p_l\partial_k + \partial_l p_k)\left(\frac{\delta H}{\delta p_l}\right) + \rho\partial_k\left(\frac{\delta H}{\delta\rho}\right)\right] + \frac{\delta F}{\delta\rho}\partial_l\rho\left(\frac{\delta H}{\delta p_l}\right) \tag{4.25a}$$
$$+ \frac{\delta F}{\delta\pi_k}\left[(\pi_l\partial_k + \partial_l\pi_k)\left(\frac{\delta H}{\delta\pi_l}\right) + \sigma\partial_k\left(\frac{\delta H}{\delta\sigma}\right)\right] + \frac{\delta F}{\delta\sigma}\partial_l\sigma\left(\frac{\delta H}{\delta\pi_l}\right). \tag{4.25b}$$

Since the superfluid velocity $\mathbf{v}^s$ equals $\mathbf{a} - \nabla\alpha$, the corresponding superfluid momentum density $\mathbf{P}$ is given by

$$P_k = \rho a_k + p_k, \tag{4.26}$$

so that

$$\Phi(P_k) = \rho a_k + \Phi(p_k) = \rho a_k - \rho\alpha_{,k}\,[\text{by } (4.23)] = \rho(a_k - \alpha_{,k}) = \rho v_k^s.$$

Therefore, to get the second Poisson bracket for rotating ^{4}He, we have to make an invertible change of variables $\mathbf{P} = \rho\mathbf{a} + \mathbf{p}$, resulting [by formula (2.20)] in the desired expression

$$\{H,F\} \sim \frac{\delta F}{\delta M_k}\left[(M_l\partial_k + \partial_l M_k)\left(\frac{\delta H}{\delta M_l}\right) + (\partial_l a_k - a_{l,k})\left(\frac{\delta H}{\delta a_l}\right) + (P_l\partial_k + \partial_l P_k)\left(\frac{\delta H}{\delta P_l}\right)\right.$$
$$\left. + \rho\partial_k\left(\frac{\delta H}{\delta\rho}\right)\right] + \left[\frac{\delta F}{\delta a_k}(a_{k,l} + a_l\partial_k) + \frac{\delta F}{\delta P_k}(P_l\partial_k + \partial_l P_k) + \frac{\delta F}{\delta\rho}\partial_l\rho\right]\left(\frac{\delta H}{\delta M_l}\right) \tag{4.27a}$$
$$+ \frac{\delta F}{\delta P_k}\left[(P_l\partial_k + \partial_l P_k)\left(\frac{\delta H}{\delta P_l}\right) + \rho\partial_k\left(\frac{\delta H}{\delta\rho}\right)\right] + \frac{\delta F}{\delta\rho}\partial_l\rho\left(\frac{\delta H}{\delta P_l}\right) + \frac{\delta F}{\delta P_k}\rho(a_{l,k} - a_{k,l})\frac{\delta H}{\delta P_l} \tag{4.27a'}$$
$$+ \frac{\delta F}{\delta\pi_k}\left[(\pi_l\partial_k + \partial_l\pi_k)\left(\frac{\delta H}{\delta\pi_l}\right) + \sigma\partial_k\left(\frac{\delta H}{\delta\sigma}\right)\right] + \frac{\delta F}{\delta\sigma}\partial_l\sigma\left(\frac{\delta H}{\delta\pi_l}\right). \tag{4.27b}$$

The bracket (4.27) is no longer linear, having the quadratic term (4.27a′) in it.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation and by the U.S. Department of Energy.

[1] V. V. Lebedev and I. M. Khalatnikov, "Hamiltonian hydrodynamic equations for a quantum fluid in the presence of solitons," Sov. Phys. JETP **48**, 1167 (1978).
[2] I. E. Dzyaloshinskii and G. E. Volovick, "Poisson brackets in condensed matter physics," Ann. Phys. (NY) **125**, 67 (1980).
[3] D. D. Holm and B. A. Kupershmidt, "Poisson structures of superfluids," Phys. Lett. A **91**, 425 (1982).
[4] B. A. Kupershmidt, "Mathematical aspects of quantum fluids. I. Generalized two-cocycles of ^{4}He type," J. Math. Phys. **26**, 2754 (1985).
[5] B. A. Kupershmidt, "Mathematical aspects of quantum fluids. II. ^{4}He, and Clebsch representations of symplectic two-cocycles," J. Math. Phys. **27**, 2437 (1986).
[6] B. A. Kupershmidt, *Discrete Lax Equations and Differential-difference Calculus* (Asterisque, Paris, 1985).